

Marcus Weigl, BSc

Abschätzung der Unsicherheiten von Belastungsszenarien einer Ortsnetzstation mittels maschinellen Lernens für ein dezentrales Verfügbarkeitsmanagement

Masterarbeit

Ausgeführt zum Zweck der Erlangung des akademischen Grades

Diplom-Ingenieur (Dipl.-Ing.)

Eingereicht an der

Technischen Universität Wien

Fakultät für Elektrotechnik und Informationstechnik

Unter der Leitung von

Univ.-Prof. Dr.-Ing. Wolfgang Gawlik
und

Projektass. Dipl.-Ing. Irina Lupandina
Projektass. Dipl.-Ing. Michael Schrammel

Institut für Energiesysteme und Elektrische Antriebe

Wien, März 2022



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Danksagung

Ein großes Danke gebührt meinen Betreuern Dipl.-Ing. Irina Lupandina und Dipl.-Ing. Michael Schrammel, die mich über das vergangene Jahr meiner Masterarbeit mit ihrer Expertise und den fachlichen Diskussionen unterstützten, viel Zeit aufbrachten und damit einen wichtigen Beitrag zum Gelingen dieser Arbeit beisteuerten. Weiters möchte ich mich bei meiner Mutter Josefine bedanken, die mich über die komplette Studienzeit unterstützte und stets positiv darauf einwirkte. Zu guter Letzt gebührt ein außerordentlicher Dank meiner lieben Frau Maria, die diese Extended-Version meines Studiums mitrug und mir in den guten wie den schlechten Phasen zur Seite stand.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass die vorliegende Arbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen von mir selbstständig erstellt wurde. Alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur, sind in dieser Arbeit genannt und aufgelistet. Die aus den Quellen wörtlich entnommenen Stellen sind als solche kenntlich gemacht.

Das Thema dieser Arbeit wurde von mir bisher weder im In- noch Ausland einer Beurteilerin/einem Beurteiler zur Begutachtung in irgendeiner Form als Prüfungsarbeit vorgelegt. Diese Arbeit stimmt mit der von den Begutachterinnen/Begutachtern beurteilten Arbeit überein.

Ort, Datum

Unterschrift

Kurzfassung

Der Ausbau an erneuerbaren Energien, der steigende Anteil an E-Mobilität oder die Sektorkopplung von Elektrizität und Wärmeversorgung sind Herausforderungen, welche die Stromnetze von morgen zu bewältigen haben. Die Netzteilnehmer bringen neue Unsicherheiten in Bezug auf die Auslastung der Netzkomponenten mit sich, wodurch traditionelle Planungs- und Betriebskonzepte der Stromnetze an ihre Grenzen stoßen.

In dieser Arbeit wird eine Methode vorgestellt, in welcher mittels maschinellen Lernens die Unsicherheiten von Belastungsszenarien abgeschätzt werden können. Dabei werden zwei verschiedene Clusteralgorithmen untersucht und auf einen netzbeschreibenden Input-Datensatz angewendet. Die resultierenden Cluster werden anschließend für die Prognose und Abschätzung von Belastungsszenarien verwendet und die Unsicherheiten mittels Dichtefunktion dargestellt.

Die Ergebnisse zeigen, dass die verwendeten Clusteralgorithmen den Input-Datensatz in mehrere Cluster unterteilen können. Die Unterscheidbarkeit zwischen den Clustern ist vor allem bei der Tageszeit zu erkennen. Die tageszeitabhängige Unterscheidbarkeit lässt sich auch in den restlichen Variablen des Input-Datensatzes erkennen und unterstreicht diese nur, als dass ein zusätzliches Unterscheidungsmerkmal gezeigt wird. Die darauf aufbauende Abschätzung der Unsicherheiten zeigt anhand unterschiedlicher Belastungsszenarien, dass die Unsicherheiten trotz Clusterings nur bedingt abzuschätzen sind. Die Unsicherheiten der Transformatorauslastung können im Vergleich zur Jahresauslastung verbessert abgeschätzt werden, jedoch nicht wesentlich im Vergleich mit erzeugten Test-Daten der Transformatorauslastung.

Abstract

The expansion of renewable energies, the increasing share of e-mobility or the sector coupling of electricity and heat supply are challenges that the electricity grids of tomorrow will have to overcome. The grid participants bring new uncertainties with regard to the utilisation of the grid components, which means that traditional planning and operating concepts for electricity grids are reaching their limits.

In this paper, a method is presented in which machine learning can be used to estimate the uncertainties of load scenarios. Two different cluster algorithms are investigated and applied to a grid-describing input data set. The resulting clusters are then used for forecasting and estimating load scenarios and the uncertainties are represented by means of a density function.

The results show that the cluster algorithms used were able to divide the input dataset into several clusters. The discriminability between the clusters can be seen especially for the time of day. The time of day dependent distinctiveness can also be seen in the rest of the variables of the input data set and only underlines them as showing an additional distinguishing feature. The estimation of uncertainties based on this shows, using different load scenarios, that the uncertainties can only be estimated to a limited extent despite clustering. The uncertainties of the transformer load can be estimated better in comparison with the annual load, but not significantly in comparison with generated test data of the transformer load.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Inhalt

1.	Einleitung	1
1.1.	Motivation und Ziele dieser Arbeit	4
1.2.	Forschungsfragen	4
1.3.	Aufbau der Arbeit	4
2.	Theoretische Grundlagen	5
2.1.	Elektrische Energiesysteme	5
2.1.1.	Planungshorizonte	5
2.1.2.	Engpassmanagement	6
2.1.3.	Standardisierte Standardlastprofile	7
2.1.4.	Einflussfaktoren auf den Energiekonsum	8
2.1.5.	Aggregieren von Lastprofilen	9
2.1.6.	SimBench, ein Benchmark-Datensatz	10
2.2.	Statistische Auswertung der Belastungsszenarien	14
2.2.1.	Stochastische Parameter	14
2.2.2.	Monte-Carlo-Simulation	15
2.2.3.	Informationskriterium	16
2.3.	Maschinelles Lernen	16
3.	Methodik	20
3.1.	Auswahl des untersuchten Stromnetzes	20
3.2.	Durchführung der Lastflussberechnungen	28
3.3.	Generierung eines Input-Datensatzes	29
3.4.	Input-Datensatz für entsprechende Transformatorstation zusammensetzen	32
3.5.	Anwendung des Clusteralgorithmus	33
3.5.1.	Erwartungs-Maximierungs-Algorithmus	34
3.5.2.	Distanzmatrix-Algorithmus	36
3.6.	Berechnen der Cluster-Dichtefunktion	41
3.7.	Prognose für Belastungsszenario	42
3.8.	Evaluierung der Prognose	43
4.	Ergebnisse	45
4.1.	Clustering	47
4.1.1.	Erste Evaluierungsrunde	47
4.1.2.	Zweite Evaluierungsrunde	70
4.1.3.	Fazit Clustering	74
4.2.	Prognose der Unsicherheiten von Belastungsszenarien	74
5.	Schlussfolgerung und Ausblick	78
	Literaturverzeichnis	81
	Abbildungsverzeichnis	84
	Tabellenverzeichnis	86



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1. Einleitung

Die Geschichte der elektrischen Energieversorgung begann in Österreich vor mittlerweile über 130 Jahren. Im Jahr 1886 wurde in der Gemeinde Scheibbs in Niederösterreich das erste öffentliche Stromnetz für eine elektrisch betriebene Straßenbeleuchtung errichtet. Mit diesem Startschuss wurden nach und nach immer mehr und zunächst noch regional betriebene Stromnetze in Österreich gegründet. Die Nachfrage nach elektrischer Energie stieg und in den darauffolgenden Jahren erfuhren diese privaten und kommunalen Stromnetze verschiedenste politische und wirtschaftliche Umstrukturierungen, wuchsen mit zunehmendem Energiebedarf zu immer größeren Elektrizitätsgemeinschaften über mehrere Spannungsebenen hinweg zusammen und bildeten folglich ein gesamtheitlich verbundenes Stromnetz [1]. Aus den anfänglich regionalen Stromnetzen wurde ein überregionales und schlussendlich kontinentales europäisches Verbundnetz.

Wie an der kurz gefassten, aber langen Historie von Stromnetzen dargelegt werden soll, entwickelten sich die Energienetze von sehr einfach aufgebauten Netzen hin zu weit vermaschten und komplexen Stromnetzen. Aus wenigen Erzeugereinheiten und Verbrauchern wurde eine fast unüberschaubare Anzahl. Trotz dieser Komplexität ist der Versorgungssicherheit und -qualität von elektrischer Energie allerhöchste Priorität beigemessen und wird auch auf einem sehr hohen Niveau in Österreich erreicht [2]. In diesem Kontext sind besonders zwei Bereiche von essenzieller Bedeutung: Netzplanung und Netzbetrieb.

Netzplanung und Netzbetrieb für eine hohe Versorgungssicherheit

Seitdem elektrische Energie der breiten Öffentlichkeit zugänglich gemacht wurde und die Nachfrage stetig anstieg, sind Netzplanung und Netzbetrieb die Grundlage für eine sichere und zuverlässige Versorgung mit elektrischer Energie. Die Mehrheit der elektrischen Energie wird in größeren Erzeugerzentren erzeugt und über die Übertragungs- und Verteilernetze zu den Verbrauchern transportiert. Diese unidirektionale Energieversorgung, also ein Energiefluss rein vom Erzeuger hin zum Verbraucher, war über viele Jahre hinweg die zentrale oder auch einzige Funktionsart der Stromnetze. Die simplen Netzteilnehmer ermöglichten eine gute Planbarkeit von Kraftwerksfahrplänen (Netzbetrieb) oder von Ausbauten der Netzinfrastruktur (Netzplanung). Diese konstanten Annahmen spiegeln sich deshalb auch in der Dimensionierung der Netze und deren Komponenten wider.

Bei der Auslegung der verschiedenen Netzkomponenten wie Transformatoren, Freileitungen oder Erdkabel, hat sich vor allem die Berechnungsmethode mit einem deterministischen Ansatz und dazu Annahmen von Extremzuständen etabliert und findet auch heute noch überwiegend Anwendung [3]. Bei diesen Extremzuständen (Worst-Case-Szenarien) werden beispielsweise Betriebszustände mit maximalem Verbrauch und minimaler Erzeugung oder minimalem Verbrauch und maximaler Erzeugung angenommen. Unter Berücksichtigung von Gleichzeitigkeitsfaktoren müssen alle Netzkomponenten diese Betriebszustände bewältigen können. Solche Extremzustände stellen jedoch nur einen kleinen Teil aller auftretenden Lastzustände dar, wodurch die meiste Zeit die Netzkomponenten nur teilweise ausgelastet sind und diese ungenutzten Netzkapazitäten zusätzliche Kosten bedeuten. Auch die steigende Anzahl an Netzteilnehmern in den Verteilernetzen lässt diese Kosten zunehmend steigen. Was früher ohne intelligente, flexible oder vernetzte Netzteilnehmer bei der Planung nicht anders möglich war, kann heute als zusätzlicher Einflussfaktor in die Netzplanung einfließen. Welches Potential in der netzdienlichen Nutzung von Flexibilitäten steckt, zeigt die folgende Studie zum Netzausbau für das Jahr 2035 in deutschen Verteilernetzen [4]:

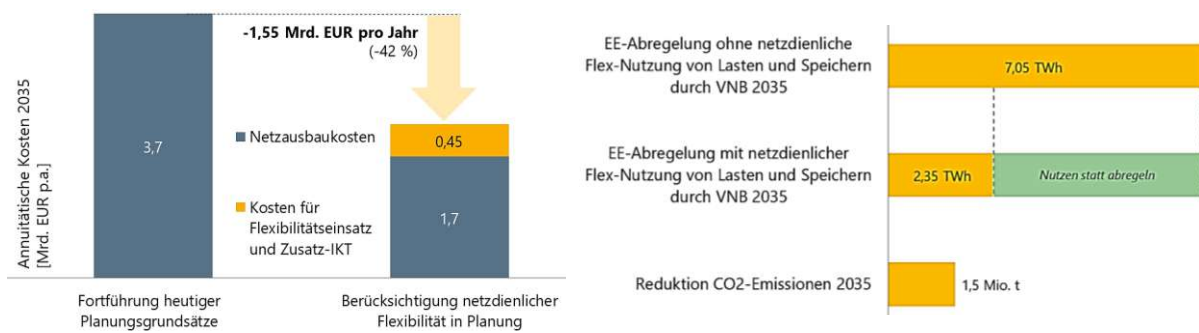


Abbildung 1: Potential netzdienlicher Flexibilitäten [4]

Abbildung 1 zeigt die Haupteigenschaften der Kurzstudie, welche im Auftrag von drei deutschen Verteilernetzbetreibern erstellt wurde. Bei einer Berücksichtigung netzdienlicher Flexibilitäten (z.B. Elektroautos, elektrische Wärmeerzeuger, Batteriespeicher, etc.) während der Planung, können Kosten in der Höhe von 42% im Vergleich zu heutigen Planungsgrundsätzen eingespart werden. Weiters kann bei der Nutzung von flexiblen Lasten und Speichern mehr erneuerbare Energie in das Verteilernetz integriert und so CO₂-Emissionen reduziert werden.

Im Bereich des Netzbetriebs spielt auch das so genannte N-1-Kriterium eine zentrale Rolle. Dieses Kriterium stellt sicher, dass auch beim Ausfall einer Netzkomponente (z.B. Unterbrechung einer Freileitung) weiterhin ein sicherer Betrieb des Netzes gewährleistet ist und die restlichen Komponenten keine Grenzwerte überschreiten. Im österreichischen Übertragungsnetz wird dieses Kriterium beispielsweise alle 5 Minuten neu berechnet und gegebenenfalls Schalthandlungen vorgenommen [5]. Um dieses Kriterium einzuhalten, werden Betriebsmittel oft nur bis zur Hälfte ihrer möglichen Leistungsgrenzen betrieben. Da dieses Kriterium durch die immer komplexeren Netze und steigende Anzahl an dezentralen Erzeugereinheiten in Zukunft schwieriger einzuhalten wird, ist auch hier eine Anpassung bzw. Neuauslegung anzudenken [6].

Ausbau der Erneuerbaren: Green Deal und Erneuerbaren-Ausbau-Gesetz

Getrieben durch den Klimawandel und den damit verbundenen politischen Zielen für einen klimaneutralen Kontinent, muss der Anteil an erneuerbaren Erzeugungsanlagen an der Stromproduktion steigen und tut dies auch in einem stetig steigenden Ausmaß. Die stufenweisen gesetzten Ziele in der Europäischen Union, wie beispielsweise das „Klima- und Energiepaket 2020“ bis zum Jahr 2020 und das Folgepaket „Fit for 55“ bis zum Jahr 2030, sind die gesetzten Meilensteine des „European Green Deal“, welche Europa als ersten klimaneutralen Kontinent der Welt bis 2050 entstehen lassen wird. Erst im Juli vergangenen Jahres hat sich die Europäische Kommission auf eine Verschärfung der Klimapolitik geeinigt, um dieses ambitionierte Ziel zu erreichen. [7]

Österreich hat sich in diesem Kontext noch straffere Ziele gesetzt und möchte schon im Jahr 2030 zu 100% Strom aus erneuerbaren Energiequellen erzeugen. Dazu wurde, ebenfalls im Juli vergangenen Jahres, das Erneuerbaren-Ausbau-Gesetz (kurz EAG) beschlossen. Neben einer neuen Förderungsstruktur und einfacheren Rahmenbedingungen für Bürgerenergiegemeinschaften, sind vor allem die gesetzten Ausbauziele der Ökostromanlagen eine Herausforderung für das Stromnetz. Bis 2030 sollen Ökostromanlagen in der Größenordnung von rund 27 TWh (das entspricht einer Steigerung um 50% zur aktuell bestehenden Ökostrom-Leistung) realisiert werden [8]. Man erkennt schnell, dass diese politischen Bestrebungen eine Umstrukturierung für das historisch gewachsene Stromnetz

bedeuten, weg von der bedarfsorientierten Stromerzeugung durch kalorische Kraftwerke und hin zu einer gebotsorientierten Stromerzeugung mit erneuerbaren Energiequellen. Wie sehr sich dieser Wandel auf die Auslastung der verschiedenen Netzkomponenten auswirkt, ist ein zentraler Punkt in dieser Arbeit.

Anforderungen an das Stromnetz von morgen

Die Stromnetze von morgen sind auch noch mit weiteren Herausforderungen der Energiewende konfrontiert. Beispielsweise kommt es durch die Förderung der E-Mobilität und den wachsenden Anteil an E-Autos zu neuen Lastspitzen und Belastungsszenarien in den Stromnetzen. Auch der Umstieg auf elektrische Energie in Bereichen wie der Wärmeerzeugung und die Nutzung von intelligent geregelten Geräten verändert den Bedarf an elektrischer Energie. Vor allem verändert sich die zeitliche Nutzung von elektrischer Energie und bringt neue Rahmenbedingungen für die Netzkomponenten und die Anwendung von Belastungsszenarien. Welche neuen Anforderungen an das Stromnetz von morgen gestellt werden, hat beispielsweise das Verbundprojekt „Smart Area Aachen“ im Zeitraum 2012 bis 2017 untersucht:

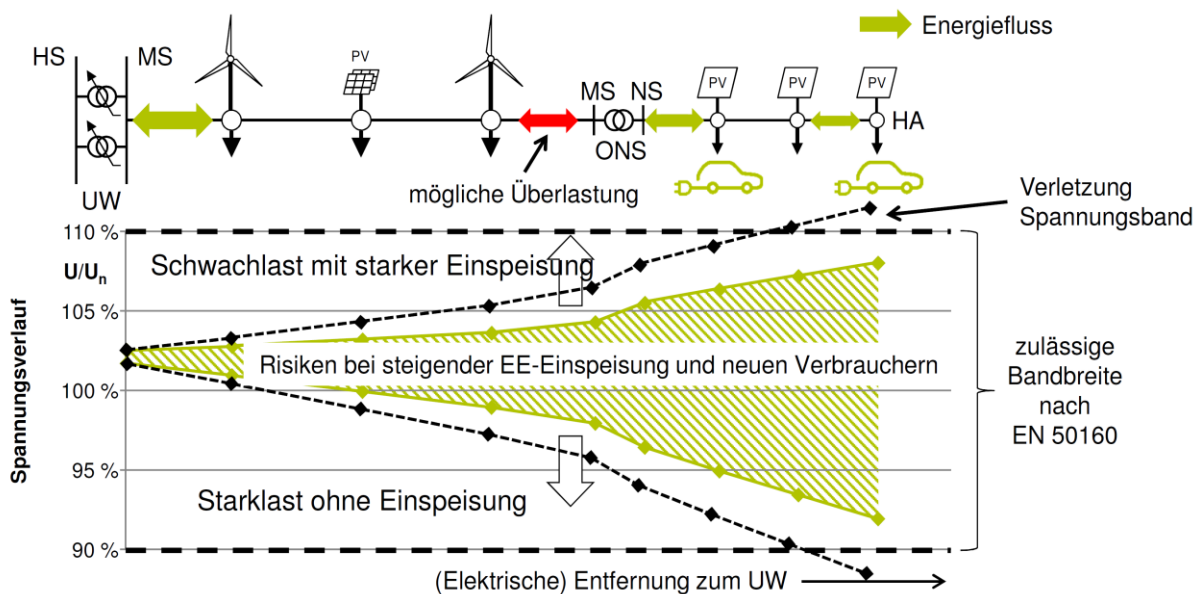


Abbildung 2: Neue Anforderungen an das Stromnetz von morgen [9]

Abbildung 2 zeigt einen möglichen Spannungsverlauf an einem Netzabgang und die dabei auftretenden Spannungsunsicherheiten durch die steigende Einspeisung der erneuerbaren Energien und neuen Verbraucher (E-Mobilität, E-Speicher, intelligente Endgeräte, etc.). Je genauer diese Belastungsszenarien und damit verbundenen Unsicherheiten abgeschätzt werden können, desto zielgerichteter kann auf die verschiedenen Belastungsszenarien von Seiten des Netzbetreibers bei Bedarf reagiert werden.

1.1. Motivation und Ziele dieser Arbeit

Die oben angesprochenen Punkte erfordern eine neue Herangehensweise bei der Netzplanung, aber vor allem eine bessere Abschätzung der Unsicherheiten der Belastungsszenarien, um die Komponenten besser ausnutzen zu können. Mittels statistischer Auswertung der Belastungsszenarien und Analyse durch maschinelles Lernen, können diese Herausforderungen intelligent ausgearbeitet werden. Mit Hilfe von Clusteranalysen können beispielsweise Ähnlichkeiten und Strukturen in einer gegebenen Menge an Daten identifiziert und diese in unterschiedliche Cluster unterteilt werden. Dadurch können zukünftige Daten einfacher verarbeitet bzw. auch vorhergesagt werden, wenn durch das Clustering Strukturen gefunden und diese Strukturen auf die Vorhersage angewendet werden.

Ziel dieser Arbeit ist es, die Anwendbarkeit von Clusteralgorithmen auf Daten in einem Stromnetz zu untersuchen. Mit den geclusterten Daten soll anschließend eine Prognose der Unsicherheiten von Belastungsszenarien getätigt werden, welche in einem dezentralen Verfügbarkeitsmanagement als Entscheidungshilfe dient.

1.2. Forschungsfragen

Die folgenden Forschungsfragen werden im Rahmen dieser Arbeit detailliert ausgearbeitet:

- Wie können die Unsicherheiten von Belastungsszenarien einer Ortsnetzstation abgeschätzt werden?
- Wie ist die Güte der angewendeten Algorithmen in Bezug auf die Clusterergebnisse?
- Welchen Mehrwert bietet die getätigte Abschätzung der Unsicherheiten von Belastungsszenarien?
- Welche Verbesserungen oder Weiterentwicklungen müssen für eine brauchbare Abschätzung gemacht werden?

1.3. Aufbau der Arbeit

Zunächst werden im Theorieteil die Grundlagen für die Bearbeitung und Beantwortung der Forschungsfragen aufgefrischt und die wichtigsten Themenbereiche festgehalten. Anschließend wird die Methodik vorgestellt, welche im Zuge dieser Arbeit entwickelt wurde und zur Durchführung dieser Arbeit diente. Im darauffolgenden Kapitel werden die Ergebnisse der durchgeführten Forschung beschrieben und diskutiert. Zuletzt erfolgt eine Interpretation der Ergebnisse und ein Anstoß für nachfolgende Arbeiten in diesem Themengebiet.

2. Theoretische Grundlagen

Im theoretischen Teil dieser Arbeit werden jene Themengebiete detaillierter beschrieben, welche für die Ausarbeitung und Analysen erforderlich waren. Es wird in den anschließenden Kapiteln auf die entsprechenden Stellen im theoretischen Teil verwiesen, damit die Analysen und Berechnungen bestmöglich nachvollzogen werden können.

2.1. Elektrische Energiesysteme

Elektrische Energie ist die wichtigste Energieform in unserer modernen Welt. Sie kann in jegliche andere Form von Energie umgewandelt werden, deckt unterschiedliche Bedürfnisse und ist verlustarm über weite Strecken zu transportieren. Für diesen verlustarmen Transport sind elektrische Energiesysteme notwendig, die mit Sorgfalt und technischer Expertise unseren modernen Lebensstil gewährleisten.

2.1.1. Planungshorizonte

Prognosen und Planungsaufgaben kommen im Zusammenhang mit Stromnetzen in unterschiedlichsten Formen und mit unterschiedlichsten Zeithorizonten vor. Beispielsweise besitzt die Lastprognose im Stromnetz eine zentrale Rolle für den stabilen Betrieb eines Stromnetzes, denn Erzeugung und Verbrauch von elektrischer Energie müssen sich zu jedem Zeitpunkt die Waage halten. Andererseits sind auch Langzeitplanungen, wie beispielsweise Netzentwicklungspläne oder Kraftwerksplanungen, unverzichtbar, da diese Aufgaben eine längere Vorlaufzeit bis zur Realisierung benötigen. Man kann die Planungszeithorizonte in 5 Abschnitte unterteilen, wobei jedem Abschnitt unterschiedliche Aufgaben zugeordnet werden und sich gegebenenfalls auch auf einen nachfolgenden Abschnitt/Zeithorizont auswirken:

Strategische Langzeitplanungen betreffen, wie schon oben eingangs erwähnt, beispielsweise Netzausbauten, Kraftwerksplanungen oder deren Stilllegungen und die damit verbundenen Finanz- und Kostenplanungen. Der Zeithorizont erstreckt sich hierbei über 5 bis 20 Jahre und beinhaltet sehr oft Unsicherheiten oder Annahmen von Szenarien, wodurch vereinfachte Verfahren für diese Planungen ausreichend sind. **Langfristige Planungen** betreffen Bezugs- und Lieferverträge von Kraftwerken und die damit verbundenen Regelbereiche der Erzeugungseinheiten. Die Jahreshöchstlasten und Jahresenergiemengen werden dazu als Jahreshöchstlasten für den Zeithorizont von 1 bis 5 Jahren erstellt und liefern wichtige Prognoseparameter wie beispielsweise Kosten, Emissionen oder Versorgungsrisiken. **Mittelfristige Planungen** betreffen vor allem die Brennstoffplanung, deren Bevorratung samt Bewirtschaftung und erstrecken sich zwischen 1 Woche und 1 Jahr. Die aus den vorhergehenden Planungen verwendeten Eingangsparameter resultieren hier in Tageslastganglinien. **Kurzfristige Planungen** sind im Bereich von 1 Tag bis zu 1 Woche vorhanden und liefern anhand der Bezugs- und Liefervereinbarungen Lastganglinien im 15min- oder 1h-Intervall. Diese Planungen erfordern aufgrund des finanziellen Risikos eine sehr hohe Genauigkeit der Prognose, um die Risiken so gut es geht zu minimieren. **Sehr kurzfristige Planungen** betreffen vor allem die Bereitstellung von Regelenergie wie die Primär- oder Sekundärregelenergieleistung oder Tertiärreserve. Der Zeithorizont betrifft hier Sekunden bis zu wenigen Stunden. Auch die Überwachung von Grenzwerten und damit verbundenen Schalthandlungen fallen in diesen Bereich. [10]

Auch im Zusammenhang mit dem Energiemarkt und dem Handel von elektrischer Energie gibt es eine solche Unterteilung der Zeithorizonte. Jeder Zeithorizont kann als eigener Markt verstanden werden, bei dem unterschiedliche Energieprodukte gehandelt werden. Ein

Energieversorgungsunternehmen muss beispielsweise durch den Erwerb und die Kombination von unterschiedlichen Energieprodukten die Energienachfrage seiner Kunden genau abdecken. Hier lässt sich einerseits zwischen Energieprodukten mit längerem Zeithorizont und andererseits mit sehr kurzfristigem Zeithorizont unterscheiden, was vergleichbar ist mit einer zunächst groben Abschätzung des Energiebedarfs und der abschließenden Feinjustierung.

Den Markt mit dem größten Zeithorizont stellt der **Terminmarkt** dar, an dem Energieprodukte über die nächsten Jahre gehandelt werden. Als Energieprodukte existieren Base- und Peak-Blöcke. Der Base-Block stellt eine kontinuierliche Energielieferung (0-24 Uhr und Montag - Sonntag) innerhalb des bezogenen Zeitraumes (Monat, Quartal, Jahr, etc.) dar. Der Peak-Block hingegen beschränkt sich auf die Energielieferung zu klassischen Geschäftszeiten zwischen 8-20 Uhr und das Montag - Freitag. Dem gegenüber steht der sogenannte **Spotmarkt**, an dem Energieprodukte sehr kurzfristig noch gehandelt werden können. Dieser Markt unterteilt sich wiederum in den Day-Ahead- und den Intraday-Markt. Wie der Name schon erahnen lässt, werden am **Day-Ahead-Markt** Energiemengen bis zum Folgetag gehandelt. Dieser Handel dient dazu, dass Energieversorgungsunternehmen ihren prognostizierten Lastgang mit dem geplanten Kraftwerksfahrplan bilanztechnisch ausgleichen. Um kurzfristig noch Fehlmengen oder Prognoseabweichungen ausgleichen zu können, können am **Intraday-Markt** noch Geschäfte bis 5 min vor Liefertermin abgeschlossen werden. Hier werden ausschließlich 15-Minuten-Kontrakte gehandelt. [11]

An den oben angeführten Energiemärkten wird Energie gehandelt, die nach vorliegenden Prognosen der Energieversorgungsunternehmen benötigt wird und deshalb bereitgestellt werden muss. Einen weiteren Energiemarkt, aber mit einer etwas anderen Intension von Energiehandel, stellt der **Regelenergiemarkt** dar. Hier wird Regel- bzw. Ausgleichsenergie und Regel- bzw. Ausgleichsleistung gehandelt, die im optimalen Betrieb eines Stromnetzes nur vorgehalten werden muss und gar nicht zum Einsatz kommt. Diese Reserve wird dennoch benötigt und ist auch von großer Wichtigkeit, da sie bei positiven oder negativen Schwankungen der Frequenz im Stromnetz schnell aktiviert werden kann, das Stromnetz wieder stabilisiert und eventuell sogar ein Blackout verhindert. Ein Anbieter muss also für den gebotenen und zugeschlagenen Zeitraum diese zugeschlagene Menge an Energie und Leistung bereitstellen können. Regelenergie stellt darüber hinaus eine marktbezogene Maßnahme dar, die Netzbetreiber im Rahmen des Engpassmanagements ergreifen und damit das Stromnetz sicher und zuverlässig betreiben können.

2.1.2. Engpassmanagement

Das Engpassmanagement zählt zu den Aufgaben der Übertragungs- und Verteilernetzbetreiber, die damit einen sicheren und zuverlässigen Betrieb gewährleisten. Beispielsweise sind alle größeren Kraftwerksbetreiber verpflichtet, ihre Fahrpläne dem Übertragungsnetzbetreiber am Vortag zu übermitteln. Kommt es bei den anschließenden Netzbelastungsberechnungen zu Lastflüssen im Stromnetz, welche die Netzkapazitäten überschreiten oder Grenzwerte verletzen würden, muss der Netzbetreiber geeignete Maßnahmen setzen, um diese Engpässe zu verhindern. Dem Netzbetreiber stehen hierzu mehrere Maßnahmen zur Verfügung. Da es sich in diesem Fall um einen zu erwartenden Engpass handelt, kann der Netzbetreiber die Kraftwerksbetreiber anweisen, ihre Einsatzplanungen nach Vorgaben anzupassen [12]. Dieses sogenannte **Redispatch** verursacht zusätzliche Kosten, da die abgeänderte Kraftwerkseinsatzplanung von wirtschaftlichen und effizienten Fahrplänen des Kraftwerksbetreibers abweicht. Die Kosten sind Teil des Netznutzungsentgeltes und werden daher vom Kunden getragen. Wie sehr diese

Kosten in den vergangenen Jahren gestiegen sind, kann aus der nachfolgenden Abbildung entnommen werden.



Abbildung 3: Redispatch-Kosten in der APG [13]

Abbildung 3 zeigt die aufgewendeten Redispatch-Kosten des österreichischen Übertragungsnetzbetreibers APG im Zeitraum 2011 bis 2021. Es ist ein enormer Anstieg seit dem Jahr 2014 zu erkennen, welcher sich auf den steigenden Stromverbrauch zurückführen lässt, aber viel mehr auf die fluktuierenden Einspeisungen der erneuerbaren Energieanlagen. Diese Kosten werden beispielsweise durch Modernisierungsmaßnahmen der Netzinfrastruktur so gering wie möglich gehalten und könnten zusätzlich auch durch eine Abschätzung der Unsicherheiten von Belastungsszenarien verbessert werden.

2.1.3. Standardisierte Standardlastprofile

Werden heutzutage Netzberechnungen, im besonderen Lastflussberechnungen durchgeführt, so sind neben den Netzparametern (Kenngrößen aller Betriebsmittel im Stromnetz) auch Daten über das Verbraucherverhalten der angeschlossenen Verbrauchertypen notwendig. Auch für die Abrechnung der netzabhängigen Stromkosten wird das Verbraucherverhalten benötigt. Da es sich für kleinere Verbraucher (weniger als 100.000 kWh Jahresenergieverbrauch) bis vor wenigen Jahren nicht auszahlte, eine eigene und detaillierte Leistungsmessung durchzuführen, wurden sogenannte standardisierte Lastprofile für die unterschiedlichen Verbrauchertypen bereitgestellt. Die am gängigsten und am häufigsten verwendeten Lastprofile sind die des Bundesverbands der Energie- und Wasserwirtschaft e. V. BDEW (bis 2007 noch Verband der Elektrizitätswirtschaft e. V., kurz VDEW). Sie beinhalten Datensätze von 11 Lastgängen und sind dabei jeweils auf einen Jahresenergieverbrauch von 1.000 kWh normiert. Da es sich um standardisierte bzw. aggregierte Lastprofile handelt, folgt ein reales Lastprofil nicht immer dem Standardlastprofil. Die Gültigkeit bei der Verwendung der Standardlastprofile ergibt sich erst ab einer Mindestanzahl an Verbrauchern. Über eine Vielzahl an Verbrauchern gesehen (aggregierte Last) stellt sich ein einzelnes Lastprofil näherungsweise als Standardlastprofil dar (siehe auch Punkt 2.1.5). Beispielsweise ergibt sich

bereits bei einer Anzahl von 200-300 Haushalten eine sinnvolle Anwendbarkeit, da sich die stochastischen Effekte durch die Mittelung der Viertelstundenwerte ausgleichen. [14]

An dieser Stelle soll kurz auf die Intelligente Messgeräte-Einführungsverordnung eingegangen werden, welche die Ausrollung der intelligenten Strommessgeräte (Smart Meter) in Österreich festlegt. Smart Meter zeichnen den Stromverbrauch detailliert auf (15min-Intervall bei Opt-In), übertragen diesen an die Netzbetreiber und legen so den Grundstein für ein modernes Stromnetz. Vorteile sind dabei beispielsweise eine intelligente Netzsteuerung oder die Realisierung von Energiegemeinschaften. Derzeit liegt das Ziel bei einem 95% Anteil an Smart Metern bis zum Ende des Jahres 2024 [15]. Durch die gewonnenen Lastprofile, auch von kleineren Verbrauchern, können die verwendeten standardisierten Lastprofile überarbeitet und netzspezifisch angepasst werden.

2.1.4. Einflussfaktoren auf den Energiekonsum

Betrachtet man die Stromlastgänge von Ein- oder Mehrfamilienhäusern, Bürokomplexen, Gewerbe- oder Landwirtschaftsbetrieben, Kinos etc., so zeigt sich jeweils ein unterschiedlicher Lastgang: Ein- oder Mehrfamilienhäuser haben eine morgendliche und abendliche Lastspitze, Bürokomplexe besitzen eine Mittagsspitze an Werktagen, Gewerbe- und Landwirtschaftsbetriebe eine im Vergleich etwas ausgeglichene Lastkurve und Kinotheater einen stärkeren Verbrauch an Wochenenden. Was man hier selbst mit einem geschulten Blick erkennen kann, ist die Abhängigkeit des Energiekonsums anhand der Zeit. Tages-, Wochen- oder Jahreszeit sind ein wesentlicher Einflussfaktor auf den Energiekonsum in allen Bereichen. Je nach Verbraucherkategorie ergeben sich dadurch die unterschiedlichen Stromlastgänge. Aber nicht nur der Faktor Zeit spielt beim Verhalten des Energiekonsums (im konkreten Fall Elektrizität bzw. Strom) eine Rolle. Um herauszufinden, welche Einflussfaktoren sich auf den Stromverbrauch und das Verbrauchsverhalten auswirken, wird beispielsweise die jeweilige Verbraucherkategorie in anteilmäßige Bereiche aufgeteilt. Diese Bereiche können anschließend einfacher den Einflussfaktoren zugeordnet werden. Wie schwierig eine eindeutige Zuordnung dennoch ist, zeigt das folgende Beispiel anhand des Verbrauchertyps Haushalt. Abbildung 4 zeigt die Aufteilung in unterschiedliche Bereiche der Einflussfaktoren anhand 5 Studien aus unterschiedlichen Ländern und Jahren: Europäische Union [16], Norwegen [17], Taiwan [18], USA [19] und China [20]

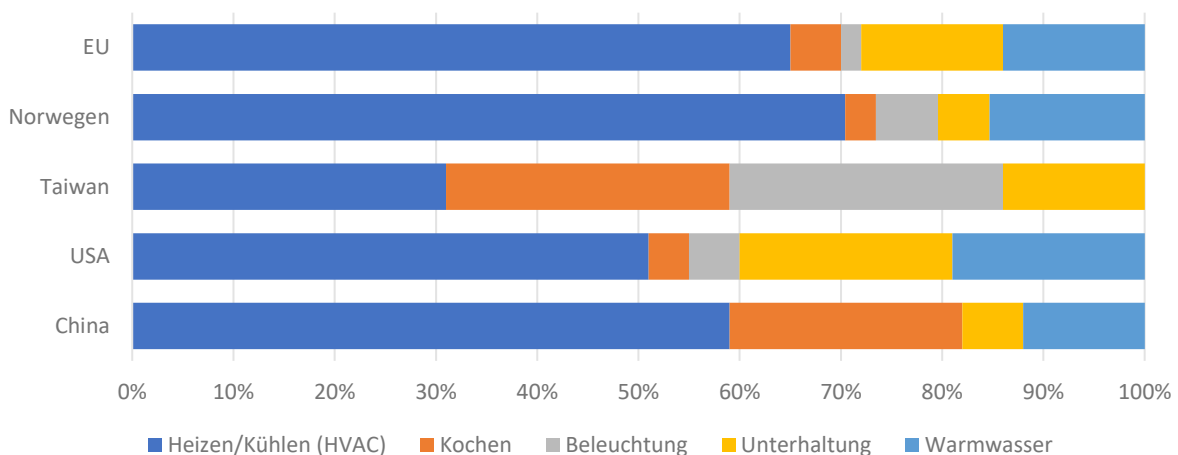


Abbildung 4: Endverbrauchskategorien der elektrischen Energie in Haushalten

Wie zu erkennen ist, ergeben die Daten aus den unterschiedlichen Studien keine einheitliche Aufteilung zwischen den Einflussbereichen. Den größten Anteil macht aber immer der Bereich Heizen/Kühlen aus. Dieser Bereich hängt zu einem sehr hohen Maße von der Außentemperatur und der jeweiligen Jahreszeit ab. Der Bereich Kochen ist im Gegensatz dazu nicht von der Temperatur abhängig, sondern stärker von den Faktoren Tageszeit oder Wochentag.

2.1.5. Aggregieren von Lastprofilen

Wenn im energietechnischen Kontext von Lastprofilen die Rede ist, dann wird, je nach Detailgrad der Betrachtung, ein aggregiertes Lastprofil eines Verbrauchertyps, eines Netzknotens oder eines ganzen Versorgungsgebietes gemeint. Die in der Praxis verwendeten aggregierten Profile weisen eine Samplerate mit bis hinunter auf 15min-Intervalle auf und besitzen einen durchwegs glatten bzw. kontinuierlichen Lastverlauf. Doch betrachtet man ein reales elektrisches Baumittel (z.B. Kühlschrank), so besitzt es keinen so glatten Lastverlauf und zeigt mehr eine Ein-Aus-Charakteristik. Wie kommt es also zu den resultierenden Lastprofilen mit ihrem glatten Verlauf?

Betrachtet man das Lastprofil eines Versorgungsgebietes, so besteht dieses aus unterschiedlichen Verbrauchertypen, wie Gewerbebetrieben, Industriestandorten, Haushalten etc. Diese Verbrauchertypen können weiter in ihre Endverbrauchskategorien (siehe auch Punkt 2.1.4) unterteilt werden, wie Beleuchtung, Raumheizung, Warmwasser etc. Die Raumwärme wird in Einfamilienhäusern beispielsweise mittels Wärmepumpe, Boiler oder Radiator erzeugt. Abbildung 5 soll diesen Sachverhalt bildlich darstellen.

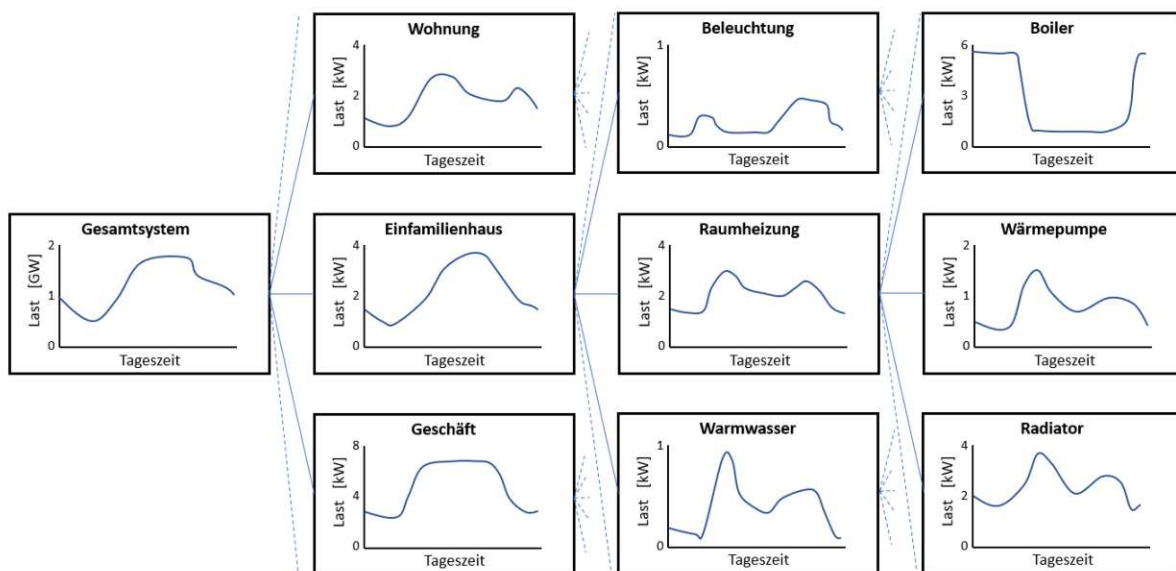


Abbildung 5: Analyse von aggregierten Lastprofilen [21]

Betrachtet man anhand Abbildung 5 beispielsweise den Tageslastgang des Warmwasserboilers, so weist dieser einen höheren Verbrauch in den Morgen- und Abendstunden auf und einen eher niedrigeren Verbrauch über den restlichen Tag. Das reale Betriebsverhalten eines Warmwasserboilers folgt jedoch nicht dieser glatten Lastkurve, sondern funktioniert nach dem Prinzip des Zweipunktreglers. Der reale Lastgang zeigt somit diskrete Lastspitzen, bei denen auf eine vordefinierte Temperatur geheizt wird und Lastpausen, in denen die Temperatur (exakter wäre Temperaturbereich) aktuell erreicht ist (siehe obere Diagramme in Abbildung 6). Würde man dieses reale Lastprofil einfach auf die

Anzahl an Warmwasserboilern in einem Netzgebiet hochskalieren, so würde das resultierende Lastprofil nicht dem gemessenen Lastprofil in diesem Netzgebiet entsprechen und es kommt der Begriff „aggregiert“ ins Spiel. Beim einfachen Hochskalieren von einzelnen Verbrauchern wird nämlich der Gleichzeitigkeitsfaktor außer Acht gelassen. Ein aggregiertes Lastprofil hingegen berücksichtigt den Fakt, dass einzeln betrachtete Verbraucher ein diskretes Lastprofil mit Ein-Aus-Phasen besitzen, aber diese Phasen von Verbraucher zu Verbraucher zeitlich variieren. Dadurch resultiert ab einer größer betrachteten Verbraucheranzahl ein Lastprofil, welches dem aggregierten Profil in Abbildung 6 ähnelt. [21]

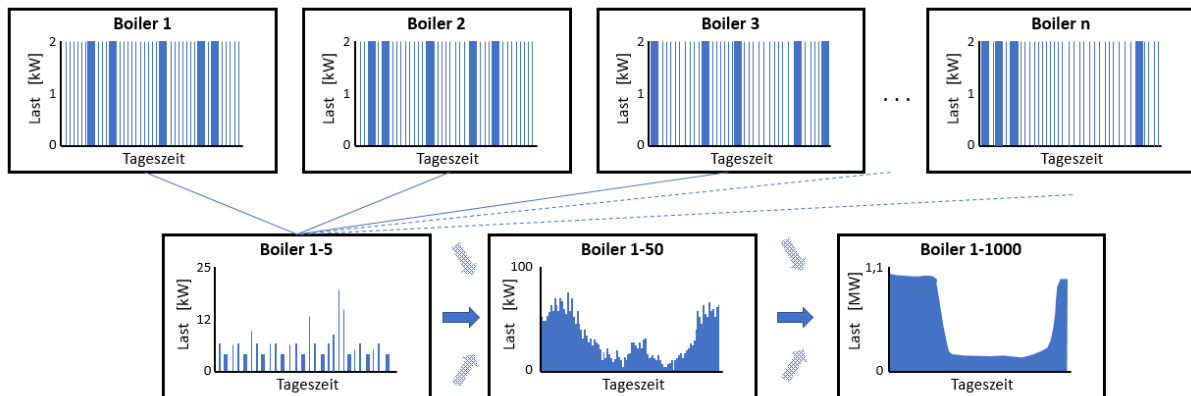


Abbildung 6: Einzellast zu aggregierter Last [21]

Jedes Lastprofil, ob Einzellast oder aggregierte Last, hat seine Gültigkeit in dem jeweiligen Kontext. Das aggregierte Lastprofil stellt demnach auch eine Dichtefunktion dar, zu welchen Zeiten ein Wasserboiler eher in Betrieb ist und wann dieser weniger Energie verbraucht.

2.1.6. SimBench, ein Benchmark-Datensatz

Hinter dem Forschungsprojekt „SimBench - Elektrische Benchmarknetzmodelle“ steht die Motivation, dass durch die Umstellung der elektrischen Netze (Integration von erneuerbaren Energiequellen, wachsende E-Mobilität, Integration von E-Speichern, etc.) auch eine Umstellung der Planung/Berechnung von Netzen erforderlich ist. Es müssen die Netzdaten diesen Anforderungen angepasst bzw. aktualisiert werden, damit zukünftige Stromnetze nach aktuellem Stand der Technik, aber mit Blick in die Zukunft ausgeführt werden können.

Dieses Benchmarknetz, also ein Stromnetz mit der Möglichkeit die Daten mit unterschiedlichen Softwareprogrammen zu verwenden und zu vergleichen, bietet eine umfangreiche Simulationsdatenbasis für Bereiche der Netzanalyse, Netzplanung und -betriebsführung. Nachstehend einige hervorzuhebende Eigenschaften des SimBench-Datensatzes:

- Abgestimmt auf viele Anwendungsfälle, wie beispielsweise eine probabilistische Lastflussberechnung
- Eindeutig zugeordnete Jahreszeitreihen von Verbrauchern und Erzeugern
- Dokumentierter Designprozess für eine nachvollziehbare Datenbasis
- Verfügbarer Datensatz über alle Netzebenen von Höchst- bis Niederspannung

Nachfolgend wird auf die beiden verwendeten Netzebenen (Mittel- und Niederspannungsebene) näher eingegangen. Die Informationen stammen allesamt aus der SimBench Dokumentation [22].

Mittelspannungsebene - MS

Für die Mittelspannungsebene stehen im SimBench-Datensatz vier unterschiedliche Netze zur Verfügung: ländlich, vorstädtisch, städtisch und gewerblich. Diese Netzvarianten unterscheiden sich beispielsweise hinsichtlich ihrer Topologie, Abgangslängen oder Knotenanzahl. Abbildung 7 gibt hierzu einen Überblick der unterschiedlichen MS-Netze.

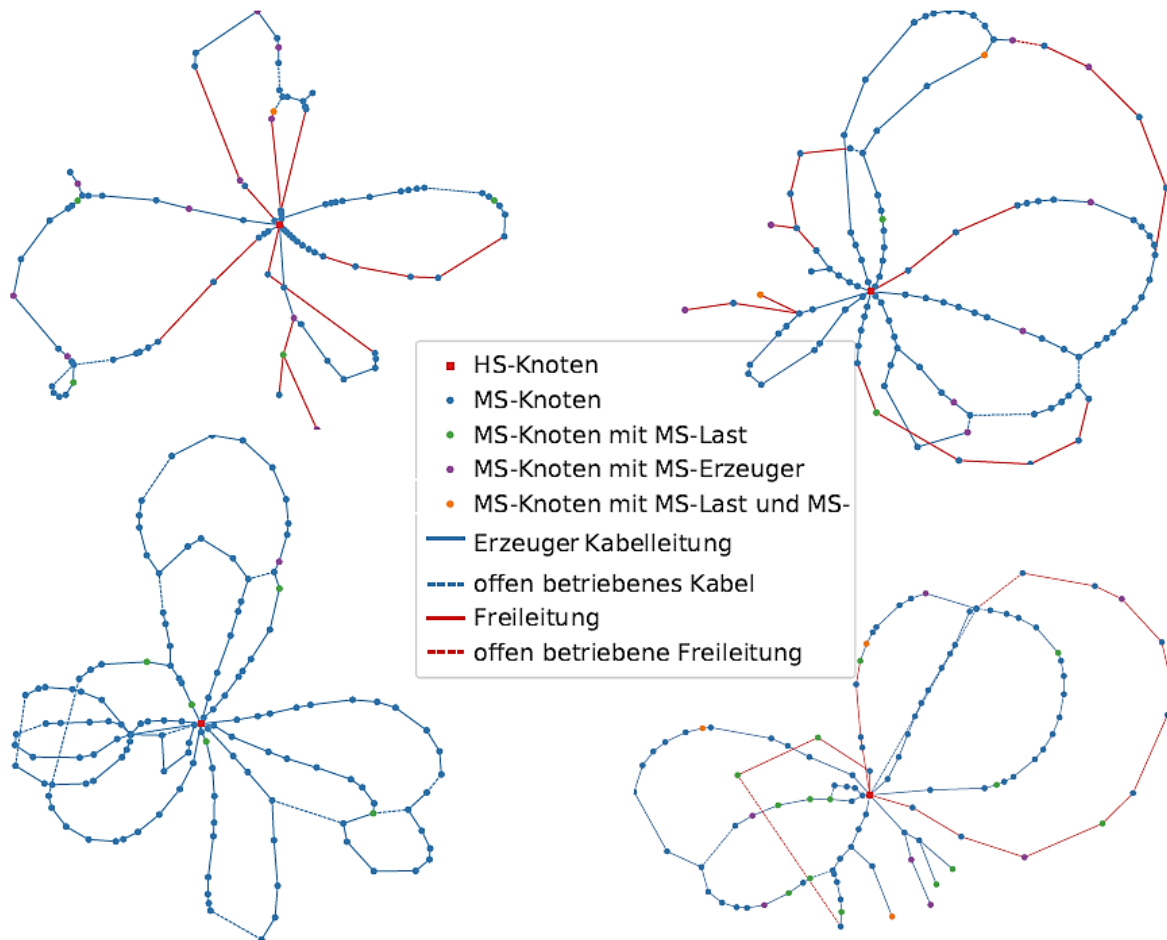


Abbildung 7: Darstellung der MS-Netze: ländlich (links oben), vorstädtisch (rechts oben), städtisch (links unten) und gewerblich (rechts unten)

Anders als bei Höchstspannungs- oder Hochspannungsnetzen, sind öffentliche Informationen oder andere Quellen für eine realitätsnahe Herleitung der Mittelspannungsnetze nicht verfügbar, wodurch diese nicht automatisiert entstanden sind. Auch eine Herleitung, wie bei Niederspannungsnetzen anhand von Straßenzügen, Gebäude- oder Bevölkerungsverteilungen, wäre mit großen Unsicherheiten verbunden. Die Datengrundlage für die Generierung der Mittelspannungsnetze waren 74 separat betriebene MS-Netze von 5 verschiedenen Netzbetreibern. Die so entstandenen Netze sind keinem realen Versorgungsgebiet zugeordnet und rein synthetisch entstanden. Dennoch liegen die Geodaten relativ zu den Einspeisepunkten der Hochspannungsebene, welche aus Daten von realen Übertragungsnetzen abgeleitet wurden. Die Netzparameter der Mittelspannungsnetze wurden so gewählt, dass eine gute Realitätsnähe besteht, wie beispielsweise Abgangslängen wie sie in der Realität vorkommen. Die Verteilung der angeschlossenen Niederspannungsnetze folgt jedoch keinen realistischen Annahmen.

Niederspannungsebene - NS

Für die Niederspannungsebene stehen im SimBench-Datensatz sechs unterschiedliche Netze zur Verfügung: 3x ländlich, 2x vorstädtisch und 1x städtisch. Diese Netzvarianten unterscheiden sich beispielsweise hinsichtlich ihrer Knotenanzahl oder Betriebsmitteltypen. Abbildung 8 gibt hierzu einen Überblick der unterschiedlichen NS-Netze.

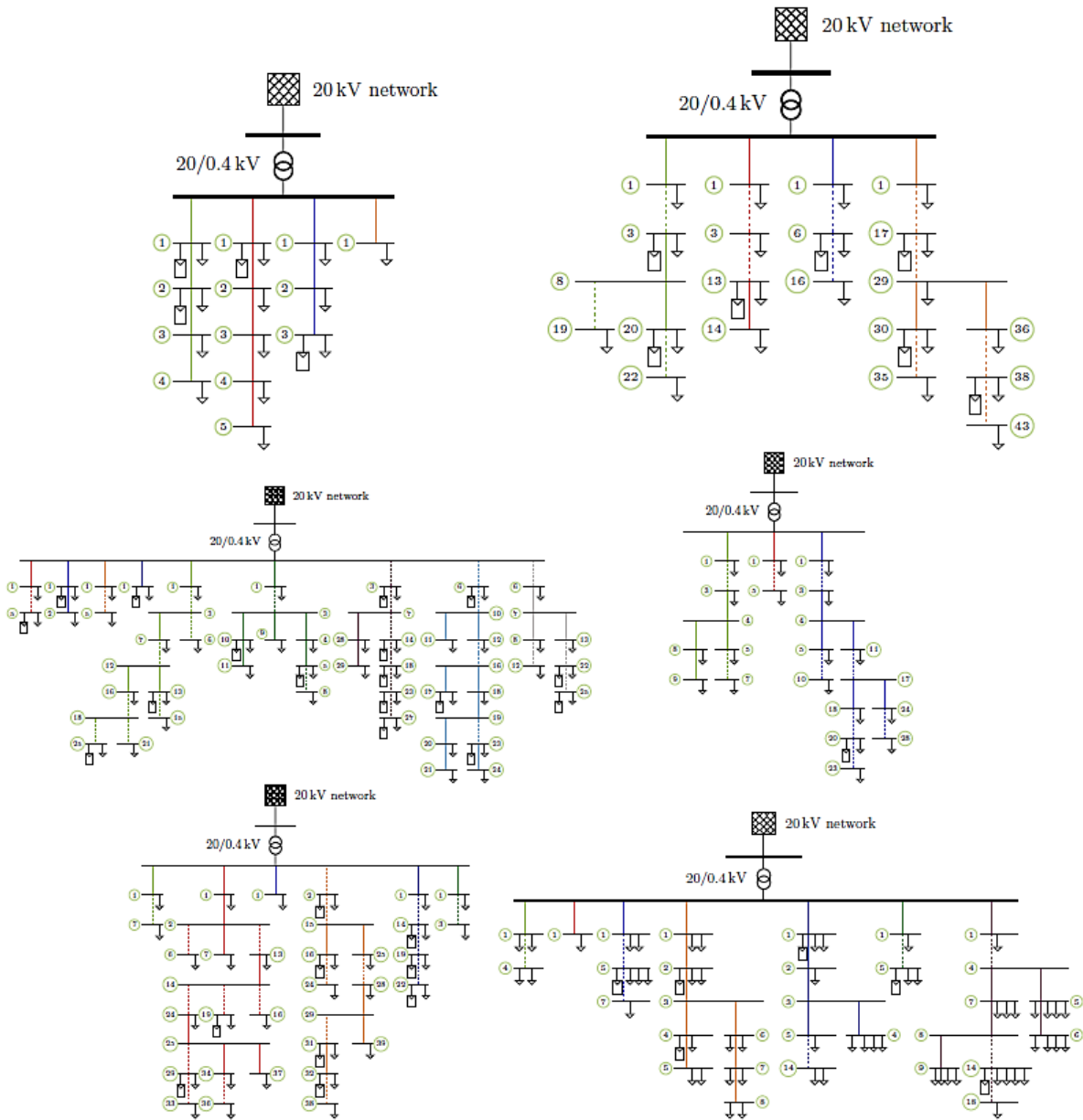


Abbildung 8: Schematische Darstellung der NS-Netze (von oben links nach unten rechts): 3x ländlich, 2x vorstädtisch und 1x städtisch

Die entstandenen Niederspannungsnetze wurden aus OpenStreetMap-Daten generiert und entsprechen prinzipiell realen Geokoordinaten. Da der Fokus im SimBench-Datensatz jedoch mehr auf repräsentativen und durchschnittlichen Netzen liegt, wurden die Geokoordinaten nicht weiterverwendet und verworfen. Um dennoch realitätsnahe Netze zu erhalten, wurden synthetische Koordinaten in Bezug auf die Mittelspannungsebene erzeugt. Die

Erzeugungsanlagen in den einzelnen Niederspannungsnetzen beschränken sich rein auf Photovoltaikanlagen, da Windkraftanlagen in Niederspannungsnetzen die Ausnahme darstellen.

Erzeugung der Jahreszeitreihen

Die Jahreszeitreihen der Gewerbelasten und der landwirtschaftlichen Profile stammen aus der registrierenden Leistungsmessung, welche in einem anonymisierten Datensatz vorliegen und sind aus dem Jahr 2016. Aus diesem Datensatz wurden Zeitreihen ausgewählt, welche am besten den Standardlastprofilen entsprechen und so als repräsentative Lastverläufe dienen. Für die Haushaltslasten kam ein Datensatz der HTW-Berlin zum Einsatz, welcher aus dem Jahr 2010 stammt. Die Jahreszeitreihen der erneuerbaren Erzeugungsanlagen wurden mit Hilfe des Simulationswerkzeugs SIMONA erzeugt. Hierzu wurden als Eingangsparameter Wetterdaten des Deutschen Wetterdienstes aus dem Jahr 2011 und 2012 herangezogen und daraus Zeitreihen für Wind- und Photovoltaikanlagen erzeugt. Zuletzt wurden die Zeitstempel auf das Jahr 2016 angepasst, um sie konsistent mit den anderen Zeitreihen zu haben. Die durch diese Methodik entstandenen Jahreszeitreihen sind übersichtlich in Abbildung 9 dargestellt.

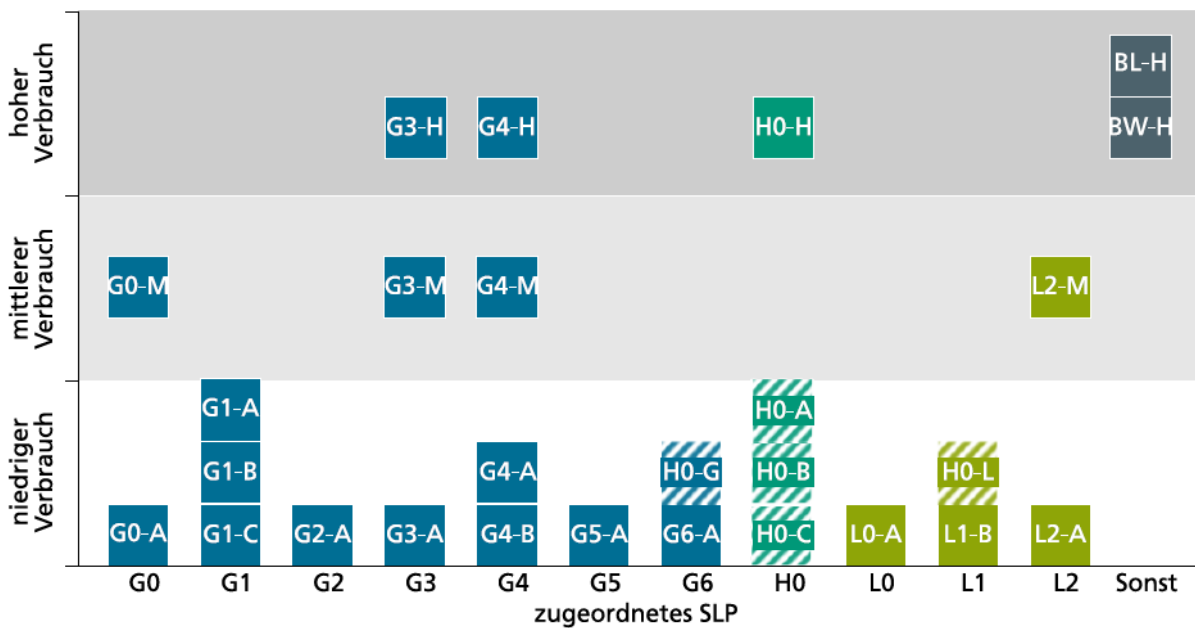


Abbildung 9: Gesamtübersicht aller zur Verfügung gestellter Profile für Sammelverbraucher

Bei den Haushaltslasten wurden drei unterschiedliche Profile ausgewählt, welche am ehesten dem H0-Profil entsprechen. Zusätzlich wurde jeweils ein Haushaltsprofil mit der Ähnlichkeit zum L1- und G6-Profil ausgewählt. Für Gewerbe- und Landwirtschaftslasten wurde ein G0-Profil gewählt und für die Mittelspannungsebene zusätzlich noch jeweils eines mit der Ähnlichkeit zum G0-, G3 und G4-Profil.

2.2. Statistische Auswertung der Belastungsszenarien

Zur Beschreibung der Unsicherheiten von Belastungsszenarien wird eine statistische Auswertung durchgeführt und die Ergebnisse anschließend mit Hilfe von stochastischen Kenngrößen beschrieben. Es werden einerseits grundlegende stochastische Parameter zur Beschreibung der Wahrscheinlichkeitsverteilung verwendet und andererseits Bewertungsmethoden herangezogen, welche die Modelle validieren und deren Aussagekraft bewerten. Zum bestmöglichen Verständnis werden die wichtigsten verwendeten stochastischen Kenngrößen im Detail erläutert.

2.2.1. Stochastische Parameter

Um eine Menge an Datenpunkten (z.B. Auslastungswerte einer Transformatorstation) zu beschreiben, kann diese mit Hilfe einer **Wahrscheinlichkeitsverteilung** erklärt werden. Diese Verteilung gibt an, welche Werte bzw. Wertebereiche der Datenmenge häufiger vorkommen, sprich eine höhere Wahrscheinlichkeit besitzen und welche Werte bzw. Wertebereiche weniger vorhanden sind und eine niedrigere Wahrscheinlichkeit besitzen. Am bekanntesten ist die Wahrscheinlichkeitsverteilung einer Normalverteilung, wessen Dichtefunktion die Form einer Gaußschen Glockenkurve besitzt. Die Normalverteilung beschreibt Zufallsvariablen wie beispielsweise die Streuung von Messwerten oder Stichproben. Abbildung 10 zeigt dazu die Dichtefunktion $f(x)$ der Normalverteilung und die dazugehörige kumulierte Dichtefunktion, welche als Verteilungsfunktion $F(x)$ bezeichnet wird. In diesem Zusammenhang ist auch der mit der Normalverteilung verbundene zentrale Grenzwertsatz zu erwähnen. Dieser besagt, dass die resultierende Gesamtverteilung, bestehend aus einer Menge an unterschiedlich verteilten Zufallsvariablen, eine Normalverteilung ergibt [23].

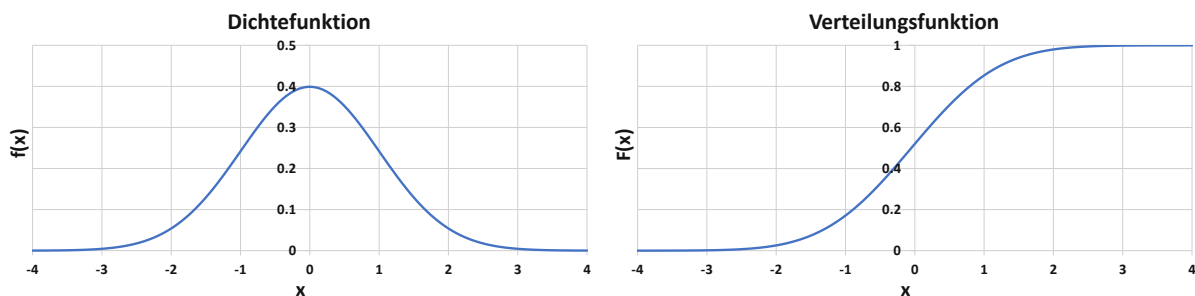


Abbildung 10: Dichte- und Verteilungsfunktion einer Normalverteilung

Zur Beschreibung der Wahrscheinlichkeitsverteilung stehen die Parameter **Erwartungswert** μ und die **Standardabweichung** σ zur Verfügung. Der Erwartungswert sagt aus, welchen Wert eine Zufallsvariable annimmt, wenn man aus einer unendlichen Menge an Variablen zieht. Oft wird auch der Begriff Mittelwert verwendet, welcher aber nicht mit dem arithmetischen Mittel zu verwechseln ist. Das arithmetische Mittel wird im Unterschied dazu aus einer definierten Anzahl an Variablen berechnet. Die Standardabweichung gibt an, wie weit die Variablen im Durchschnitt vom arithmetischen Mittel entfernt sind. Je niedriger der Wert, desto weniger streuen die Zufallsvariablen.

Je nach Verteilung einer Variable, kann diese besser oder schlechter durch eine Wahrscheinlichkeitsverteilung beschrieben werden. Um herauszufinden, welche Wahrscheinlichkeitsverteilung die jeweilige Menge an Variablen am besten beschreibt, kann beispielsweise mittels Maximum-Likelihood-Methode die Anpassung einer Dichtefunktion an die Ausgangsdaten geschätzt werden. Die passendste Dichtefunktion wird dann für die

statistische Beschreibung weiterverwendet. In dieser Arbeit waren für die Auswertung der Unsicherheiten von Belastungsszenarien die Dichtefunktionen der Verbraucher- und Erzeugereinheiten in einem Stromnetz notwendig. Nach ausgiebiger Recherche wurden folgende Funktionen verwendet:

- **Haushaltsprofile** sind am besten beschreibbar mittels **Log-Normalverteilung** [24]
- Die Erzeugung von **PV-Anlagen** folgt am besten einer **Beta-Verteilung** [25]
- Windgeschwindigkeiten der **Windkraftanlagen** werden mittels **Weibull-Verteilung** approximiert [25]

Eine weitere Darstellungsform der Wahrscheinlichkeitsverteilung von Lasten und Verbrauchern wäre die Jahresdauerlinie. Sie stellt die auftretenden Leistungswerte in sortierter Reihenfolge über ein Jahr gesehen dar. Mit ihrer Hilfe kann sehr einfach abgeschätzt werden, wie viele Lastspitzen über ein Jahr auftreten oder welche Grundlast eine Verbraucher- oder Erzeugereinheit besitzt.

2.2.2. Monte-Carlo-Simulation

Bei der Erzeugung von Zufallszahlen, sei es durch einfaches Würfeln, einen Münzwurf oder mit Hilfe eines digitalen Zufallsgenerators, besteht für alle möglichen Ereignisse dieselbe Wahrscheinlichkeit. Es handelt sich somit um eine gleichverteilte Zufallsvariable.

Möchte man hingegen Zufallszahlen erzeugen, welche einer bekannten Wahrscheinlichkeitsverteilungen unterliegen, so bietet die Monte-Carlo-Simulation eine einfache stochastische Methode. Ausgangslage stellt eine Verteilungsfunktion (kumulierte Dichtefunktion) dar, welche an der Ordinate die Wahrscheinlichkeit $F(x)$ zwischen 0–1 angibt. Es wird eine gleichverteilte Zufallszahl $F(x)$ erzeugt und über die Verteilungsfunktion die Zahl x ermittelt. Die Zahl x entspricht nun der zugrundeliegenden Verteilungsfunktion. Abbildung 11 zeigt den beschriebenen Vorgang anhand der Verteilungsfunktion einer Standardnormalverteilung ($\mu=0; \sigma=1$). Die gewürfelte gleichverteilte Zufallszahl 0,7 wurde mit Hilfe der Verteilungsfunktion in die normalverteilte Zahl 0,5 umgewandelt.

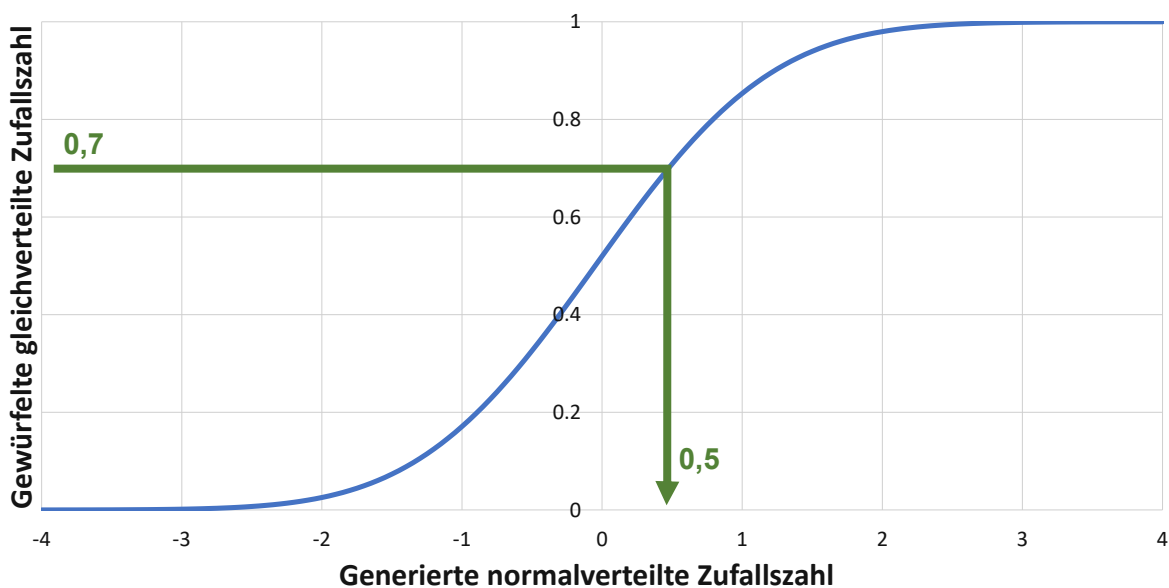


Abbildung 11: Beispiel Monte-Carlo-Simulation

2.2.3. Informationskriterium

Für die Auswahl eines geeigneten Modells zur Beschreibung der Unsicherheiten von Belastungsszenarien, werden das in der Statistik gebräuchliche Akaike- und Bayessche Informationskriterium (kurz AIC und BIC) verwendet. Beide Kriterien bewerten die Modelle anhand ihrer Anpassungsgüte und Komplexität, wobei die Anpassungsgüte als positive Bewertung und die Komplexität des Modells als negative Bewertung miteinfließt. Die Formeln für das AIC und BIC lauten dazu wie folgt:

$$AIC = 2 \cdot N \log L + 2 \cdot p \quad (1)$$

$$BIC = 2 \cdot N \log L + p \cdot \log(n) \quad (2)$$

Dabei entspricht $N \log L$ der negativen Likelihood-Funktion, p der Anzahl an geschätzten Parametern und n der Anzahl an Beobachtungspunkten. Das Modell mit dem niedrigeren AIC bzw. BIC ist jeweils zu bevorzugen. Die Kriterien haben nur eine Aussagekraft im Vergleich der Modelle untereinander und dürfen nicht als absolutes Maß hergenommen werden. [26]

2.3. Maschinelles Lernen

Der Begriff künstliche Intelligenz ist seit einigen Jahren immer öfters in Bereiche des öffentlichen Lebens vorgedrungen und stellt für viele den Inbegriff von moderner Technik dar. Allein die Nutzung und Akzeptanz von Sprachassistenten, Gesichtserkennung, automatischen Suchvorschlägen und vieles mehr zeigt, wie sehr dieses Thema im Alltag angekommen ist. In diesen Bereich fällt auch das sogenannte maschinelle Lernen, wessen Begrifflichkeit nicht so bekannt ist, aber auch genau die Eigenschaften von einer künstlichen Intelligenz besitzt. Beim maschinellen Lernen wird ein System bzw. Algorithmus mit Daten trainiert, anhand dessen es Schlussfolgerungen, Zusammenhänge, Ähnlichkeiten etc. erkennt und dadurch auf neue und unbekannte Daten dieses Wissen anwenden kann. Beispielsweise erkennt eine Gesichtserkennungssoftware einzelne Gesichtsmarkmalen wieder und folgert daraus, dass es sich um eine bereits bekannte, besser gesagt gelernte Person handeln muss. Die Software hat also selbstständig Eigenschaften definiert und eine Logik aufgebaut, anhand derer sie ein neues Gesicht analysiert und ihre Entscheidungen trifft.

Maschinelles Lernen wird üblicherweise in zwei Teilbereiche unterteilt: überwachtes und unüberwachtes Lernen. Beim **überwachten Lernen** bekommt der Algorithmus Lerndaten und zusätzlich eine definierte Einteilung der Daten übergeben. Der Algorithmus baut sich anschließend selbst eine dahinterliegende Logik. Es muss jedoch keine eigenständige Klassifikation durchgeführt werden. Neue Daten können dann anhand der gebauten Logik in die vorgegebenen Klassen geordnet werden. Bekannte Anwendungsgebiete sind beispielsweise Regressionsanalysen oder Klassifikationsverfahren. Dem gegenüber steht das **unüberwachte Lernen**, bei dem ein angewendeter Algorithmus selbstständig, also ohne vorgegebene Klassifikationen, innerhalb der Daten Muster erkennt und damit gewisse Datenbereiche voneinander unterscheiden kann. Es können dadurch unbekannte Strukturen in den Daten gefunden und weitere Kenntnisse daraus gewonnen werden. Am bekanntesten ist in diesem Feld die Segmentierung (Clustering) von Daten. In der vorliegenden Arbeit wurde mit Hilfe von Clusteralgorithmen gearbeitet, weshalb dieser Bereich des maschinellen Lernens detaillierter beschrieben wird.

Clusteranalyse

Die Clusteranalyse stellt einen Bereich des maschinellen unüberwachten Lernens dar und zielt darauf ab, Ähnlichkeiten und Strukturen in einer gegebenen Menge an Daten zu finden. Die Clusteranalyse kann dabei in ein weiches und hartes Clustering unterteilt werden. Beim weichen Clustering werden die Daten mittels Gewichtungen unterschiedlichen Clustern zugeordnet. Ein Datenpunkt kann somit mehreren Clustern angehören, zu denen er eine unterschiedlich starke Zugehörigkeit besitzt. Beim harten Clustering werden die Daten stets nur einem Cluster zugewiesen. Dadurch kommt es zu einer eindeutigen Zuordnung zwischen jedem Datenpunkt und seinem Cluster. Nachfolgend soll auf einige gängige Clusteralgorithmen eingegangen werden, welche auch in dieser Arbeit verwendet wurden.

k-Means-Clustering

Der k-Means-Algorithmus ist ein sehr einfacher Clusteralgorithmus, bei dem eine Anzahl an Clustern in dem Datensatz vorgegeben werden muss. Anschließend werden zufällige Clusterzentren gesetzt und jeder Datenpunkt dem nächstgelegenen Clusterzentrum zugeordnet. Anschließend wird aus den gruppierten Datenpunkten ein neues Clusterzentrum berechnet und die Zuordnung der Datenpunkte wiederholt. Sobald sich das neu berechnete Clusterzentrum nicht mehr stark von dem aktuellen Clusterzentrum unterscheidet (Abbruchkriterium), wird der Vorgang beendet. Die zu minimierende Zielfunktion lautet dabei wie folgt:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} \text{dist}(c_i, x)^2 \quad (3)$$

Dabei entspricht k die Anzahl der Cluster, C_i dem i -ten Cluster, c_i dem Zentrum des i -ten Clusters und x dem jeweiligen Datenpunkt. Es muss Formel (3) zufolge die Summe der quadrierten Distanzen SSE minimiert werden. Neben der quadratischen euklidischen Distanz werden auch andere Metriken für die Distanzberechnung angewendet, wie beispielsweise die Manhattan- oder Mahalanobis-Distanz.

Ein wichtiger Punkt für eine erfolgreiche Zuordnung der Datenpunkte ist die Wahl der Start- bzw. Anfangszentren, da es durch unterschiedliche Startpunkte zu unterschiedlichen Ergebnissen kommen kann. Es bietet sich daher an, den Algorithmus mehrmals mit zufälligen Startzentren zu starten und das Ergebnis mit dem niedrigsten SSE zu wählen. Schwachpunkte dieses sehr einfach gehaltenen Algorithmus sind einerseits das Verhalten bei Rauschen, wodurch die resultierenden Cluster unnatürlich „aufgeblasen“ werden und andererseits Cluster ohne eine kugelförmige Struktur, weil diese aufgrund der Zielfunktion nicht ausreichend erkannt werden. [27]

Hierarchisches Clustering

Beim (agglomerierenden) hierarchischen Clustering wird zunächst jeder Datenpunkt allein einem Cluster zugewiesen und sukzessive zwei benachbarte Cluster zusammengeführt, bis alle Datenpunkte in einem Cluster enthalten sind. Das Resultat ist ein sogenanntes Dendrogramm, woran man den Vorgang des Zusammenführens gut nachverfolgen kann. Pro Iteration werden immer die zwei nächstgelegenen Cluster zusammengeführt und anschließend die neuen Distanzen zwischen den Clustern berechnet. Aus Abbildung 12 kann

man beispielsweise entnehmen, dass die Datenpunkte 14 und 24 als erstes zusammengeführt wurden, weil ihre Distanz (Höhe im Dendrogramm) am geringsten ist.

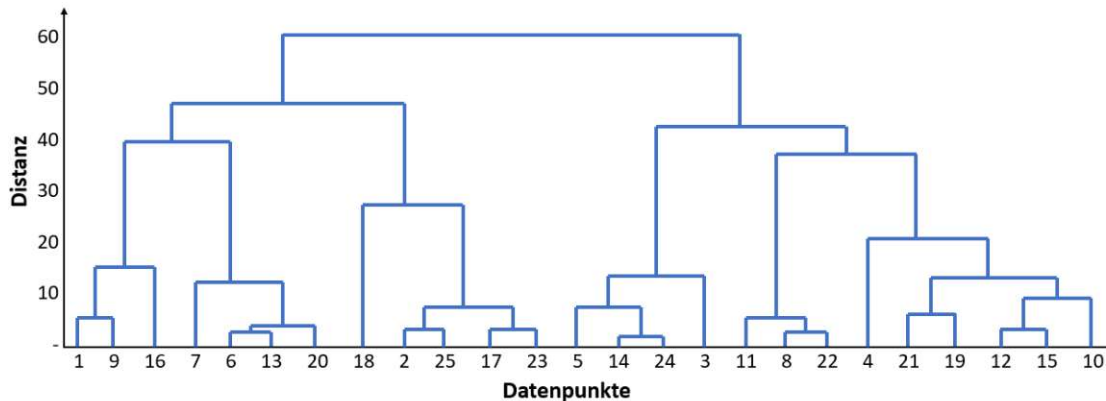


Abbildung 12: Dendrogramm beim hierarchischen Clustering

Die Clusteranzahl muss beim hierarchischen Clustering ebenfalls vorgegeben werden. Dafür können unterschiedliche Kriterien definiert werden. Entweder wird eine fixe Anzahl vorgegeben oder es wird eine Distanz zwischen den Clustern definiert, unter dieser zwei Cluster nicht kombiniert werden dürfen und damit der Algorithmus selbständig stoppt. Die Abstandsberechnung zwischen den Clustern kann unterschiedlich definiert werden. Dies ist nicht zu verwechseln mit der ausgewählten Metrik. Die Abstandsberechnung bezieht sich darauf, welcher Punkt oder welche Punkte von einem Cluster für die Berechnung herangezogen werden. Gängige Abstandsmaße sind beispielsweise die kürzeste Distanz (die zwei am nächstgelegenen Datenpunkte), die maximale Distanz (die zwei am weitesten entfernten Datenpunkte) oder ein durchschnittlicher Abstand zwischen zwei Clustern. [27]

Gaußsches Mischmodell mittels Erwartungs-Maximierungs-Algorithmus

Die beiden oben genannten Algorithmen arbeiten beide mit der Minimierung der Distanzen innerhalb der Cluster. Enthält ein Datensatz keine einfachen kompakten Clusterstrukturen, sondern setzt sich aus dichteren Bereichen und Bereichen mit weniger zusammenhängenden Datenpunkten (z.B. Rauschen) zusammen, so stoßen diese Algorithmen oft an ihre Grenzen. Ein Algorithmus, der einen Datensatz durch mehrere überlagerte Dichtefunktion beschreiben kann, kann mit solch unterschiedlichen Strukturen im Datensatz besser umgehen. Ein Spezialfall dieses Algorithmus mittels Mischverteilungen stellt hier die Variante mit Gaußschen Normalverteilungen dar.

Beim Gaußschen Mischmodell stellt jede Normalverteilung einen Cluster dar und beinhaltet die Datenpunkte aus dieser Verteilung. Die Parametrierung der Verteilungsfunktionen erfolgt mittels Erwartungs-Maximierungs-Algorithmus. Die Abfolge und mathematische Formulierung des Algorithmus lassen sich wie folgt beschreiben:

Die Normalverteilungen der einzelnen Cluster g lauten

$$f_G(x|\mu_g, \sigma_g) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_g}} \cdot e^{-\frac{1}{2} \left(\frac{x - \mu_g}{\sigma_g} \right)^2} \quad (4)$$

mit dem Mittelwert μ und der Standardabweichung σ des jeweiligen Clusters. Die Dichtefunktion einer Gaußschen Mischverteilung dazu lautet

$$f(x|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{g=1}^G \pi_i \cdot f(x|\mu_g, \sigma_g) \quad (5)$$

mit der Clusteranzahl G und der jeweiligen Clustergewichtung π . Der EM-Algorithmus versucht nun die drei unbekannt Parameter π , μ und σ in einem abwechselnden Erwartungs- und Maximierungs-Schritt zu approximieren. Dazu wird den Parametern zunächst ein Startwert $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\mu}^{(0)}$, $\boldsymbol{\sigma}^{(0)}$ zugeordnet. In jeder Iteration r wird nun die a-posteriori-Wahrscheinlichkeit p im Erwartungs-Schritt berechnet durch

$$p_{ig}^{(r)} = \frac{\pi_g^{(r-1)} \cdot f(x_i|\mu_g^{(r-1)}, \sigma_g^{(r-1)})}{f_G(x_i|\boldsymbol{\pi}^{(r-1)}, \boldsymbol{\mu}^{(r-1)}, \boldsymbol{\sigma}^{(r-1)})} \quad (6)$$

und anschließend die unbekannt Parameter im Maximierungs-Schritt ermittelt durch

$$\hat{\pi}_g^{(r)} = \frac{1}{n} \cdot \sum_{i=1}^n p_{ig}^{(r)} \quad (7)$$

$$\hat{\mu}_g^{(r)} = \frac{p_{ig}^{(r)}}{\sum_{j=1}^n p_{jg}^{(r)}} \cdot x_i \quad (8)$$

$$\hat{\sigma}_g^{(r)} = \sqrt{\sum_{i=1}^n \frac{p_{ig}^{(r)}}{\sum_{j=1}^n p_{jg}^{(r)}} \cdot x_i \cdot (x_i - \mu_g^{(r)})^2} \quad (9)$$

Zuletzt wird geprüft, ob die Lösung konvergiert und im Konvergenzfall der Iterationsprozess abgebrochen. [28]

3. Methodik

In diesem Kapitel wird die methodische Herangehensweise an die Forschungsfragen vorgestellt und die durchgeführten Arbeitspakete näher erläutert. Die Methodik beschreibt nicht nur den Weg zur Beantwortung der Forschungsfragen, sie begründet auch die ausgewählten Verfahren, Daten, Berechnungen und Annahmen und stellt diese im Detail dar. Dadurch kann eine möglichst genaue Reproduzierbarkeit dieser Arbeit gewährleistet werden und die erlangten Ergebnisse und Erkenntnisse für verwandte Forschungsfragen in späteren Arbeiten dienlich werden. Abbildung 13 zeigt das Flussdiagramm der durchgeführten Forschung, welches sich in 8 zentrale Punkte unterteilen lässt. Im Nachfolgenden wird auf jeden dieser Punkte im Detail eingegangen.

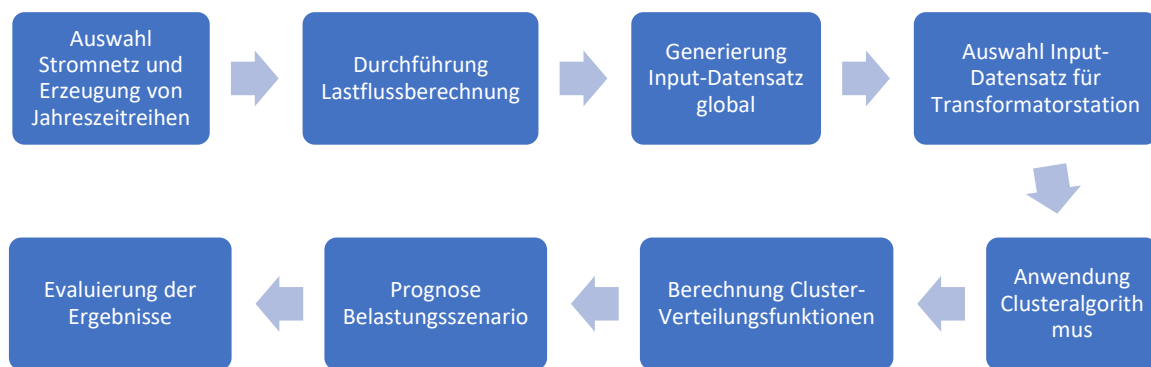


Abbildung 13: Flussdiagramm Methodik

Zunächst wurde ein Stromnetz ausgewählt, welches für die Analyse am geeignetsten erschien und zusätzliche Jahreszeitreihen erzeugt (1). Als nächstes wurde mit diesem Stromnetz eine Quasi-Dynamische Lastflussrechnung durchgeführt, welche die Auslastungswerte aller Ortsnetzstationen lieferte (2). Anschließend wurden diese Auslastungswerte mit Zeit- und Wetter-Variablen kombiniert, wodurch ein (globaler) Input-Datensatz erzeugt wurde (3). Da die Prognose von Belastungsszenarien für eine spezifische Ortsnetzstation in dem Stromnetz erfolgt, wurden nach der Auswahl einer Station die entsprechenden Daten aus dem globalen Input-Datensatz gefiltert und der lokale Input-Datensatz erzeugt (4). Dieser Datensatz diente als Eingangsgröße für die Clusteranalyse, bei der mittels Clusteralgorithmus die Daten in mehrere Clustergruppen unterteilt wurden (5). Bei der anschließenden Prognose der Belastungsszenarien wurden die Daten aus den Clustern verwendet, welche am besten dem jeweiligen Belastungsszenario entsprechen und daraus eine Prognose der Unsicherheiten (Dichtefunktion) erstellt (6+7). Die abschließende Evaluierung diente der Kontrolle der Ergebnisse und der Abschätzung zur Anwendbarkeit dieser Methode (8).

3.1. Auswahl des untersuchten Stromnetzes

Für die Auswahl eines zu untersuchenden elektrischen Stromnetzes sollte ein frei verfügbarer Datensatz gewählt werden, welcher in ausreichend hohem Detaillierungsgrad verfügbar stand. Weiters wurden rein fiktive Datensätze nicht in Betracht gezogen, da diese für die Implementierung des Clusteralgorithmus ungeeignet erschienen. Wie schon im Punkt 2.1.6 erläutert, vereint der SimBench-Datensatz viele dieser Eigenschaften, welcher im Rahmen eines Forschungsprojektes zwischen Universitäten, Forschungsinstituten und Verteilernetzbetreibern zwischen 2015 und 2019 erarbeitet wurde.

Das verwendete Netzmodell aus dem SimBench-Datensatz ist das Mittelspannungs-Niederspannungsnetz mit der Bezeichnung *1-MVLV-semiurb-all-0-no_sw* und dessen Ausbaustufe bzw. Szenario 2 mit der Bezeichnung *1-MVLV-semiurb-all-2-no_sw*. Dabei bezeichnet „MVLV“ die verwendeten Spannungsebenen, „semiurb“ den Urbanisierungsgrad des Mittelspannungsnetzes, „all“ die Anzahl der Niederspannungsnetze, „0“ bzw. „2“ das gewählte Szenario und „no_sw“, dass keine Netzschalter verwendet wurden. Mit anderen Worten bedeutet diese Bezeichnung, dass es sich um ein Mittelspannungs- (eng. MV - medium voltage) und Niederspannungsnetz (eng. LV - low voltage) handelt, bei dem das Mittelspannungsnetz einen vorstädtischen (eng. semiurban) Netzcharakter aufweist. Zudem wurden alle (für dieses Netz verfügbaren) Niederspannungsnetze verwendet und die beiden Szenarien Basisszenario-0 und Entwicklungsszenario-2 ausgewählt und keine Netzschalter verwendet (eng. no_sw – no switches). Die Entscheidung fiel auf dieses Netzmodell, da dieses vorstädtische Mittelspannungsnetz die unterschiedlichste Konstellation von angeschlossenen Niederspannungsnetzen aufwies und sich so sehr gut für die Validierung des Clusteralgorithmus eignete.

Das Mittelspannungsnetz wird dabei als offener 20kV-Ring mit Vermaschungen bzw. „Triple“ betrieben. Dieses Stromnetz mit seinen zwei Ausbaustufen (Basisszenario-0 und Entwicklungsszenario-2) soll einerseits den aktuellen Stand in einem Stromnetz darstellen und andererseits die Entwicklungen durch den Ausbau an erneuerbaren Energien und die wachsende E-Mobilität aufzeigen. Ausgewählte Parameter des Mittelspannungsnetzes sind aus Tabelle 1 zu entnehmen.

Tabelle 1: Netzbeschreibende Parameter des Mittelspannungsnetzes

	Vorstädtisch Basisszenario-0	Vorstädtisch Entwicklungsszenario-2
<i>Anzahl Leitungen</i>	121	126
<i>Summe Leitungslängen</i>	67,28 km	74,14 km
<i>Extrema Leitungslängen</i>	0,15-1,70 km	0,15-2,34 km
<i>Verkabelungsgrad</i>	69 %	72 %
<i>Anzahl Lasten im MV-Netz</i>	5	12
<i>Anschlussleistung Lasten</i>	1,12 MW	1,80 MW
<i>Extrema Lasten</i>	0,09-0,34 MW	0,01-0,51 MW
<i>Typen von Lasten</i>	Gewerbe, Landwirtschaft, Bandlast	Gewerbe, Landwirtschaft, Bandlast, Luft-WP, Erd-WP, E-Fahrzeuge
<i>Anzahl Erzeuger im MV-Netz</i>	11	17
<i>Anschlussleistung Erzeuger</i>	12,30 MW	52,68 MW
<i>Extrema Erzeuger</i>	0,08-2,10 MW	0,08-14,00 MW
<i>Typen von Erzeugern</i>	PV, Wind, Biogas, Wasser	PV, Wind, Biogas, Wasser, Batterie
<i>Angeschlossene Niederspannungsnetze</i>	110	

Wie aus Tabelle 1 zu entnehmen ist, besitzt das Mittelspannungsnetz im Entwicklungsszenario-2 im Vergleich zum Basisszenario-0 mehrere Bereiche, in denen Ausbauten getätigt werden. Es werden zusätzliche Leitungen in sogenannte Vorrangflächen gelegt, um die zusätzliche Erzeugung durch Wind und PV ins Netz integrieren zu können. Auch neue Lasten (z.B. Erd-Wärmepumpen, E-Fahrzeuge) werden im Mittelspannungsnetz

angeschlossen. Vor allem aber ist der Ausbau der erneuerbaren Energien zu erkennen, bei denen die Einspeiseleistung um mehr als das Vierfache in Bezug auf das Basisszenario vergrößert wird.

Die an dem Mittelspannungsnetz angeschlossenen Niederspannungsnetze besitzen unterschiedliche Netztypen (siehe auch Punkt 2.1.6). Diese NS-Netze werden als Strahlennetze betrieben und setzen sich aus den folgenden Netztypen zusammen:

- LV1 ländlich 6x
- LV2 ländlich 25x
- LV3 ländlich 18x
- LV4 vorstädtisch 36x
- LV5 vorstädtisch 23x
- LV6 städtisch 2x

Es existieren demnach in beiden Szenarien insgesamt 110 Niederspannungsnetze, welche idente Leitungsparameter aufweisen, sich jedoch stark in der Größe der angeschlossenen Erzeuger und Verbraucher unterscheiden.

Tabelle 2: Netzbeschreibende Parameter der Niederspannungsnetze

	Vorstädtisch Basisszenario-0	Vorstädtisch Entwicklungsszenario-2
<i>Anzahl Leitungen</i>	8872	
<i>Summe Leitungslängen</i>	152,54 km	
<i>Extrema Leitungslängen</i>	0,19-139,53 m	
<i>Verkabelungsgrad</i>	100 %	
<i>Anzahl Lasten in den LV-Netzen</i>	8.767	11.197
<i>Anschlussleistung Lasten</i>	30,53 MW	43,31 MW
<i>Extrema Lasten</i>	1,00-31,00 kW	1,00-49,40 kW
<i>Typen von Lasten</i>	Haushalt, Gewerbe, Landwirtschaft	Haushalt, Gewerbe, Landwirtschaft, Luft-WP, Erd-WP, E-Fahrzeuge
<i>Anzahl Erzeuger in den LV-Netzen</i>	783	2.415
<i>Anschlussleistung Erzeuger</i>	11,53 MW	43,12 MW
<i>Extrema Erzeuger</i>	1,80-78,38 kW	1,40-128,80 kW
<i>Typen von Erzeugern</i>	PV	PV, Batterie

Aus Tabelle 2 ist zu entnehmen, dass im Basisszenario-0 in allen Niederspannungsnetzen insgesamt 8.767 Lasten und 783 Erzeuger angeschlossen sind und diese Zahl im Entwicklungsszenario-2 auf 11.197 Lasten und 2.415 Erzeuger anwächst. Abbildung 14 zeigt das Einliniendiagramm des Stromnetzes im Basisszenario-0, nachdem es in die Simulationssoftware PowerFactory importiert wurde. Jede Farbe entspricht einem Niederspannungsnetz, welches von einem der 110 Ortsnetzstationen gespeist wird. Die etwas untypische Anordnung der Netzkomponenten in dem Einliniendiagramm (normalerweise wird ein Einliniendiagramm hierarchisch angeordnet; siehe auch Abbildung 8) ist den hinterlegten geographischen Koordinaten geschuldet, weshalb sich in der Mitte liegend das Umspannwerk zwischen Hochspannungs- und Mittelspannungsebene befindet. Von dort aus sind die

Ortsnetzstationen entlang des offen betriebenen Rings angeschlossen (vergleiche mit Abbildung 7):

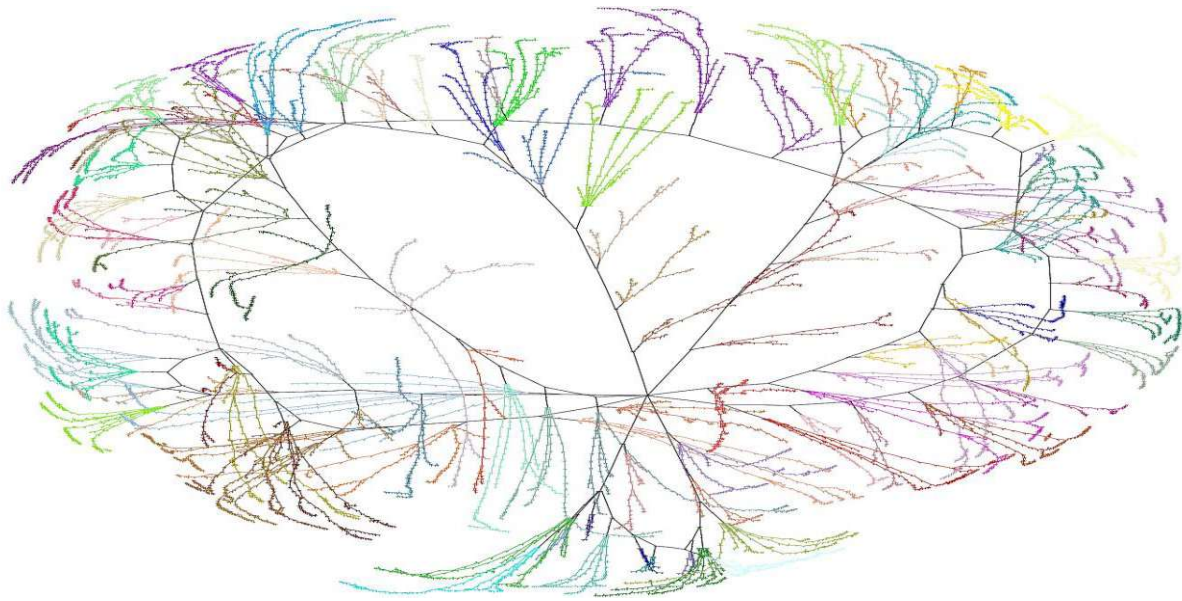


Abbildung 14: Netztopologie vom verwendeten Netzmodell im Basisszenario-0

Nach der Auswahl des Stromnetzes mussten auch die Jahreszeitreihen, welche den Erzeugungs- und Verbrauchereinheiten hinterlegt sind, untersucht werden, da der SimBench-Datensatz nur Jahreszeitreihen für ein Jahr beinhaltet. Aufgrund dessen wurden zusätzliche statistische Jahreszeitreihen generiert, um bei der Clusteranalyse und Prognose von Belastungsszenarien zusätzliche Lern- und Test-Daten zur Verfügung zu haben.

Generierung weiterer Jahreszeitreihen basierend auf dem SimBench-Datensatz

Wie schon in Punkt 2.1.6 erläutert wurde, enthält der SimBench-Datensatz für die unterschiedlichen Verbrauchertypen (Haushalt, Gewerbe, etc.) und Erzeugungseinheiten (PV, Wind, etc.) mehrere Jahreszeitreihen, wie beispielsweise 5 Profile für Haushaltslasten oder 8 Profile für Photovoltaikanlagen. In einem Stromnetz werden demnach jedem Verbraucher und Erzeuger eines dieser Profile zugeordnet und kann dadurch eindeutig bestimmt werden. Da der SimBench-Datensatz Zeitreihen für jeweils ein Jahr (nämlich das Jahr 2016) enthält, können Lastflussberechnungen oder Clusteranalysen nur für dieses eine Jahr durchgeführt werden.

Um dennoch für die vorliegende Forschungsarbeit mehr als ein Jahr in der Auswertung berücksichtigen zu können, wurden zusätzliche Jahreszeitreihen mit vorangegangener statistischer Auswertung erzeugt. Für die Parametrierung der Verteilungsfunktion der verschiedenen Lastprofile, wurde ein Ansatz mittels Monte-Carlo Methode verwendet, welcher in [24] beschrieben wurde. Die angewendete Methodik soll hier im Detail erläutert werden.

Schritt 1: Jahreszeitreihen im verwendeten SimBench-Datensatz identifizieren

Nach der Auswahl des SimBench-Datensatzes, mussten zunächst die verfügbaren Jahreszeitreihen identifiziert werden. In dieser Arbeit wurde der SimBench-Datensatz im CSV-Datenformat heruntergeladen, wodurch für die Lastzeitreihen und die Erzeugerzeitreihen

jeweils eine eigene Datenbank bzw. eigene Datei zur Verfügung gestellt wurde. Die Erzeugerzeitreihen beschränken sich dabei auf Zeitreihen für erneuerbare Energiequellen bzw. Batteriespeicher und enthalten keine Zeitreihen für konventionelle Erzeugungseinheiten oder übergeordnete Netze, da diese über die Residuallast ermittelt werden. Die Lastzeitreihen beinhalten die drei typischen Kategorien von Verbrauchertypen: Haushalt, Gewerbe und Landwirtschaft und zusätzlich die Zeitreihen für Wärmepumpen und E-Fahrzeuge. Für die Erzeugung eines weiteren Datensatzes von Jahreszeitreihen wurden in dieser Arbeit nur Zeitreihen für Lasten erzeugt, da diese mit statistischen Funktionen beschrieben und parametrisiert werden können. Die Zeitreihen für Erzeuger können ebenso statistisch beschrieben werden, hängen jedoch sehr stark mit realen Wetterdaten zusammen bzw. wurden selbst daraus generiert. Dieser Zusammenhang wurde bei der späteren Cluster-Analyse nützlich gemacht und sollte daher nicht verloren gehen.

Für das verwendete Netzmodell *1-MVLV-semiurb-all-0-no_sw* wurden somit 20 Jahreszeitreihen identifiziert. Für das Entwicklungsszenario-2 existieren noch 4 weitere Zeitreihen für E-Fahrzeuge und Wärmepumpen. Diese Lastprofile wurden bei der Generierung zusätzlicher Jahreszeitreihen ausgenommen und nur die 20 Jahreszeitreihen aus Tabelle 3, welche in beiden Ausbaustufen vorhanden sind, verwendet. Dabei steht BL für eine Bandlast, H für ein Haushaltsprofil, G für ein Gewerbeprofil und L für ein landwirtschaftliches Profil. Der abschließende Buchstabe identifiziert die Größe des Stromverbrauchs (siehe auch Abbildung 9).

Tabelle 3: Identifizierte Lastzeitreihen im verwendeten SimBench-Datensatz

BL-H	G2-A	G4-B	H0-A	H0-L
G1-A	G3-A	G4-M	H0-B	L1-A
G1-B	G3-M	G5-A	H0-C	L2-A
G1-C	G4-A	G6-A	H0-G	L2-M

Schritt 2: Berechnung des Leistungsfaktors

Jede Lastzeitreihe besteht aus einem Wertepaar (Real- und Imaginärteil) der Scheinleistung (S), also aus einem Wirk- (P) und Blindanteil (Q). Bei der Erzeugung von neuen Lastzeitreihen wurde nur die Wirkleistung P neu erzeugt und über den Leistungsfaktor $\cos(\varphi)$ die Blindleistung nachgezogen. Der Leistungsfaktor entspricht dabei dem ursprünglichen Wert aus den Ausgangs-Lastzeitreihen zu der jeweiligen Viertelstunde. Es wurde also für jeden 15min-Wert der Leistungsfaktor mit folgender Formel berechnet und abgespeichert:

$$S = \sqrt{P^2 + Q^2} \quad (10)$$

$$\cos(\varphi) = \frac{P}{S} \quad (11)$$

Schritt 3: Unterteilung in 3x3 Tagestypen

Jede Lastzeitreihe beinhaltet 35.136 Einzelwerte (366 Tage * 24 Stunden * 4 Viertelstunden), welche in eine der 9 Tagestypen eingeteilt wurden. Die Unterteilung in die Jahreszeiten Winter, Übergang und Sommer und in die drei Tage Werktag, Samstag und Sonn-/Feiertag entspricht hierbei derselben Einteilung, wie sie bei den VDEW-Standardlastprofilen verwendet wird. Jeder Einzelwert kann so eindeutig in eine der 9 Tagestyp-Kategorien eingeteilt werden. Aus Tabelle 4 kann entnommen werden, wie groß der Anteil der eingeteilten Daten in einer der neun Tagestyp-Kategorien ist. Beispielsweise fallen 19% der Tage im Jahr in die Kategorie *Werktag-Übergang*.

Tabelle 4: Einteilung Tagestypen

	Winter 01.11. – 20.03.	Übergang 21.03. – 14.05. und 15.09. – 31.10.	Sommer 15.05. – 14.09.
Werktag	27 %	19 %	24 %
Samstag	5 %	4 %	5 %
Sonn- & Feiertag	6 %	5 %	5 %

Schritt 4: Parametrierung der Verteilungsfunktion

Innerhalb jeder der neun Tagestyp-Kategorien standen nun mehrere Tageszeitreihen in 15min-Auflösung zur Verfügung. Beispielsweise beinhaltet die Kategorie *Werktag-Übergang* 19% der Tage im Jahr (Tabelle 4), was 70 Tage zu je 96 Stück 15min-Werte entspricht. Für jeden dieser 15min-Wert innerhalb der Tagestyp-Kategorie wurde die passende Verteilungsfunktion mit dessen Parametern Mittelwert μ und Standardabweichung σ ermittelt. Als passende Verteilungsfunktion wurde die logarithmische Normalverteilung angewendet (siehe auch Punkt 2.2.1). Die Parameter Mittelwert und quadratische Standardabweichung wurden wie folgt berechnet:

$$\mu = \frac{1}{n} \cdot \sum_{i=1}^n \ln(X_i) \quad (12)$$

$$\sigma^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (\ln(X_i) - \mu)^2 \quad (13)$$

Dabei entspricht n der Anzahl der 15min-Werte und X_i dem i -ten 15min-Wert. Es ergeben sich dadurch in jeder der neun Tagestyp-Kategorien jeweils 96 dieser Parameter-Paare für jeden 15min-Wert.

Schritt 5: Erzeugung einer neuen Jahreszeitreihe der Wirkleistung mittels Monte-Carlo Methode

Als nächstes wurden die zuvor statistisch ausgewerteten 15min-Werte herangezogen und damit mehrere neue Jahreszeitreihen der Wirkleistung erzeugt. Für jeden dieser 15min-Werte existierte das Wertepaar aus Mittelwert und Standardabweichung mit denen die dazugehörige kumulierte Dichtefunktion bzw. Verteilungsfunktion erzeugt wurde. Mittels Monte-Carlo-Methode wurde anschließend eine Zufallszahl zwischen 0 und 1 gewürfelt und diese Zahl über die Verteilungsfunktion in den dazugehörigen Leistungswert umgewandelt. Dieses Verfahren wurde für alle 35.136 Einzelwerte wiederholt, bis eine komplette Jahreszeitreihe der Wirkleistung vorhanden war.

Schritt 6: Erzeugung einer neuen Jahreszeitreihe der Blindleistung

Die zuvor erzeugten Jahreszeitreihen der Wirkleistung wurden nun zusammen mit den abgespeicherten Leistungsfaktoren aus Schritt 2 verwendet, um den jeweiligen Wert der Blindleistung zu jeder Viertelstunde zu berechnen. Die Formeln dazu lauten:

$$S = \frac{P}{\cos(\varphi)} \quad (14)$$

$$Q = \sqrt{S^2 - P^2} \quad (15)$$

Nach diesem Schritt war das Wertepaar der Scheinleistung komplett.

Schritt 7: Analyse der erzeugten Jahreszeitreihen

Zuletzt wurden die erzeugten Jahreszeitreihen analysiert, um sicherzustellen, dass diese noch den ursprünglichen Jahreszeitreihen entsprechen und nur statistische Schwankungen bzw. Prognoseunsicherheiten abbilden. Folgende Gegenüberstellungen wurden durchgeführt:

- Kumulierter Energieverbrauch über ein Jahr

Der kumulierte Energieverbrauch E über ein Jahr wurde mit Hilfe der 35.136 15min-Leistungswerte P_i und dem Zeitintervall t_i wie folgt berechnet:

$$E = \sum_{i=1}^{35136} P_i \cdot t_i \quad (16)$$

Der Jahresenergieverbrauch unterscheidet sich dabei zwischen den ursprünglichen und erzeugten Profilen entsprechend der nachstehenden Tabelle:

Tabelle 5: Lastprofilevergleich - Kumulierter Energieverbrauch

Profil	Differenz	Profil	Differenz	Profil	Differenz	Profil	Differenz
BL-H	+3,7%	G3-A	-0,1%	G5-A	+6,2%	H0-G	+0,3%
G1-A	+1,4%	G3-M	+0,6%	G6-A	+0,8%	H0-L	-2,0%
G1-B	+4,3%	G4-A	+0,7%	H0-A	+0,5%	L1-A	+0,5%
G1-C	+10,1%	G4-B	±0,0%	H0-B	-0,6%	L2-A	+0,1%
G2-A	±0,0%	G4-M	±0,0%	H0-C	+0,2%	L2-M	+0,1%

Der Großteil der in Tabelle 5 verglichenen Profile (16 von 20 Profilen) weisen eine Differenz innerhalb von $\pm 2\%$ auf, was eine hohe Güte bedeutet. Dieser Sachverhalt lässt sich auch gut mit Hilfe der Jahresdauerlinie veranschaulichen. Abbildung 15 zeigt die Jahresdauerlinie exemplarisch für das Lastprofil G6-A, welches das am häufigsten zugewiesene Gewerbelastprofil in dem untersuchten Stromnetz *1-MVLV-semiurb-all-0-no_sw* ist. Hierbei entspricht der blaue Graph dem ursprünglichen und der orange Graph dem erzeugten Profil. An der Ordinate können die 15min-Leistungswerte in Per-Unit abgelesen werden und an der Abszisse die Anzahl der 15min-Werte. Die Leistungswerte sind deshalb in Per-Unit, weil das Lastprofil erst mit der jeweiligen Nennleistung der Last skaliert wird und ab dann einen Leistungswert besitzt.

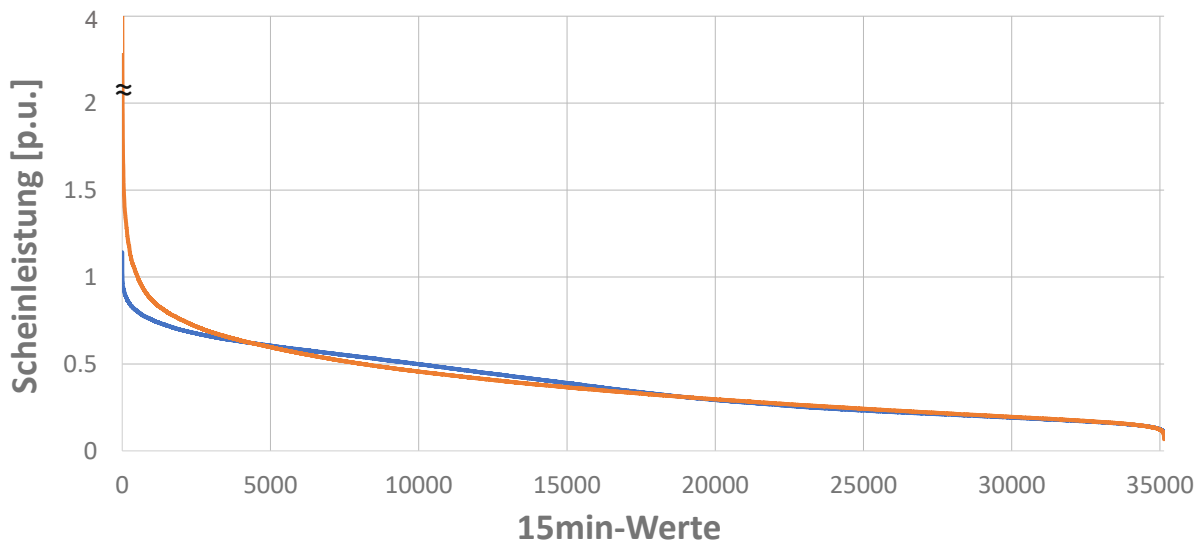


Abbildung 15: Lastprofilevergleich – Jahresdauerlinie

Die Jahresdauerlinie stellt die 15min-Leistungswerte der Größe nach in absteigender Reihenfolge dar. Man kann dementsprechend sehr einfach quantitativ feststellen, wie viele Spitzenlast-Zeiträume ein Verbraucher aufweist. Die Jahresdauerlinie des erzeugten Profils weist mehr und größere Lastspitzen als das ursprüngliche Profil auf. Diese zusätzlichen Lastspitzen stellen die statistischen Ausreißer dar, die durch die Generierung über die Verteilungsfunktion entstanden sind. Die Fläche unter den Graphen entspricht wieder der kumulierten Jahresenergiemenge. Diese kann als ident angesehen werden.

- Tagesverläufe

Auch ein Blick auf die jeweiligen Verläufe der ursprünglichen und generierten Lastprofile bietet einen brauchbaren Vergleich. Abbildung 16 zeigt exemplarisch den Verlauf des Lastprofils *H0-L*, welches das am häufigsten zugewiesene Profil in dem untersuchten Stromnetz *1-MVLV-semiurb-all-0-no_sw* ist. Hierbei entspricht wieder der blaue Graph dem ursprünglichen und der rote Graph dem erzeugten Profil. An der Ordinate können die 15min-Leistungswerte in Per-Unit abgelesen werden und an der Abszisse die Anzahl der 15min-Werte.

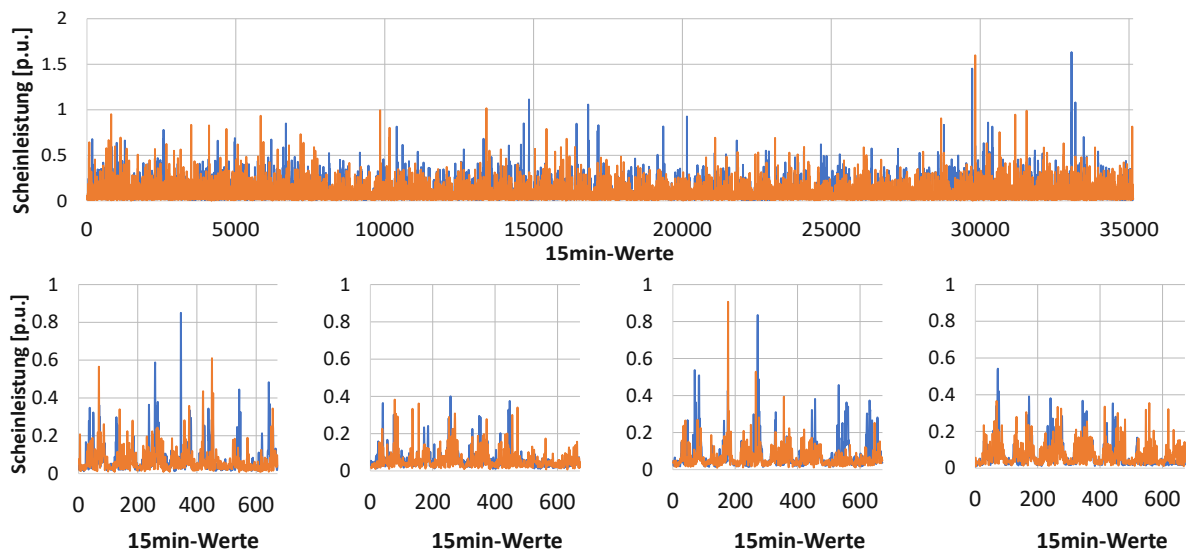


Abbildung 16: Lastprofilevergleich – Jahreslastgang und Wochenlastgänge

In Abbildung 16 zeigt das obere Diagramm den Jahreslastgang und die unteren Diagramme zeigen jeweils einen Wochenlastgang von exemplarisch ausgewählten Wochen:

- Frühling: Montag, 7.März - Sonntag, 13.März
- Sommer: Montag, 11.Juli - Sonntag, 17.Juli
- Herbst: Montag, 24.Oktober - Sonntag, 30.Oktober
- Winter: Montag, 19.Dezember - Sonntag, 25.Dezember

Die Verläufe der erzeugten Profile zeigen eine gute Übereinstimmung mit den ursprünglichen Profilen und weisen auch die gewünschten statistischen Schwankungen auf.

3.2. Durchführung der Lastflussberechnungen

Die Lastflussberechnungen wurden mit der Simulationssoftware DigSILENT PowerFactory durchgeführt. Dazu wurde der ausgewählte SimBench-Datensatz zunächst im CSV-Datenformat heruntergeladen, abgespeichert und anschließend über Python (SimBench2PowerFactory_Converter) in PowerFactory geladen. Die Berechnungen wurden dabei als symmetrischer Lastfluss (Mitsystem) und als eine Quasi-Dynamische Simulation im 1-Stunden-Intervall ausgeführt. Es ergaben sich so für den Simulationszeitraum 2016 (366 Tage * 24 Stunden) eine Berechnungsanzahl von 8784. Für die Quasi-Dynamische Simulation musste darauf geachtet werden, dass die hinterlegten Zeitcharakteristiken (Jahreszeitreihen) der Verbraucher und Erzeuger als „in Betrieb“ markiert wurden, da ansonsten nur ein Betriebspunkt in den Berechnungen berücksichtigt worden wäre. Weiters musste bei den Erzeugereinheiten die Wahl der Blindleistungsregelung vorgenommen werden. Hierbei konnte beispielsweise zwischen einer konstanten U, Q(P)-, Q(U)- oder cosphi(P)-Charakteristik gewählt werden. Für die Blindleistungsregelung der Erzeugereinheiten wurde eine Q(U)-Charakteristik ausgewählt, welche mit einem Totband von 0,97 p.u. bis 1,03 p.u. operierte. Dadurch wurde die Blindleistungserzeugung und der Blindleistungsverbrauch zur Spannungsstützung anhand des Spannungswertes am jeweiligen Anschlussort angepasst und dieser Wert innerhalb der gewählten Grenzen gehalten.

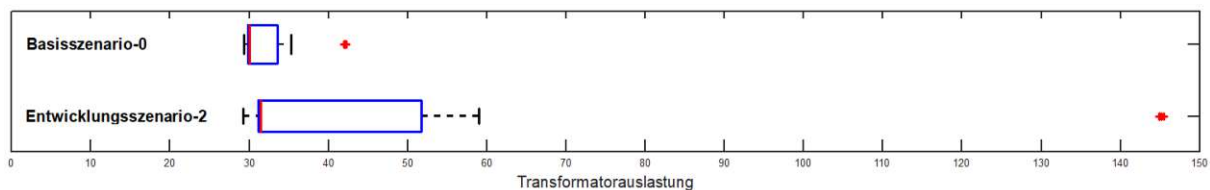


Abbildung 17: Boxplot der maximalen Transformatorauslastungen

Abbildung 17 zeigt eine erste Auswertung der Ergebnisse der Quasi-Dynamischen Lastflussberechnung. Die hier dargestellte jährliche Maximalauslastung aller Ortsnetzstationen weist im Basisszenario-0 einen Medianwert von etwas mehr als 30% auf. Einzelne Ausreißer besitzen eine maximale Auslastung von rund 42%. Im Entwicklungsszenario-2 kommt es zu noch größeren Auslastungen, wobei der Medianwert mit 31% nur knapp über dem im Basisszenario-0 liegt. Dafür gibt es Ausreißer bis hin zu 145% Transformatorauslastung. Diese erste Analyse bestätigt den Sachverhalt, dass Betriebsmittel wie Transformatoren über große Reservekapazitäten verfügen, welche oft ungenutzt bleiben.

Als Ergebnisse aus den durchgeführten Quasi-Dynamischen Simulationen wurden ausschließlich die Auslastungswerte der Transformatoren bzw. Ortsnetzstationen der

einzelnen Niederspannungsnetze benötigt. Andere Berechnungswerte wie beispielsweise Spannungswerte an den Sammelschienen, Energieflüsse entlang der Leitungen, der Verlauf der Netzfrequenz oder Auslastungswerte anderer Betriebsmittel wurden hier nicht weiter betrachtet. Die Auslastungswerte der Transformatoren stellen damit einen Teil des benötigten Input-Datensatzes für die Cluster-Analyse dar. Die restlichen Teile dieses Input-Datensatzes setzen sich aus den unterschiedlichen Zeit- und Wettervariablen zusammen.

3.3. Generierung eines Input-Datensatzes

Der Input-Datensatz besteht aus ausgewählten Eigenschaften, welche sich auf die Auslastung von Transformatoren auswirken (siehe auch Punkt 2.1.3) und die auch frei bzw. leicht verfügbar waren. Dieser Datensatz kann als Matrix verstanden werden und besteht dabei aus einzelnen mehrdimensionalen Vektoren, welche jeweils die ausgewählten Eigenschaften beinhalten. Eine Eigenschaft ist beispielsweise die zuvor berechnete Transformatorauslastung oder die Temperatur zum entsprechenden Zeitpunkt. Ein Vektor stellt mehrere Eigenschaften dar, die alle einen Zeitpunkt beschreiben. Abbildung 18 soll diesen Input-Datensatz bildlich veranschaulichen.

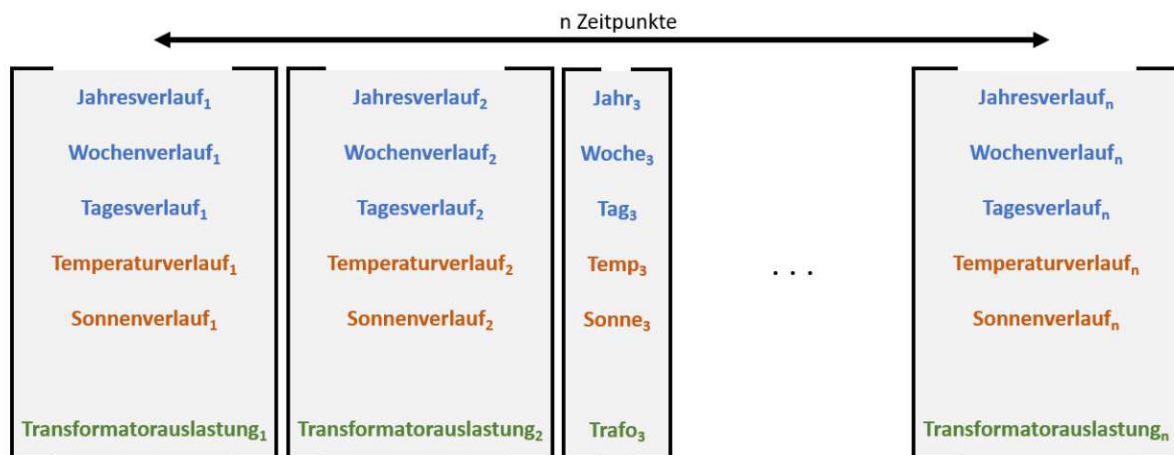


Abbildung 18: Matrix des Input-Datensatzes

Ein Vektor besteht folglich aus den drei Zeit-Variablen Jahresverlauf – Wochenverlauf – Tagesverlauf, aus den zwei Wetter-Variablen Temperaturverlauf – Sonnenverlauf und aus der Betrieb-Variable Transformatorauslastung. Die Größe des Input-Datensatzes bzw. die Anzahl n der Zeitpunkte beträgt hier 8784, was einem Jahr entspricht. Es wird somit ein einjähriger Input-Datensatz verwendet und nur im Punkt 4.1.2 ein mehrjähriger.

Der Input-Datensatz kann in globale und lokale Variablen unterteilt werden. Die globalen Variablen stellen die Zeit-Variablen dar, da diese an jeder der 110 Transformatorstationen ident sind. Lokale Variablen stellen Wetter- und Betrieb-Variablen dar, da diese transformatorspezifisch sind. Damit unterschiedliche Transformatorstationen untersucht werden können, wurde ein sogenannter kompletter bzw. globaler Input-Datensatz erzeugt, welcher die globalen Zeit-Variablen und alle lokalen Wetter- und Betrieb-Variablen beinhaltet. Aus diesem globalen Input-Datensatz wird, sobald eine Transformatorstation ausgewählt wurde, der lokale Input-Datensatz extrahiert, damit für die Clusteranalyse alle Variablen transformatorspezifisch sind. Im Nachfolgenden wird der globale Input-Datensatz vorgestellt, welcher für das ganze betrachtete Stromnetz zur Verfügung steht. Die anschließende Anpassung zum lokalen Input-Datensatz erfolgt dann im Unterkapitel 3.4.

Zeit-Variablen

Die drei Variablen mit Bezug auf die Zeit wurden durch den Zeitstempel ermittelt, welcher bei jeder der 8764 Lastflussberechnungen der Quasi-Dynamischen Simulation geschrieben wurde. Die Form des Zeitstempels ist „JJJJ-MM-TT hh:mm:ss“, wobei die Minuten- und Sekundenwerte irrelevant sind, da die Lastflussberechnungen im 1-Stunden-Intervall ausgeführt wurden. Weiters wurden für die Variablen Jahresverlauf und Wochenverlauf zwei Formen der Darstellung erzeugt, die als kontinuierliche und diskrete Variablen angesehen werden können.

Tabelle 6: Zeit-Variablen

	Variable	Wertebereich	Schrittgröße
Jahresverlauf	Jahrestag (kontinuierlich)	[1-366]	1
	Monat (diskret)	[1-12]	1
Wochenverlauf	Wochenstunde (kontinuierlich)	[1-168]	1
	Wochentag (diskret)	[1-7]	1
Tagesverlauf	Uhrzeit	[1-24]	1

Tabelle 6 zeigt die unterschiedlichen Darstellungsformen von Jahres-, Wochen- und Tagesverlauf, die dazugehörigen Wertebereiche und Schrittgrößen. Beispielsweise wurde der Zeitstempel „2016-07-13 15:00:00“ mit den Werten Jahrestag = 195, Monat = 7, Wochenstunde = 63, Wochentag = 3 und Uhrzeit = 15 ausgestattet. Die Schrittgröße 1 bei den Zeit-Variablen bedeutet, dass diese aus natürlichen Zahlen bestehen und es beispielsweise keinen Wert wie Wochentag = 1,5 gibt (Montag 0-23 Uhr entspricht dem Wochentag 1). Durch diese Zuordnung konnte jeder Vektor eindeutig einem Wert der Zeit-Variable zugeordnet werden.

Für das Clustering wurde pro Zeit-Variable entweder die kontinuierliche oder diskrete Form verwendet, da es sonst zu einer doppelten Beschreibung gekommen wäre. Die Anwendung der unterschiedlichen Darstellungsformen wird im Punkt 4.1.1 näher erörtert.

Wetter-Variablen

Die Variablen mit Bezug auf das Wetter wurden durch Daten vom Deutschen Wetterdienst DWD ermittelt. Diese frei zugängliche Klimadatenbank bietet Daten von über 600 Wetterstationen, welche im Bundesgebiet Deutschland verteilt sind und unterschiedliche Wetterphänomene messen. Für diese Arbeit wurden Wettergrößen verwendet, welche eine Auswirkung auf die Energienachfrage bzw. Energieproduktion im betrachteten Stromnetz haben und dadurch eine Information für die Cluster-Analyse liefern konnten. Es wurden dabei alle Wetterstationen berücksichtigt, die im gesamten Jahr 2012 (dieses Jahr diente im SimBench-Datensatzes als Wetterreferenz) in Betrieb waren und Messungen durchgeführt haben. Einzelne Messpunkte konnten dennoch auf Grund verschiedenster Ursachen (Stationswartung, Fehlwerte, etc.) nicht zur Verfügung stehen, was aber für die Cluster-Analyse keine Relevanz hatte und diese Wetterstationen trotzdem berücksichtigt werden konnten. Unvollständige Datenpunkte wurden einfach bei der Cluster-Analyse verworfen, wodurch die tatsächliche Anzahl an Input-Vektoren für ein Jahr kleiner gleich 8784 beträgt. Die Größen der beiden Wetter-Variablen lauten wie folgt:

Tabelle 7: Wetter-Variablen

Variable		Einheit	Wertebereich	Schrittgröße
Temperaturverlauf	Temperatur	°C	unbeschränkt	1 °C
	Globalstrahlung	J/cm ²	unbeschränkt	0,1 J/cm ²
Sonnenverlauf	Relative Helligkeit	%	[0 – 100]	1 %

Tabelle 7 zeigt die unterschiedlichen Wetter-Variablen. Ähnlich zu der Auswahl bei den Zeit-Variablen, gibt es auch bei der Wetter-Variable Sonnenverlauf zwei Formen der Darstellung. Einmal als Globalstrahlung und einmal als relative Helligkeit.

Die Wetter-Variable „Relative Helligkeit“ ist keine Messgröße aus einer Wetterstation, sondern wurde eigens für die Cluster-Analyse erzeugt. Dieser Wert sollte das relative Verhältnis zu einer mittleren Tageshöchstwert-Kurve darstellen. Die Intension dieser Variable war es, dass Zeitpunkte aus Winter und Sommer, welche zwar unterschiedliche Globalstrahlungswerte aufweisen, aber dennoch zu ihrer Jahreszeit als hell oder dunkel gelten, leichter in einen Cluster gegeben werden können. Der Prozess für die Erzeugung der Helligkeitswerte kann anhand Abbildung 19 erläutert werden.

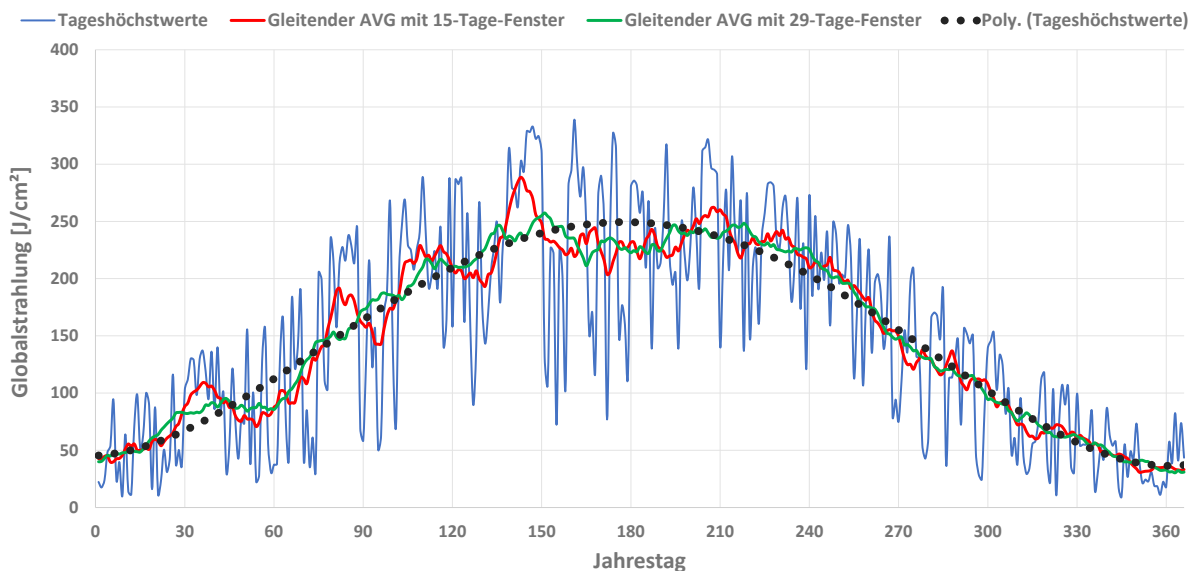


Abbildung 19: Jahresverlauf der Globalstrahlung und der erzeugten mittleren Tageshöchstwert-Kurve

Die blaue Kurve stellt die täglichen Höchstwerte der Globalstrahlung an einer Messstation dar (hier Messstation Lübeck) und diente als Ausgangskurve für die Erstellung der mittleren Tageshöchstwert-Kurve. Mit Hilfe eines gleitenden Durchschnitts bzw. Tiefpass-Filters, konnten die Minima und Maxima der Tageshöchstwerte unterdrückt und die Ausgangskurve geglättet werden. Beim gleitenden Durchschnitt wird ein definierter Betrachtungszeitraum, im konkreten Fall ein 15-Tage-Fenster (± 1 Woche) und ein 29-Tage-Fenster (± 2 Wochen), verwendet und entlang der Kurve geschoben. Der Mittelwert innerhalb dieses Fensters ergibt jeweils den neuen Wert an der aktuellen Position. Je größer der Betrachtungszeitraum, desto stetiger wird die resultierende Kurve (Vergleich rote und grüne Kurve in Abbildung 19). Das Konvergenzziel dieser Methode stellt eine Polynomfunktion 4. Grades (gepunktete Kurve) dar, welche mit Hilfe der Methode der kleinsten Quadrate in MS Excel erstellt wurde. Die Kurvengleichung dazu lautet wie folgt:

$$y = 1,9 \cdot 10^{-8} x^4 - 1,4 \cdot 10^{-4} x^3 + 2,4 \cdot 10^{-2} x^2 + 1,3 \cdot 10^1 x + 45,28 \quad (17)$$

Diese mittlere Tageshöchstwert-Kurve diente als Referenzwert, an welchem die tatsächlichen Globalstrahlungswerte bemessen wurden. Jeder Stundenwert der gemessenen Globalstrahlung wurde anhand dieser mittleren Tageshöchstwert-Kurve bewertet und mit einem Prozentwert zwischen 0% und 100% versehen. Werte oberhalb der mittleren Tageshöchstwert-Kurve wurden ebenfalls mit maximal 100% bewertet.

Wie schon bei den Zeit-Variablen, wurde für das Clustering pro Wetter-Variable nur eine Form verwendet (Globalstrahlung oder relative Helligkeit), da es sonst zu einer doppelten Beschreibung gekommen wäre. Die Anwendung der unterschiedlichen Darstellungsformen wird im Punkt 4.1.1 näher erörtert.

Betrieb-Variablen

Die Variable mit Bezug auf den Betrieb wurde durch die Berechnungswerte ermittelt, welche bei jeder der 8764 Lastflussberechnungen der Quasi-Dynamischen Simulation ermittelt wurden. Sie entspricht Transformatorauslastung in Prozent (siehe auch Unterkapitel 3.2).

3.4. Input-Datensatz für entsprechende Transformatorstation zusammensetzen

In diesem Schritt wurden die transformatorspezifischen Variablen (Wetter und Betrieb) aus dem globalen Input-Datensatz geladen und dadurch der lokale Input-Datensatz erzeugt.

Zeit-Variablen

Die Zeit-Variablen sind an allen Transformatorstationen ident und mussten nicht selektiert werden.

Wetter-Variablen

Für die Wetter-Variablen mussten die jeweiligen Wetterdaten aus dem globalen Input-Datensatz geladen werden. Dazu wurde zunächst die nächstgelegene Wetterstation ermittelt, von der die Wetterdaten übernommen werden sollten. Da nicht immer jede Wetterstation alle Wetter-Variablen misst, konnten die Daten auch von unterschiedlichen Wetterstationen verwendet und zusammengesetzt werden. Der Abstand zwischen den geografischen Koordinaten wurde mit der Harvesine Formel berechnet und lautet wie folgt:

$$a = \sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat_1) \cdot \cos(lat_2) \cdot \sin^2\left(\frac{\Delta lon}{2}\right) \quad (18)$$

$$c = \text{atan2}(\sqrt{a}, \sqrt{1-a}) \quad (19)$$

$$\text{Abstand} = R \cdot c \quad (20)$$

Dabei entspricht *lat* dem Breitengrad und *lon* dem Längengrad der geografischen Koordinaten. Der Erdradius R wurde hier mit 6371 km angewendet, was für Abstandsberechnungen von geringen Distanzen eine gute Genauigkeit ergibt. Abschließend wurde die Wetterstation ausgewählt, welche den minimalen Abstand zur Transformatorstation aufwies und die Wetterdaten in den Input-Datensatz geladen. Falls bei dieser Wetterstation nicht alle Wetter-Variablen zur Verfügung standen (z.B. nur Temperatur aber keine Globalstrahlung), wurden die fehlenden Variablen der nächstgelegenen Wetterstation abgerufen und verwendet.

Anmerkung:

Die beschriebene Vorgangsweise zur Zusammensetzung der Wetter-Variablen wurde entwickelt, um bei einer Anwendung in realen Stromnetzen die richtigen Wetterdaten zu verwenden. Im hier verwendeten SimBench-Datensatz wurden die Wetter-Daten vereinfacht angenommen und nur einzelne Wetterstationen (vorwiegend Norddeutschland) verwendet und nicht die Wetterdaten der nächstgelegenen Wetterstation. Bezugnehmend auf diese verwendeten Wetterstationen, reduzierte sich die Auswahl von Wetterstationen auf die Standorte Hannover und Lübeck. Da die Wetterdaten jeweils einer Zeitreihe der Photovoltaikanlage zugeordnet sind, erfolgte die Zuordnung der Wetterdaten anhand dessen (siehe Tabelle 8):

Tabelle 8: Zuordnung der Wetterdaten zu Zeitreihen der Photovoltaikanlagen

<i>PV1</i>	<i>Hannover</i>
<i>PV2</i>	<i>Lübeck</i>
<i>PV3</i>	<i>Hannover</i>
<i>PV4</i>	<i>Hannover</i>
<i>PV5</i>	<i>Lübeck</i>
<i>PV6</i>	<i>Lübeck</i>
<i>PV7</i>	<i>Hannover</i>
<i>PV8</i>	<i>Lübeck</i>

Betrieb-Variable

Für die Betrieb-Variable wurden die Ergebnisse aus der Quasi-Dynamischen Simulation für die ausgewählte Transformatorstation aus dem globalen Input-Datensatz geladen und nur die jeweilige Transformatorstation ausgewählt. Nach diesem Schritt waren alle Variablen von der ausgewählten Transformatorstation verfügbar und der lokale Input-Datensatz bereit für die Cluster-Analyse.

3.5. Anwendung des Clusteralgorithmus

Der Clusteralgorithmus ist der zentrale Schritt bei der Abschätzung der Unsicherheiten von Belastungsszenarien. Der vorbereitete Input-Datensatz diente dabei als Ausgangspunkt und stellt die Lern-Phase des Algorithmus dar. Der Input-Datensatz wird dabei auf Ähnlichkeitsstrukturen untersucht und anhand dessen in ein oder mehrere Cluster unterteilt. Da für jeden Transformator ein eigener Input-Datensatz existiert (die Wetter-Variablen und Betrieb-Variablen müssen auf die jeweilige Transformatorstation abgestimmt sein), wurde der Input-Datensatz vor der Cluster-Analyse individuell für den zu untersuchenden Transformator zusammengesetzt. Nur die Zeit-Variablen können als globale Größen angesehen werden, welche bei allen Transformatorauswertungen gleich sind.

Es wurden zwei Algorithmen für das Clustering des Datensatzes angewendet. Diese unterscheiden sich einerseits bei der Verarbeitung des Datensatzes (ein- oder

mehrdimensional) und andererseits bei der Ermittlung der Clusterzugehörigkeit (Euklidische Distanz oder Gaußsche Verteilungskurve).

Beide Algorithmen wurden in der Mathematiksoftware MATLAB® Version R2021b geschrieben und ausgewertet. Da einige Funktionen schon in MATLAB® verfügbar sind, wurde zusätzlich zu dem Basispaket der Software auch das Paket Statistics and Machine Learning Toolbox™ verwendet.

3.5.1. Erwartungs-Maximierungs-Algorithmus

Der Erwartungs-Maximierungs-Algorithmus, kurz EM-Algorithmus, arbeitet mit einem Gaußschen Mischmodell, welches in Punkt 2.3 beschrieben wurde. Zum besseren Verständnis ist nachfolgend das Flussdiagramm dargestellt.

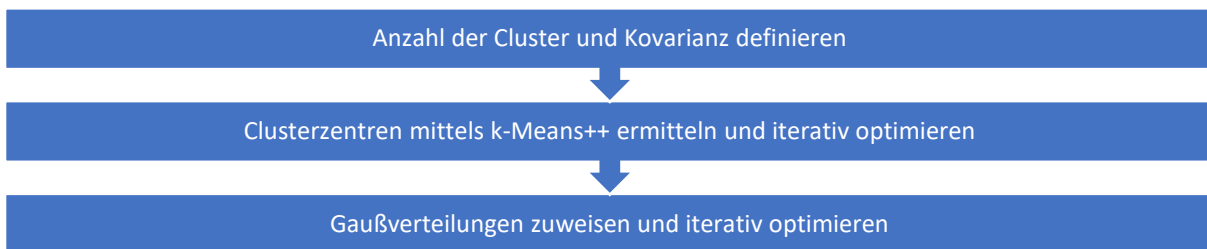


Abbildung 20: Flussdiagramm EM-Algorithmus

Im ersten Schritt muss die Clusteranzahl definiert und die Kovarianzmatrix konfiguriert werden. Anschließend arbeitet der EM-Algorithmus selbstständig und ermittelt die optimalen Clusterzentren iterativ mittels k-Means++. Zuletzt werden die Gaußverteilungen jedes Clusters ebenfalls iterativ optimiert und bei Konvergenz als finale Lösung gewählt.

Der Algorithmus wurde in dieser Arbeit mit Hilfe der MATLAB-Funktion *fitgmdist* realisiert, welche als Eingangsparameter den Input-Datensatz, die Anzahl an Clustern und die Konfiguration der Kovarianzmatrix benötigt.

Die Kovarianzmatrix besteht aus zwei Parametern, welche für das optimale Clustering ausgewählt werden musste. Einmal musste die Form der Kovarianzmatrix definiert werden, ob diese diagonal oder voll ist. Eine diagonale Kovarianzmatrix eignet sich für Modelle, bei denen die Eigenschaften unkorreliert sind. Die so entstandenen Ellipsen liegen in ihrer Haupt- und Nebenachse orthogonal (erstes Diagramm in Abbildung 21). Bei einer vollen Kovarianzmatrix können korrelierende Eigenschaften mit den Ellipsen frei gruppiert werden und sich die entstandenen Ellipsen frei im Raum ausbilden (zweites Diagramm in Abbildung 21). Der zweite Parameter der Kovarianzmatrix definiert, ob die Ellipsen der einzelnen Cluster eine unterschiedliche Struktur aufweisen können (viertes Diagramm in Abbildung 21) oder ob alle gleich groß sind und auch dieselbe Ausrichtung haben (drittes Diagramm in Abbildung 21). Welche dieser Konfiguration sich am besten für den vorliegenden Datensatz eignete, wurde mit Hilfe der beiden Informationskriterien AIC und BIC bewertet.

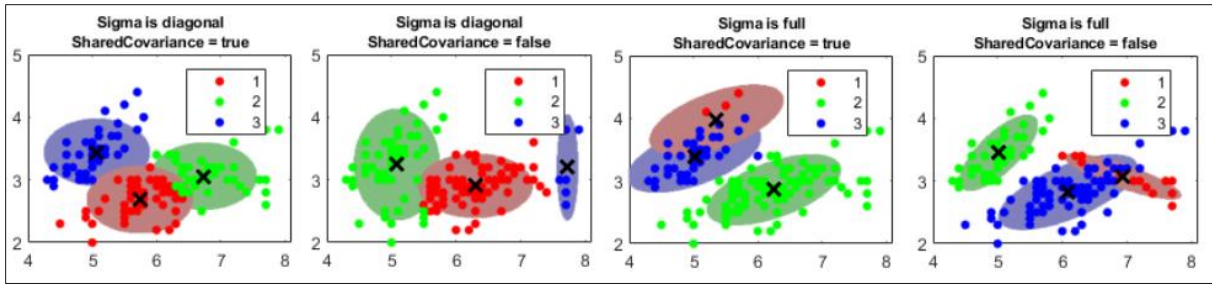


Abbildung 21: Arten der Konfiguration der Kovarianzmatrix [29]

Auswahl der Clusterzentren

An dieser Stelle soll auch noch auf die Auswahl der Startparameter für die Clusterzentren eingegangen werden, die für den EM-Algorithmus verwendet wurden. Da die Funktionsweise des EM-Algorithmus darin besteht, die Parameter für Mittelwert, Kovarianzmatrix und Mischungsverhältnis iterativ zu optimieren, bis diese zu ihrem lokalen Optimum konvergieren, kann sich das Resultat je nach Wahl des Clusterzentrums unterscheiden. Das Clusterzentrum wird hier mit Hilfe des k-Means++ Algorithmus und Mahalanobis-Distanzen ermittelt. Damit sich die Resultate nach dem Clustering durch die Konvergenz zum lokalen Optimum nicht unterscheiden und auch das globale Optimum als finales Resultat verwendet werden kann, wurde der EM-Algorithmus samt Auswahl der Clusterzentren mehrmalig angewendet (in der MATLAB-Funktion *fitgmdist* wurde *Replicates* auf 1000 gesetzt). Der Durchgang mit dem größten Loglikelihood wurde anschließend als finales Resultat weiterverwendet. Da die erzielten Resultate auch bei mehrmaliger Anwendung des EM-Algorithmus ident waren, konnte vom globalen Optimum ausgegangen werden.

3.5.2. Distanzmatrix-Algorithmus

Der hiergenannte Distanzmatrix-Algorithmus ist ein Clusteralgorithmus und wurde eigens für die Anwendung auf den vorliegenden Input-Datensatz entwickelt. Er kann in die Kategorie der hierarchischen Clusteranalysen eingeteilt werden und verwendet für die Berechnung und Ermittlung der Clusterzugehörigkeit eine eindimensionale euklidische Distanzmatrix. Die genaue Abfolge kann mit nachstehendem Flussdiagramm beschrieben werden.

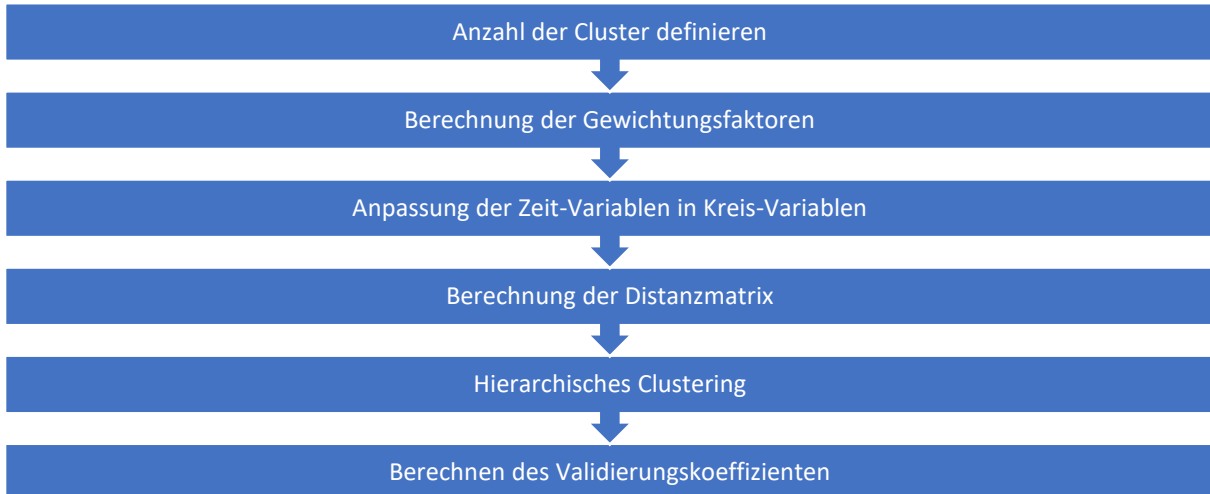


Abbildung 22: Flussdiagramm Distanzmatrix-Algorithmus

Im ersten Schritt muss die Clusteranzahl definiert werden. Im zweiten Schritt werden die Gewichtungsfaktoren berechnet, die den prozentuellen Einfluss der Variablen beschreibt und diese demgemäß gewichtet. Im dritten Schritt werden die Zeit-Variablen in Kreis-Variablen umgerechnet, damit diese Variablen in ihrer realen Struktur berücksichtigt werden können. Im vierten Schritt wird für jede Variable die Distanzmatrix berechnet und anschließend mit Hilfe der Gewichtungsfaktoren zu einer totalen Distanzmatrix zusammengesetzt. Im fünften Schritt wird die totale Distanzmatrix verwendet, um mittels hierarchischen Clusterings einen Clusterbaum zu erzeugen. Im sechsten Schritt wird der Clusterbaum verwendet und mit Hilfe der Validierungskoeffizienten die optimale Clusteranzahl ermittelt.

Clusteranzahl definieren

Die optimale Clusteranzahl wurde mit Hilfe des Silhouettenkoeffizienten ermittelt. Dazu wurde der Algorithmus mehrmals mit einer unterschiedlichen Clusteranzahl angewendet und die optimale Anzahl ermittelt. Dieser Vorgang wird im Punkt 4.1.1 näher erläutert.

Berechnung der Gewichtungsfaktoren

Die Gewichtungsfaktoren wurden verwendet, um den einzelnen Variablen eine Gewichtung zu geben und so den unterschiedlichen Einfluss auf die Transformatorauslastung zu beschreiben. Beispielsweise soll so bei Vorhandensein einer Photovoltaikanlage die Gewichtung vor allem auf die Sonnenscheindauer oder bei Windkraftanlagen vor allem auf die Windstärke fallen. Um die Gewichtungsfaktoren für die Transformatorauslastung zu generieren, wurden zwei Wirkungsbereiche kombiniert: einerseits die Einflussfaktoren auf Erzeugung und Verbrauch von elektrischer Energie und andererseits die Zusammensetzung des Versorgungsgebietes des ausgewählten Transformators.

Einflussfaktoren auf Erzeugung und Verbrauch

Um eine möglichst realitätsnahe Abschätzung der Einflussfaktoren zu erhalten, wurden die Verbraucher Haushalt, Gewerbe und Landwirtschaft zunächst in verschiedene Kategorien geteilt. Diese Kategorien sollen die Energieschwerpunkte innerhalb dieser Verbrauchergruppen darstellen, wodurch eine bessere Gewichtung vorgenommen werden konnte (siehe auch Punkt 2.1.4). Je Kategorie wurden nun alle 5 Einflussfaktoren (Tagesverlauf, Temperaturverlauf, etc.) quantitativ bewertet und einer Gewichtung zwischen 1-wenig und 3-stark zugeteilt. Falls keine Abhängigkeit eines Einflussfaktors zutraf, wurde dieses Feld ausgelassen. Die zugewiesenen Gewichtungsfaktoren sind in Tabelle 9 eingetragen. Für den Bereich Haushalt-Kochen wurde beispielsweise eine starke Abhängigkeit vom Tagesverlauf und eine mäßige Abhängigkeit vom Wochenverlauf zugeteilt.

Tabelle 9: Einflussfaktoren auf Erzeugung und Verbrauch

		Haushalt					Gewerbe			Landwirtschaft			Photovoltaik	Windkraft	Wasserkraft	Biomasse
		Heizen/Kühlen (HVAC)	Kochen	Beleuchtung	Unterhaltung	Warmwasser	Heizen/Kühlen	Warmwasser	Geräte	Heizen/Kühlen	Warmwasser	Geräte				
Zeit	Jahresverlauf	1				2	2	2		2	2				3	1
	Wochenverlauf		2		3				3			3				2
	Tagesverlauf		3	2	3				3			3	2			2
Wetter	Temperaturverlauf	3				3	3	3		3	3		2		2	1
	Sonnenverlauf			3								3			1	

Nachdem die 5 Einflussfaktoren innerhalb eines Energieschwerpunktes gewichtet wurden, mussten noch die Energieschwerpunkte innerhalb einer Verbraucherkategorie gewichtet werden. Je nach Literatur werden diese Energieschwerpunkte unterschiedlich gewichtet. Beispielsweise hat [19] für die Haushaltsbereiche HVAC-Kochen-Beleuchtung-Unterhaltung-Warmwasser eine Gewichtung 51%-4%-5%-21%-19% vorgeschlagen und die Studie [17] eine Gewichtung von 69%-3%-6%-5%-15%. Es ist schnell zu erkennen, dass die unterschiedlichen Gewichtungen einerseits durch klimatische Unterschiede zustande kommen und andererseits sicherlich auch auf den Grad der Elektrifizierung im jeweiligen Sektor zurückzuführen sind. Es wurde daher auf keine Werte aus der Literatur zurückgegriffen, sondern eine eigene Auswertung durchgeführt. Dazu kam die Methode des Paarweisen Vergleichs zur Anwendung, bei dem jeweils zwei Energieschwerpunkt gegeneinander abgewogen wurden. Die resultierenden Faktoren sind in Tabelle 10 und Tabelle 11 aufgelistet.

Tabelle 10: Paarweiser Vergleich - Haushalt

	Heizen/Kühlen	Kochen	Beleuchtung	Unterhaltung	Warmwasser	ERGEBNIS
Heizen/Kühlen		2	2	2	2	40%
Kochen	0		2	2	1	25%
Beleuchtung	0	0		1	0	5%
Unterhaltung	0	0	1		0	5%
Warmwasser	0	1	2	2		25%

Tabelle 11: Paarweiser Vergleich - Gewerbe und Landwirtschaft

	Heizen/ Kühlen	Warmwasser	Geräte	ERGEBNIS
Heizen/Kühlen		1	0	17%
Warmwasser	1		0	17%
Geräte	2	2		67%

Der Paarweise Vergleich ergab bei den Haushaltslasten die größte Relevanz im Bereich Heizen/Kühlen, gefolgt von Kochen und Warmwasser und zum Schluss Beleuchtung und Unterhaltung. Bei den Gewerbe- und Landwirtschaftslasten ist der Bereich Geräte am stärksten ausgeprägt gefolgt von Heizen/Kühlen und Warmwasser. Die ermittelten Verhältnisse der Energieschwerpunkte innerhalb einer Verbraucherkategorie wurden zuletzt mit den quantitativen Bewertungen kombiniert. Diese Operation lässt sich als Matrix-Vektor-Produkt darstellen, wofür nachfolgend ein Beispiel für den Verbrauchertyp Haushalt angeführt ist:

$$\begin{array}{cccccc|c}
 & & & & & & 40\% \\
 & & & & & & 25\% \\
 & & & & & & 5\% \\
 & & & & & & 5\% \\
 & & & & & & 25\% \\
 \hline
 1 & 0 & 0 & 0 & 2 & 0,9 \\
 0 & 2 & 0 & 3 & 0 & 0,7 \\
 0 & 3 & 2 & 3 & 0 & 1 \\
 3 & 0 & 0 & 0 & 3 & 2 \\
 0 & 0 & 3 & 0 & 0 & 0,2
 \end{array} \quad (21)$$

Die Matrix entspricht den quantitativen Bewertungen für den Haushalt aus Tabelle 9 und der Vektor entspricht den Verhältnissen der Energieschwerpunkte ebenfalls für Haushalte aus Tabelle 10. Die blau hinterlegten Werte aus Formel (21) sind die resultierenden Einflussfaktoren der Verbraucherkategorie Haushalt. Zusammen mit den Einflussfaktoren aus den anderen Verbraucher- und Erzeugerkategorien ergab sich somit die Einflussfaktor-Matrix **E**. Diese Matrix wurde je Kategorie normiert, sodass die Summe der Einflussfaktoren innerhalb einer Verbraucherkategorie 100% ergab. Die komplette Einflussfaktor-Matrix **E** setzt sich wie folgt zusammen:

Tabelle 12: Einflussfaktor-Matrix E

	H	G	I	PV	Wind	Wasser	Bio
Zeit-Jahresverlauf	19%	12%	12%	0%	50%	60%	14%
Zeit-Wochenverlauf	14%	35%	35%	0%	0%	0%	29%
Zeit-Uhrzeit	22%	35%	35%	29%	0%	0%	29%
Wetter-Temperatur	42%	18%	18%	29%	25%	0%	14%
Wetter-Sonnenverlauf	3%	0%	0%	43%	25%	40%	14%

Zusammensetzung des Versorgungsgebietes

Der zweite Teil des Gewichtungsfaktors betrachtet das Versorgungsgebiet des ausgewählten Transformators. Je nach Zusammensetzung des Versorgungsgebietes, sprich mehr oder weniger Haushalte oder mehr oder weniger Photovoltaikanlagen etc., werden sich die Variablen unterschiedlich stark auf die Auslastung des Transformators auswirken und sollten dadurch auch unterschiedlich stark gewichtet sein. Die Information zum Versorgungsgebiet

kann aus der Transformatorbezeichnung gewonnen werden. Beispielsweise lautet die Bezeichnung eines MV/LV-Transformators *MV1.101-LV1.105-Trafo 1*, wobei die Zahl nach den Buchstaben *LV* das entsprechende Versorgungsgebiet bezeichnet und im konkreten Beispiel 1.105 entspricht. Die Zahl nach den Buchstaben *MV* bezeichnet das überlagerte Mittelspannungsnetz (hat hier keine Relevanz). Nun wurden alle Verbraucher und Erzeuger, die in diesem Versorgungsgebiet angeschlossen sind, identifiziert, indem die bereitgestellte SimBench-Datenbank für Erzeuger und Verbraucher (jeweils eine eigene CSV-Datei) nach diesem Versorgungsgebiet gefiltert wurde. Die Summe der Anschlussleistung der jeweiligen Erzeuger- und Verbrauchertypen wurde anschließend normiert und ergab so den Versorgungsgebiet-Vektor V . Beispielsweise würde ein Versorgungsgebiet mit 6 Haushalten, 4 Gewerbelasten, 10 Landwirtschaftslasten und keinerlei Erzeugungseinheiten folgenden Versorgungsgebiet-Vektor aufweisen:

$$V = [6 \ 4 \ 10 \ 0 \ 0 \ 0 \ 0]^T = [0,3 \ 0,2 \ 0,5 \ 0 \ 0 \ 0]^T \quad (22)$$

Die beiden Größen Einflussfaktor-Matrix E und Versorgungsgebiet-Vektor V wurden nun mit folgender Formel zum Gewichtungsfaktor-Vektor G zusammengeführt:

$$G = E \cdot V \quad (23)$$

Anpassung der Zeit-Variablen in Kreis-Variablen

Da es sich bei den Zeit-Variablen nicht um skalare Größen wie beispielsweise Temperatur oder Windstärke handelt, mussten die Zeit-Variablen vor der Cluster-Analyse angepasst werden, um eine korrekte Verarbeitung durch den Clusteralgorithmus zu gewährleisten. Denn die Monate Jänner und Dezember hätten als skalare Größen den Wert 1 und 12 zugewiesen und dadurch die maximale Distanz zueinander, obwohl sie sehr nahe beisammen liegen. Um dieses Problem zu lösen, wurden die vorhin zugeordneten Zeit-Variablen mit folgender Formel auf Kreisvariablen umgerechnet:

$$V_{\text{Zeit_kreis}} = V_{\text{Zeit_skalar}} \cdot \frac{2 \cdot \pi}{\max\{V_{\text{Zeit_skalar}}\}} \quad (24)$$

Als Winkelmaß wurde die Einheit Rad verwendet. Der Monat Jänner bekommt beispielsweise statt der skalaren Größe 1 die Kreisvariable $\pi/6$ zugewiesen oder der Monat Juni statt der skalaren Größe 6 die Kreisvariable π . Durch die erhaltenen Kreisvariablen ist sofort die Nähe bzw. der geringe Abstand der Monate Jänner und Dezember zu erkennen, da die Winkelgrößen $\pi/6$ und 0π (bzw. 2π) beisammen liegen. Das gleiche Verfahren wurde bei der Anpassung von Uhrzeit, Wochenstunde, Wochentag und Jahrestag angewendet. Diese Methodik zur Umwandlung in Kreisvariablen wurde aus der Arbeit [30] entnommen.

Berechnung der Distanzmatrix

Die Distanzmatrix dient dazu, die Unterschiede bzw. Unähnlichkeiten der Belastungsszenarien zu ermitteln und so eine Einteilung beim anschließenden Clustering zu ermöglichen. Die Distanzmatrix wird auch oft als Unähnlichkeitsmatrix (eng. Dissimilarity Matrix) bezeichnet, was diesen Sachverhalt noch besser auf den Punkt bringt.

Die Distanz-Matrix M_i wurde erzeugt indem alle Belastungsszenarien, welche jeweils einen Vektor in Abbildung 18 entsprechen, innerhalb einer Variablen-Kategorie (z.B. Wetter-Variable Temperatur) gegeneinander verglichen wurden und die Differenz berechnet wurde. Die Differenz wurde anschließend auf die Maximaldifferenz normiert, wodurch sich Werte

zwischen 1 und 0 bzw. 100% und 0% ergeben. Tabelle 13 zeigt dazu ein einfaches Beispiel, bei dem 12 Belastungsszenarien in unterschiedlichen Monaten gegenübergestellt sind.

Tabelle 13: Distanzmatrix mit Zeit-Variable Monat

	Jänner	Februar	März	April	Mai	Juni	Juli	August	September	Oktober	November	Dezember
Jänner	0%	17%	33%	50%	67%	83%	100%	83%	67%	50%	33%	17%
Februar	17%	0%	17%	33%	50%	67%	83%	100%	83%	67%	50%	33%
März	33%	17%	0%	17%	33%	50%	67%	83%	100%	83%	67%	50%
April	50%	33%	17%	0%	17%	33%	50%	67%	83%	100%	83%	67%
Mai	67%	50%	33%	17%	0%	17%	33%	50%	67%	83%	100%	83%
Juni	83%	67%	50%	33%	17%	0%	17%	33%	50%	67%	83%	100%
Juli	100%	83%	67%	50%	33%	17%	0%	17%	33%	50%	67%	83%
August	83%	100%	83%	67%	50%	33%	17%	0%	17%	33%	50%	67%
September	67%	83%	100%	83%	67%	50%	33%	17%	0%	17%	33%	50%
Oktober	50%	67%	83%	100%	83%	67%	50%	33%	17%	0%	17%	33%
November	33%	50%	67%	83%	100%	83%	67%	50%	33%	17%	0%	17%
Dezember	17%	33%	50%	67%	83%	100%	83%	67%	50%	33%	17%	0%

Die erzeugten 6 Distanz-Matrizen M_1 bis M_6 wurden anschließend zu einer gesamten totalen Distanz-Matrix D kombiniert und dabei mit dem zuvor berechneten Gewichtungsfaktor-Vektor G verknüpft. Es sind 6 Distanz-Matrizen, weil es 5 Eigenschaften (Zeit und Wetter) gibt und zusätzlich die Transformatorauslastung, also insgesamt 6 Variablen. Die Formel dazu lautet wie folgt:

$$D = \sum_{i=1}^6 M_i \cdot G(i) \quad (25)$$

Dabei entspricht M_i den Distanz-Matrizen der 6 Variablen und $G(i)$ der jeweiligen Komponente des Gewichtungsfaktor-Vektors G .

Hierarchisches Clustering

Beim Hierarchischen Clustering wird der Datensatz (in diesem Fall die totale Distanz-Matrix D) auf Abstände zwischen den einzelnen Datenpunkten untersucht und dabei Cluster gebildet, bei denen innerhalb eines Clusters geringe Distanzen zwischen den Datenpunkten bestehen, wohingegen zu anderen Clustern größere Distanzen herrschen. Als Abstandsmaß wurde hier die durchschnittliche Distanz zwischen zwei Clustern verwendet. Die genaue Abfolge beim hierarchischen Clustering ist aus Punkt 2.3 zu entnehmen.

Der entstandene Clusterbaum dient als Ausgangspunkt für die Zuweisung aller Datenpunkte in einen Cluster. Nachdem eine Clusteranzahl definiert wurde, wurde das Dendrogramm bildlich in der Höhe geschnitten (siehe Abbildung 12), in welcher sich die entsprechende Clusteranzahl befindet. Alle darunter liegenden Datenpunkte fallen somit in diese Cluster.

Berechnung des Validierungskoeffizienten

Der Validierungskoeffizient dient dazu, die optimale Anzahl von Clustern zu ermitteln bzw. in welcher Höhe das Dendrogramm sinnvollerweise abgeschnitten werden soll. Allgemein verwenden Validierungskoeffizienten dabei die Kompaktheit, Separierung und Dichte der einzelnen Cluster (oder eine Kombination daraus). Es wurde hier der Silhouettenkoeffizient verwendet, welcher die Parameter Separierung und Kompaktheit kombiniert und daraus die Differenz bildet. Anschließend wird diese Differenz mit Hilfe des maximalen Kompaktheits- oder Separierungsmaßes normiert. Die Formel dazu lautet wie folgt [31]:

$$Sill = \frac{1}{N} \sum_{c_k \in C} \sum_{x_i \in c_k} \frac{Sep(x_i, c_k) - Comp(x_i, c_k)}{\max\{Comp(x_i, c_k), Sep(x_i, c_k)\}} \quad (26)$$

$$Comp(x_i, c_k) = \frac{1}{|c_k|} \sum_{x_j \in c_k} d_e(x_i, x_j) \quad (27)$$

$$Sep(x_i, c_k) = \min_{c_l \in C \setminus c_k} \left\{ \frac{1}{|c_l|} \sum_{x_j \in c_l} d_e(x_i, x_j) \right\} \quad (28)$$

Die Kompaktheit eines Clusters wird hier als Durchschnittswert verwendet und die Separierung als minimaler Abstand zum nächstgelegenen Cluster.

3.6. Berechnen der Cluster-Dichtefunktion

Um bei der Prognose eines Belastungsszenarios die Unsicherheiten abschätzen zu können, wurden vorbereitend die eingeteilten Cluster analysiert und deren relative Häufigkeit ermittelt. Es wurden dazu die Daten innerhalb eines Clusters als Dichtefunktion der relativen Häufigkeit dargestellt, um einerseits die Unterschiede zwischen den Clustern besser zu erkennen und andererseits für die Prognose die notwendigen Werte der relativen Häufigkeit verfügbar zu machen. Die Dichtefunktion stellt dabei die relative Häufigkeit zu jedem diskreten Schritt einer Eigenschaft dar. Die Formel dazu lautet:

$$h_i(x) = \frac{H_i(x)}{n} \quad (29)$$

Dabei entspricht i der jeweiligen Eigenschaft, $h_i(x)$ der relativen und $H_i(x)$ der absoluten Häufigkeit bei dem entsprechenden Eigenschaftswert x und n der Anzahl von Datenpunkten. Beispielsweise würden 30 Werte bei der Temperatur 0°C (entspricht $H_{Temp}(0^\circ\text{C})$) und einer Gesamtmenge von 200 Datenpunkten (entspricht n) die relative Häufigkeit von 0,15 bzw. 15% (entspricht $h_{Temp}(0^\circ\text{C})$) ergeben. Die relative Häufigkeit wird immer nur auf die Anzahl der Datenpunkte innerhalb eines Clusters bezogen und nicht auf den ganzen Datensatz. Abbildung 23 zeigt beispielhaft die Dichtefunktion der Temperatur nach einem Clustering mit zwei resultierenden Clustern. Hierbei gibt es einen Cluster mit vorzugsweise niedrigeren Temperaturen (Blau) und einen Cluster mit vorzugsweise höheren Temperaturen (Orange). Die Werte an der Ordinate beziehen sich jeweils auf die Schrittgröße der Eigenschaft. Beispielsweise entspricht die relative Häufigkeit bei 0°C (Schrittgröße 1°C) etwa 3% beim blauen und etwa 1% beim orangen Cluster.

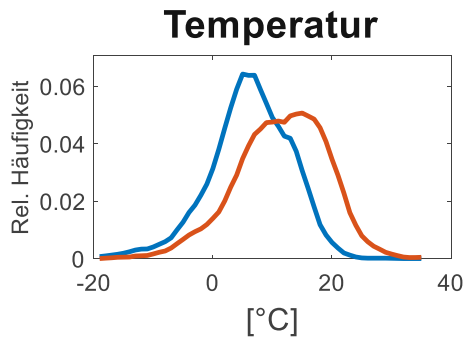


Abbildung 23: Beispiel für Dichtefunktion der eingeteilten Cluster

Abbildungen der Dichtefunktion, wie sie hier zu sehen ist, wurden auch bei der Auswertung in Kapitel 4 zur Beurteilung der Clusterergebnisse herangezogen. Da es durch die diskreten Schritte aller Eigenschaften und auch geringen Anzahl an Datenpunkten manchmal zu Sprüngen in der Dichtefunktion kam (vergleiche Abbildung 6), wurden die Abbildungen zur besseren Veranschaulichung mit Hilfe eines gleitenden Durchschnitts geglättet. Dadurch konnten die Informationen aus dem Vergleich zwischen den Dichtefunktionen besser interpretiert werden und auch der Fokus auf die wesentlichen Eigenschaften (unterschiedliche Schwerpunkte) der Dichtefunktion fallen.

3.7. Prognose für Belastungsszenario

Ein Belastungsszenario wurde in dieser Arbeit mit Hilfe der Eigenschaften aus dem Input-Datensatz definiert. Es wurde dazu eine Anzahl beliebiger Eigenschaften verwendet, die zusammen das Belastungsszenario ergaben, wie beispielsweise „Freitag, 14 Uhr und 10 °C“ oder „Dezember, 19 Uhr und 50 % rel. Helligkeit“. Wie viele Eigenschaften sinnvoll für die Beschreibung eines Belastungsszenarios sind, wurde im Anschluss an die Evaluierung der Clusterergebnisse in Punkt 4.1.3 definiert. Nachfolgenden wird deshalb die Anzahl der verwendeten Eigenschaften mit Z bezeichnet.

Nachdem das Belastungsszenario mit Hilfe der Eigenschaften definiert wurde, wurde innerhalb der ausgewählten Eigenschaften die relative Häufigkeit $h_{i,c}(x)$ bei dem entsprechenden Wert x der Eigenschaft i für jeden Cluster c (klein c) abgefragt. Anschließend wurden die abgefragten relativen Häufigkeiten pro Eigenschaft wie folgt normiert, damit sich die relativen Häufigkeiten über alle Cluster C (groß C) bei diesem Abfragewert zu 1 bzw. 100 % ergaben:

$$h_{i,c_normiert}(x) = \frac{h_{i,c}(x)}{\sum_{c=1}^C h_{i,c}(x)} \quad (30)$$

Es existierten also innerhalb jeder Eigenschaft, welche das Belastungsszenario beschrieb, für jeden Cluster (Anzahl der Cluster entspricht C) eine normierte relative Häufigkeit. Diese normierte relative Häufigkeit wurde zum Schluss für jeden Cluster über die Eigenschaften, welche das Belastungsszenario beschrieben, aufsummiert und durch die Anzahl dieser Eigenschaften dividiert. Die Division durch die Anzahl der Eigenschaften bedeutet, dass jede Eigenschaft, welche das Szenario beschreibt, gleichgewichtet berücksichtigt wurde:

$$h_c = \frac{\sum_{i=1}^Z h_{i,c_normiert}(x)}{\text{Anzahl der Szenario_Eigenschaften}} \quad (31)$$

Der resultierende Wert der Clustergewichtung h_c ist ein Prozentwert und gibt an, wie sehr ein Cluster bei dem jeweiligen Belastungsszenario vertreten ist und deshalb auch in diesem Ausmaß bei der Prognose berücksichtigt werden muss. Im abschließenden Schritt wurden diese Prozentwerte herangezogen, die Dichtefunktion der Transformatorauslastung innerhalb jedes Clusters damit gewichtet und mit den restlichen Clustern aufsummiert:

$$h_{Trafo_total}(x) = \sum_{c=1}^C h_{Trafo,c}(x) \cdot h_c \quad (32)$$

Das nachfolgende Beispiel soll diesen Vorgang noch besser veranschaulichen:

Abbildung 24 zeigt ein Belastungsszenario, welches durch die Eigenschaften Uhrzeit und Temperatur definiert wurde. Es wurde das Belastungsszenario „12 Uhr und 0°C“ abgefragt. Der Uhrzeit zufolge besitzt der orange Cluster eine relative Häufigkeit von $h_{Zeit,Orange} = 0,08$ und der blaue Cluster eine relative Häufigkeit von $h_{Zeit,Blau} = 0$. Der Temperatur zufolge besitzt der orange Cluster eine relative Häufigkeit von $h_{Temp,Orange} = 0,01$ und der blaue Cluster eine relative Häufigkeit von $h_{Temp,Blau} = 0,03$. Die dazugehörigen normierten relative Häufigkeiten wären bei der Uhrzeit 100% bzw. 0% und bei der Temperatur 25% bzw. 75%.

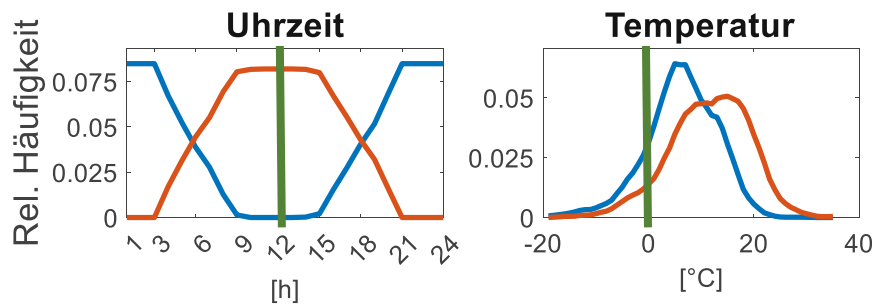


Abbildung 24: Beispiel eines Belastungsszenarios

Die resultierenden Clustergewichtungen ergaben sich bei diesem Belastungsszenario zu $h_{Orange} = 62,5\%$ und $h_{Blau} = 37,5\%$. Mit diesen Gewichtungen wurden die beiden Dichtefunktionen der Transformatorauslastung gemischt, wodurch sich die resultierende grüne Dichtefunktion nach Abbildung 25 ergab.

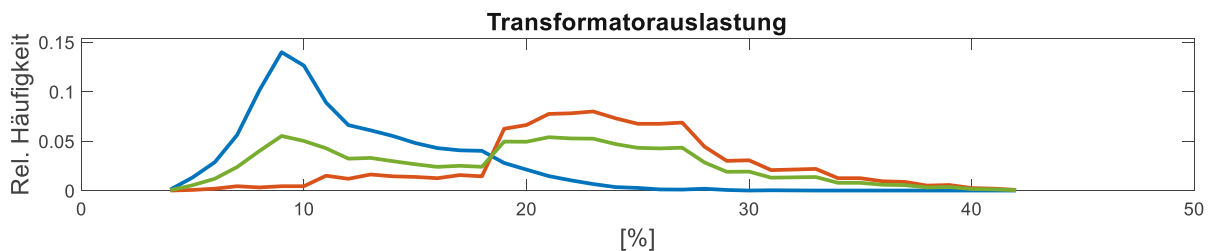


Abbildung 25: Resultierende Dichtefunktion des Beispielszenarios

3.8. Evaluierung der Prognose

Die erzeugte Dichtefunktion bei einem Belastungsszenario wurde anschließend mit „realen“ Test-Daten und der gesamten Dichtefunktion über ein Jahr verglichen und bewertet. Die Test-Daten stellen Werte der Transformatorauslastung dar, welche aus den zusätzlich generierten Jahreszeitreihen erzeugt wurden. Es wurden damit neue Jahre erzeugt, wiederum eine Quasi-Dynamische Lastflussrechnung durchgeführt und die erhaltenen Auslastungsdaten bei dem jeweiligen Belastungsszenario mit der Prognose verglichen. Für die Bewertungen wurden die

Dichtefunktionen der Prognose und der Vergleichsdaten (Test-Daten und Jahresauslastung) herangezogen und mit Hilfe der Methode der kleinsten Quadrate bewertet. Dazu wurde die Summe der quadrierten Abstände SSE aller diskreten Schritte i mit folgender Formel berechnet. Als Referenz m dienten jeweils die Werte der Test-Daten:

$$SSE = \sum_{i=1}^n (y_m - y_i)^2 \quad 33$$

Des Weiteren wurde die Unsicherheit mit Hilfe der statistischen Parameter Erwartungswert, Median, Interquartilsabstand und Spannweite bewertet.

4. Ergebnisse

Die Fragestellungen in dieser Arbeit thematisiert die Abschätzung der Unsicherheiten an einer Ortsnetzstation, welche im konkreten Fall die Auslastungsdaten betreffen. Um die Unsicherheiten qualitativ und quantitativ bewerten zu können, wurde zunächst die Ausgangslage dieser Unsicherheiten bzw. Auslastungsdaten festgehalten. Wie stark eine Transformatorstation über ein Jahr betrachtet ausgelastet ist, kann beispielsweise mit Hilfe eines Histogramms der 8784 Stundenwerte in einem Jahr veranschaulicht werden. Diese Stundenwerte entsprechen den Einzelberechnungen aus der Quasi-Dynamischen-Lastflussberechnung, welche in Punkt 3.2 beschrieben wurde.

Für die Auswertungen in diesem Kapitel wurde die Transformatorstation *MV2.101-LV1.203-Trafo 1* untersucht. Die Station besteht aus einem Transformator mit einer Nennscheinleistung von 160kVA und der Schaltgruppe YNyn5. Abbildung 26 zeigt dazu das Histogramm der Auslastungswerte an dieser Transformatorstation. In Blau sind die Auslastungswerte im Basisszenario-0 und in Orange die Auslastungswerte im Entwicklungsszenario-2 zu sehen. Die Wahl fiel auf diese Station, weil sich in dem Versorgungsgebiet im Basisszenario-0 anteilmäßig die meisten Photovoltaikanlagen, verglichen mit den restlichen Transformatorstationen im Mittelspannungs-Niederspannungsnetz, befinden und dadurch eine möglichst hohe Abhängigkeit zu den Wetter-Variablen abgeleitet wurde. An der Abszisse befinden sich die Auslastungswerte in Prozent und an der Ordinate die relative Häufigkeit der Auslastungswerte. Die Klassenbreite wurde in 1%-Schritten gruppiert.

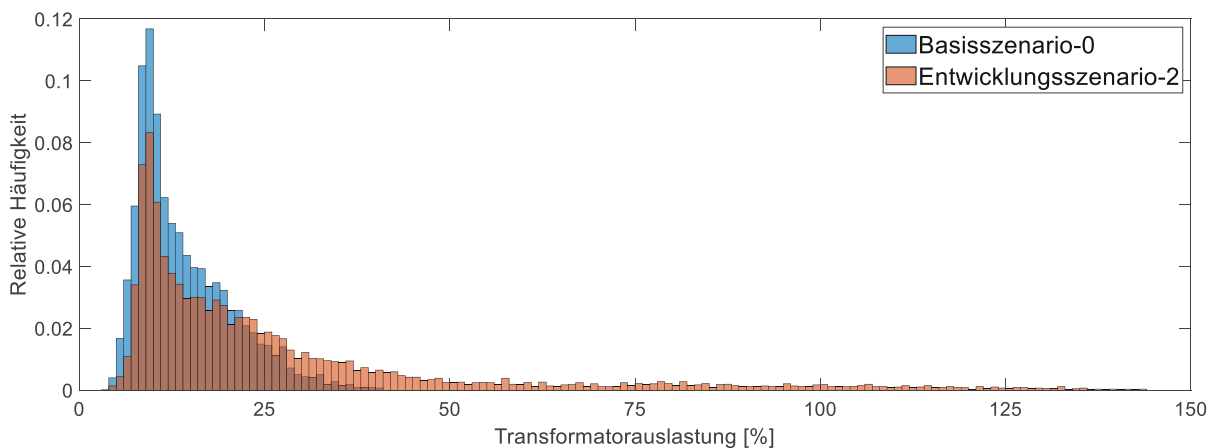


Abbildung 26: Histogramm der Transformatorauslastung über ein Jahr

Die Auslastungsdaten beider Szenarien weisen eine rechtsschiefe Verteilung (mehr Belastungsszenarien mit geringer Auslastung als mit hoher Auslastung) auf. Die Auslastungen im Basisszenario-0 befinden sich im Bereich zwischen 4% und 42%. Betrachtet man dabei die Auslastungsdaten als Dichtefunktion, so besitzt sie einen Mittelwert von 14,2% und eine Standardabweichung von 6,7%. Mit rund 12%iger Häufigkeit tritt am häufigsten ein Belastungsszenario zwischen 9-10% Transformatorauslastung auf. Die Auslastungen im Entwicklungsszenario-2 befinden sich im Bereich zwischen 4% und 145%. Betrachtet man die Auslastungsdaten wieder als Dichtefunktion, so besitzt sie einen Mittelwert von 27,1% und eine Standardabweichung von 25,6%. Mit rund 8%iger Häufigkeit tritt ebenfalls am häufigsten ein Belastungsszenario zwischen 9-10% Transformatorauslastung auf. Die Auslastungswerte unterscheiden sich zwischen den Szenarien, da sich die Anzahl bzw. Größe der angeschlossenen Lasten und Erzeuger in dem Versorgungsgebiet unterscheidet. Wie sich dabei die Lasten und Erzeuger in dem betrachteten Versorgungsgebiet zusammensetzen, ist in Tabelle 14 aufgelistet.

Tabelle 14: Zusammensetzung des Versorgungsgebietes LV1.203

		Basisszenario-0		Entwicklungsszenario-2	
		Anzahl	Summe Leistung	Anzahl	Summe Leistung
Verbraucher	Haushalt	3	7 kW	3	7 kW
	Landwirtschaft	10	73 kW	10	73 kW
	E-Ladestellen	-	-	7	77 kW
	Luft-Wärmepumpe	-	-	3	10 kW
	Erd-Wärmepumpe	-	-	5	16 kW
SUMME		13	80 kW	28	183 kW
Erzeuger	Photovoltaik	4	160 kW	8	468 kW
	Batterie	-	-	5	206 kW
SUMME		4	160 kW	13	674 kW

Schon im Basisszenario-0 ist mit der installierten PV-Leistung doppelt so viel Erzeugung in dem Netzgebiet vorhanden, als die installierte Leistung der Erzeuger. Dieser Trend weitet sich im Entwicklungsszenario-2 sogar noch aus, wonach dreimal so viel Erzeugung in dem Netz vorhanden ist wie Verbraucher. Da diese Erzeugung stark von der Sonneneinstrahlung abhängt, wird durch diese Eigenschaft ein stärkerer Einfluss vermutet. Bei den vorhandenen Verbrauchern hingegen wird eine starke Abhängigkeit der Zeit-Variablen vermutet, da diese klassisch einem Tages- und Jahresrhythmus folgen. Abbildung 27 zeigt das Einliniendiagramm des Versorgungsgebietes LV1.203. In Orange sind die Erzeuger und Lasten markiert, welche erst im Entwicklungsszenario-2 hinzugefügt wurden.

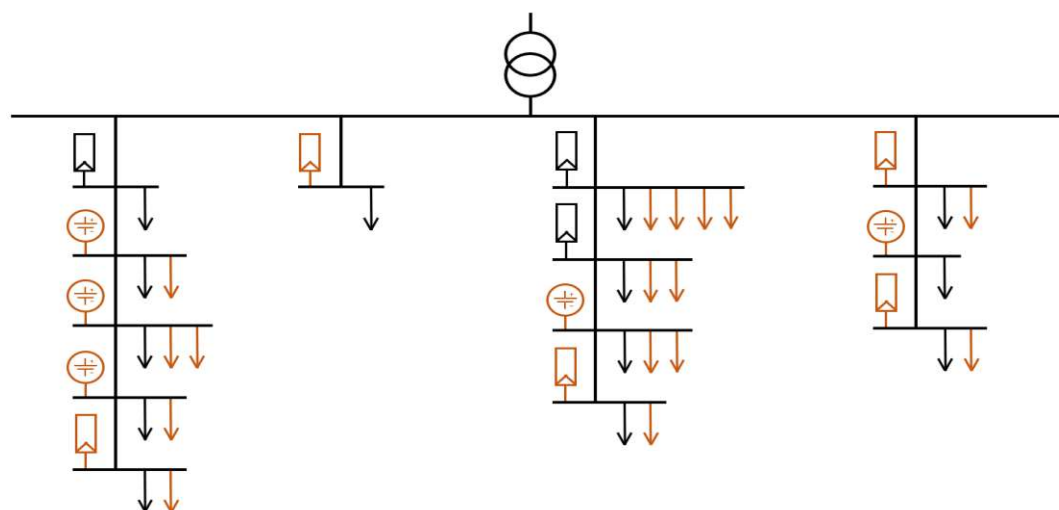


Abbildung 27: Einliniendiagramm des Versorgungsgebietes LV1.203

Ein weiterer Punkt bei der Festlegung der Ausgangslage bezieht sich auf die verschiedenen Eigenschaften, welche zusammen mit der Transformatorauslastung beim Clustering verarbeitet wurden. Abbildung 28 zeigt dazu für jede Eigenschaft die Ausgangslage der zugehörigen Dichtefunktion. Diese Dichtefunktionen treten beispielsweise auf, wenn alle Datenpunkte in einen gemeinsamen Cluster gegeben werden. An der Abszisse befinden sich die jeweiligen Wertebereiche samt Einheit und an der Ordinate die relative Häufigkeit. In dieser Abbildung sind die Werte der relativen Häufigkeit angeführt, werden aber in den nachfolgenden Abbildungen zur besseren Veranschaulichung ausgeblendet. Der Zweck

dieser Darstellung besteht mehr darin, die relative Häufigkeit zwischen den Belastungsszenarien zu vergleichen. Beispielsweise soll gezeigt werden, dass Belastungsszenarien bei Temperaturen um 10°C mehr als doppelt so häufig sind bzw. mehr als doppelt so oft auftreten, wie Belastungsszenarien bei Temperaturen um 0°C.

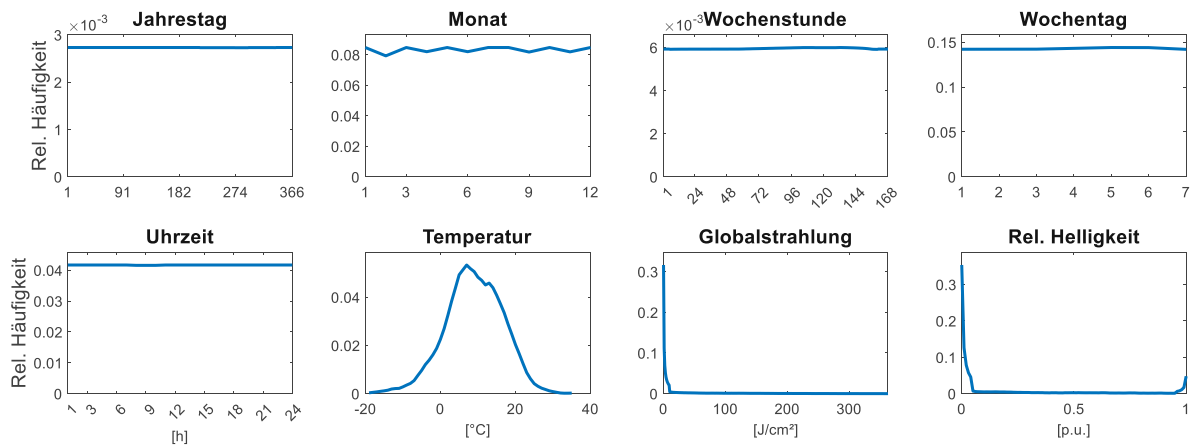


Abbildung 28: Ausgangslage der Dichtfunktionen der Eigenschaften aus dem Input-Datensatz

Alle Zeit-Variablen zeigen eine kontinuierliche Gleichverteilung, da beispielsweise jeder Jahrestag oder jede Uhrzeit gleich oft in dem Datensatz vertreten ist. Bei der Eigenschaft Monat kann man die leicht variierende Anzahl an Tagen erkennen (Februar mit 28 Tagen). Aus den Wetter-Variablen kann man erkennen, dass die durchschnittliche Temperatur bei rund 10°C liegt und die meisten Datenpunkte keine Sonneneinstrahlung aufweisen (in etwa die Hälfte aller Datenpunkte).

4.1. Clustering

In diesem Unterkapitel wurden die beiden Clusteralgorithmen auf den Datensatz angewendet, die Clusterergebnisse gegenübergestellt und die gewonnenen Erkenntnisse für eine Optimierung und Anpassung der Algorithmen verwendet. Es wurde zunächst nur der Input-Datensatz von dem Basisszenario-0 (sprich das Netz 1-MVLV-semiurb-all-0-no_sw) verwendet und dabei die zuvor erwähnte Transformatorstation MV2.101-LV1.203-Trafo 1 untersucht. Das Entwicklungsszenario-2 wurde erst anschließend verwendet, um nach einer Optimierungsrunde weitere Auswertungen durchzuführen.

4.1.1. Erste Evaluierungsrunde

In der ersten Evaluierungsrunde wurden die beiden Clusteralgorithmen konfiguriert, die optimale Clusteranzahl bzw. der optimale Suchbereich der Clusteranzahl evaluiert und die Auswahl der Eigenschaften für den Input-Datensatz untersucht. Diese Schritte dienten dazu, erste Erkenntnisse zu gewinnen und dadurch am Ende der Evaluierungsrunde Anpassungen für ein optimales Clusterergebnis zu erhalten.

Konfiguration der Algorithmen

EM-Algorithmus

Bei der Konfiguration des EM-Algorithmus musste die Art der Kovarianzmatrix definiert werden (voll/diagonal) und zusätzlich, ob sich die Kovarianzmatrix zwischen den einzelnen Clustern unterscheiden durfte (zusammen/getrennt). Um das herauszufinden, wurde die

Transformatorauslastung zunächst paarweise mit jeder Eigenschaft einzeln untersucht (der Input-Datensatz wurde auf einen 2-dimensionalen Input-Datensatz reduziert) und der EM-Algorithmus darauf angewendet. Die beste Konfiguration wurde anhand des AICs und BICs ermittelt, wobei die Ausführung mit dem kleinsten Wert bevorzugt wurde. Abbildung 29 zeigt beispielhaft die Ergebnisse der vier Konfigurationsarten der Kovarianzmatrix anhand des reduzierten Input-Datensatzes Uhrzeit-Transformatorauslastung. Die Clusteranzahl wurde zwischen 1-10 variiert.

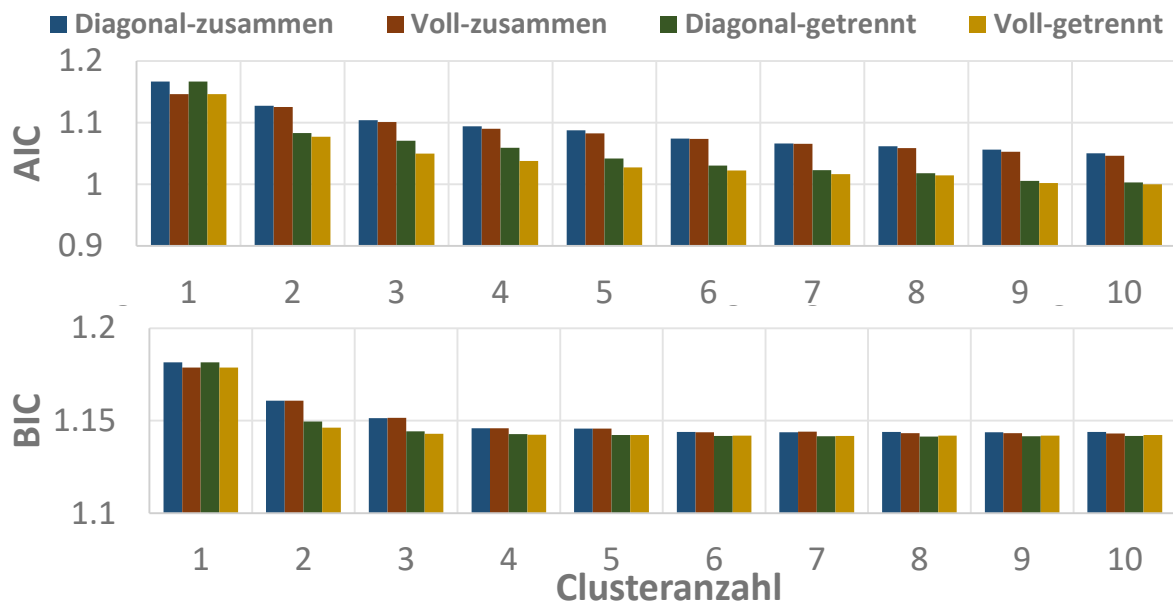


Abbildung 29: AIC und BIC der Kombination Uhrzeit-Transformatorauslastung

Die Ergebnisse des AICs und BICs weisen beide sehr ähnliche Werte bzw. Verläufe auf. Man kann gut erkennen, dass die Konfigurationen, bei denen die Kovarianzmatrix aller Cluster dieselbe Ausführung haben muss (im Diagramm die Farben Blau und Rot bzw. in der Diagrammlegende „zusammen“) tendenziell schlechtere Ergebnisse beim Clustering erzielen, als bei denen die Kovarianzmatrix aller Cluster unterschiedlich sein darf (im Diagramm die Farben Grün und Gelb bzw. in der Diagrammlegende „getrennt“). Weiters ist gut zu erkennen, dass mit steigender Anzahl an Clustern auch die Qualität des Clusterings steigt, da AIC und BIC mit einzelnen Ausnahmen monoton sinken. Nimmt man die Werte des AIC oder BIC aus Abbildung 29 her und betrachtet jeweils die beste Konfiguration bei der Clusteranzahl von 1-10, so kann man anteilmäßig beurteilen, welche Konfiguration zu bevorzugen ist. Es würde demnach laut Abbildung 29 bei jeder Clusteranzahl die Variante „Voll-getrennt“ bevorzugt werden und daher einen Anteil von 100% zugeschrieben bekommen.

Die Konfiguration der Kovarianzmatrix wurde auch bei den restlichen Eigenschaften ausgewertet. Dazu wurde wieder jede Eigenschaft paarweise mit der Transformatorauslastung geclustert und dabei die Clusteranzahl zwischen 1-10 variiert. Anhand des AICs und BICs wurde bei jeder der 10 Clusteranzahlen die bevorzugte Konfiguration ermittelt. Abbildung 30 zeigt für alle Eigenschaften die anteilmäßig bevorzugte Kovarianzmatrix. Beispielsweise ergibt sich bei der Eigenschaft Jahrestag laut BIC die bevorzugte Konfiguration zu 70% „Voll-getrennt“ und zu 30% „Diagonal-getrennt“.

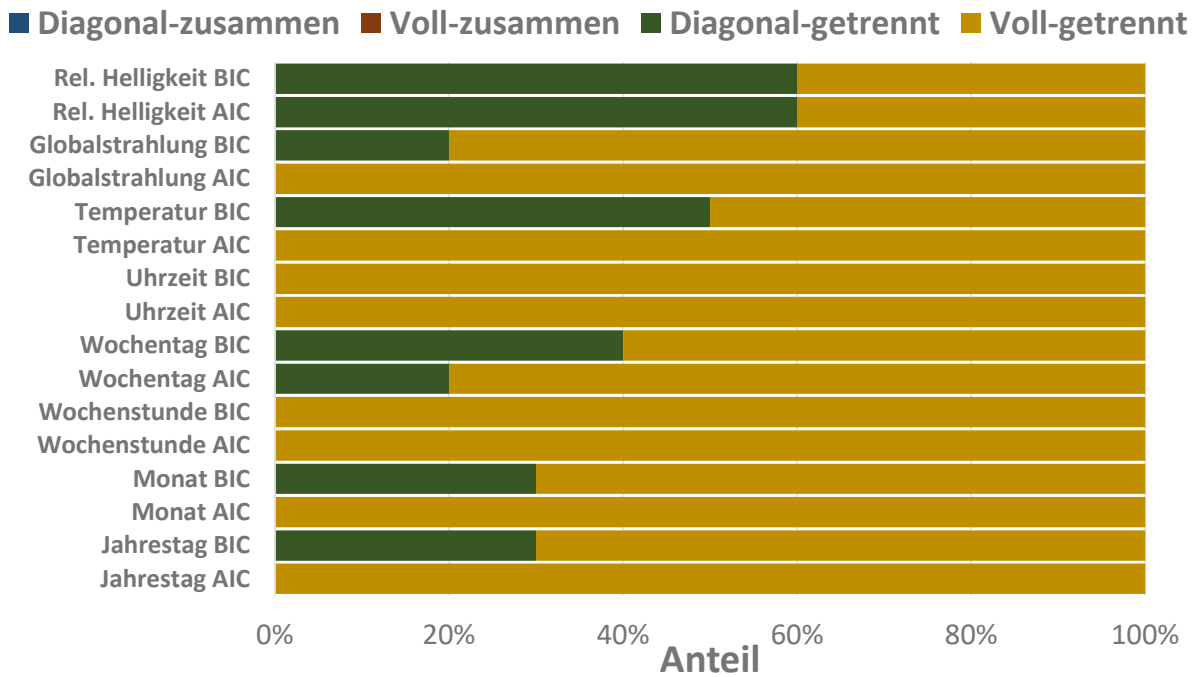


Abbildung 30: Verhältnis der bevorzugten Kovarianzmatrizen laut AIC und BIC bei einer Clusteranzahl von 1-10

Wie schon bei der Auswertung zwischen Temperatur und Transformatorauslastung, schneiden die Konfigurationen mit unterschiedlicher Ausführung der Kovarianzmatrix besser ab, sprich weisen einen höheren Anteil in Abbildung 30 auf. Über alle Eigenschaften betrachtet, schneidet die Konfiguration „Voll-getrennt“ am besten ab, weshalb bei den nachfolgenden Auswertungen die Kovarianzmatrix dementsprechend konfiguriert wurde. Ein Vergleich zwischen AIC und BIC zeigt weiters, dass das Kriterium AIC stabiler mit der Konfiguration „Voll-getrennt“ arbeitet als das Kriterium BIC. Daher wird auch nachfolgend nur mehr mit dem AIC bewertet. Soweit nicht anders angegeben, besitzt nun die Kovarianzmatrix immer eine volle-getrennte Konfiguration.

Distanzmatrix-Algorithmus

Bei der Konfiguration des Distanzmatrix-Algorithmus wurde evaluiert, ob die Gewichtungsfaktoren, welche für die Berechnung der Distanzmatrix verwendet werden, einen positiven Einfluss auf das Clustering haben und daher berücksichtigt werden sollen. Die Gewichtungsfaktoren ordnen dabei jeder der 5 Eigenschaften einen Prozentwert zu, der den Einfluss auf die Transformatorauslastung widerspiegeln und dementsprechend bei der Berechnung der Distanzmatrix miteinfließen soll. Für den untersuchten Transformator *MV2.101-LV1.203-Trafo 1* ergaben sich folgende Faktoren laut Abbildung 31.

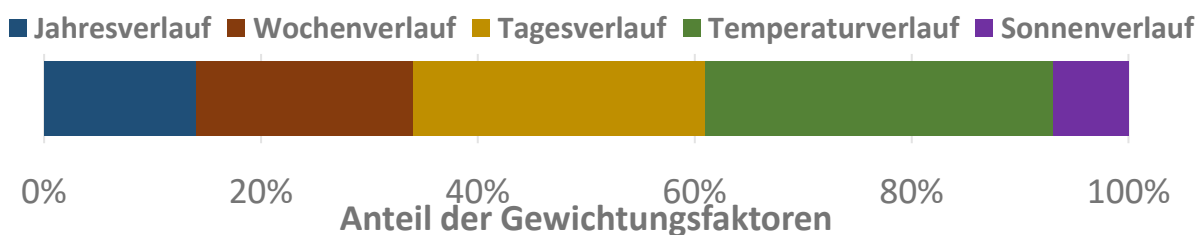


Abbildung 31: Gewichtungsfaktoren der Transformatorstation *MV2.101-LV1.203-Trafo 1*

Dabei besitzt der Jahresverlauf eine Gewichtung von 14%, der Wochenverlauf 20%, der Tagesverlauf 27%, der Temperaturverlauf 32% und der Sonnenverlauf 7%. Ob und wie sehr sich diese Faktoren auf das Clustering auswirken, wurde mit Hilfe des Silhouettenkoeffizienten bewertet. Der Distanzmatrix-Algorithmus wurde mit und ohne Gewichtungsfaktoren bei einer Clusteranzahl von 2-10 auf den Datensatz angewendet und die Ergebnisse gegenübergestellt (Clusteranzahl 1 ist beim Silhouettenkoeffizienten undefiniert). Ohne Gewichtungsfaktor bedeutet in diesem Fall, dass bei der Berechnung der Distanzmatrix in Formel (25) (siehe Punkt 3.5.2) der Gewichtungsfaktor auf 1 gesetzt wurde und daher alle Eigenschaften gleichgewichtet sind. Abbildung 32 zeigt beispielsweise den Silhouettenkoeffizient beim Clustering mit und ohne Gewichtungsfaktor, bei einer Zusammensetzung des Input-Datensatzes „Kontinuierlich 1“. Die Erläuterung zu dieser Zusammensetzung des Input-Datensatzes folgt noch später (siehe Auswahl der Eigenschaften aus dem Input-Datensatz). Hier wurde nur das Augenmerk auf die Bewertung des Gewichtungsfaktors gelegt.

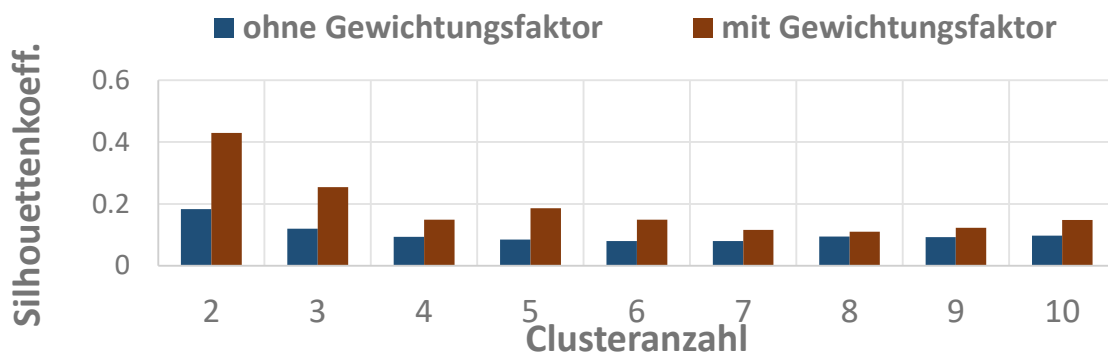


Abbildung 32: Vergleich des Clusterings mit und ohne Gewichtungsfaktor beim Distanzmatrix-Algorithmus „Kontinuierlich 1“

Wie gut zu erkennen ist, zeigt die Variante mit Gewichtungsfaktor stets bessere Ergebnisse (sprich einen höheren Silhouettenkoeffizienten) als die Variante ohne Gewichtungsfaktor. In Tabelle 15 sind zusätzlich die Maximalwerte eingetragen, die sich bei allen 4 Zusammensetzungen des Input-Datensatzes ergaben. Bei allen Varianten ist ein positiver Effekt der Gewichtungsfaktoren zu erkennen, weil der Silhouettenkoeffizient stets größer ist.

Tabelle 15: Maximalwerte des Silhouettenkoeffizient mit und ohne Gewichtungsfaktoren beim Distanzmatrix-Algorithmus

	Kontinuierlich 1	Kontinuierlich 2	Diskret 1	Diskret 2
Ohne Gewichtungsfaktor	0,18	0,23	0,16	0,18
Mit Gewichtungsfaktor	0,43(+135%)	0,38 (+69%)	0,40 (+151%)	0,38 (+110%)

Soweit im Nachfolgenden nicht anders angegeben, wird daher der Distanzmatrix-Algorithmus immer mit Berücksichtigung der Gewichtungsfaktoren verwendet.

Optimale Clusteranzahl

Die Kenntnis über die optimale Anzahl von Clustern im Datensatz ist ein wesentlicher Punkt, weil mit Hilfe der eingeteilten Daten in genau diese einzelnen Cluster die anschließende Prognose von Belastungsszenarien getätigt wird. Daher müssen die Daten möglichst gut

geclustert werden und Ähnlichkeiten in den Daten kombiniert und Unähnlichkeiten getrennt werden. Je nach Anzahl der Cluster unterscheidet sich klarerweise auch die Qualität bzw. Güte dieser Einteilung.

EM-Algorithmus

Die Güte des EM-Algorithmus wurde, wie im vorhergehenden Punkt, anhand des AICs bewertet. Da für die optimale Clusteranzahl der tatsächliche AIC-Wert nicht ausschlaggebend ist, sondern das relative Verhältnis zwischen den unterschiedlichen Varianten zählt (in diesem Fall bedeutet Variante eine unterschiedliche Clusteranzahl), wurden für die folgenden Auswertungen das AIC skaliert. Durch diese Skalierung befinden sich die Werte zwischen 0 – 1 und es kann der relative Unterschied zwischen den Varianten besser entnommen werden. Abbildung 33 zeigt den skalierten und relativen Verlauf des AICs beim Clustering jeder Eigenschaft jeweils paarweise mit der Transformatorauslastung (sprich ein 2-dimensionaler Input-Datensatz).

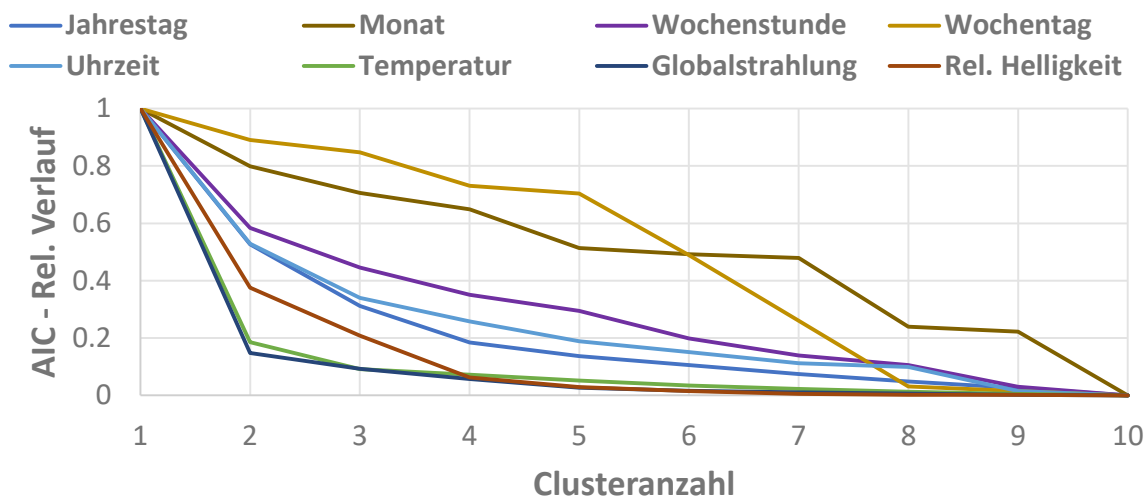


Abbildung 33: Identifikation der optimalen Anzahl an Clustern beim EM-Algorithmus

Es kann sehr leicht festgestellt werden, dass mit steigender Anzahl an Clustern auch die Güte des Clusterings steigt, da das AIC bis auf einzelne Ausnahmen stetig sinkt. Um auf keine zu große Anzahl an Clustern zu kommen, was wiederum zu einem höheren Rechenaufwand beim Clustering und der anschließenden Prognose führen würde, wurde ein Kompromiss zwischen der Güte des Clusterings und dem Rechenaufwand getroffen. Es ist etwa aus Abbildung 33 gut zu erkennen, dass die Güte zunächst stärker zunimmt (AIC sinkt stark) und ab einem Punkt nur mehr in einem geringeren Maße zunimmt (AIC sinkt wenig). Mit Hilfe der Ellbogen- oder Knie-Methode wurde somit ein Bereich definiert, ab dem die Erhöhung der Clusteranzahl nur mehr eine geringe Verbesserung der Güte bringt. Aus diesem festgelegten Bereich wird die optimale Clusteranzahl schließlich ausgewählt. Beispielsweise wäre der optimale Bereich beim Clustering von der Eigenschaft Monat mit der Transformatorauslastung (im Diagramm mit der Farbe Braun) zwischen 2 und 4 Clustern.

Mit dieser Methode zur Identifikation der optimalen Clusteranzahl wurde nachfolgend beim eigentlichen Clustering des Input-Datensatzes die Clusteranzahl angenommen.

Distanzmatrix-Algorithmus

Die optimale Clusteranzahl beim Distanzmatrix-Algorithmus wurde mit Hilfe des Silhouettenkoeffizienten ermittelt. Für die 4 Kombinationsvarianten der Eigenschaften ergaben sich folgende Werte der Koeffizienten:

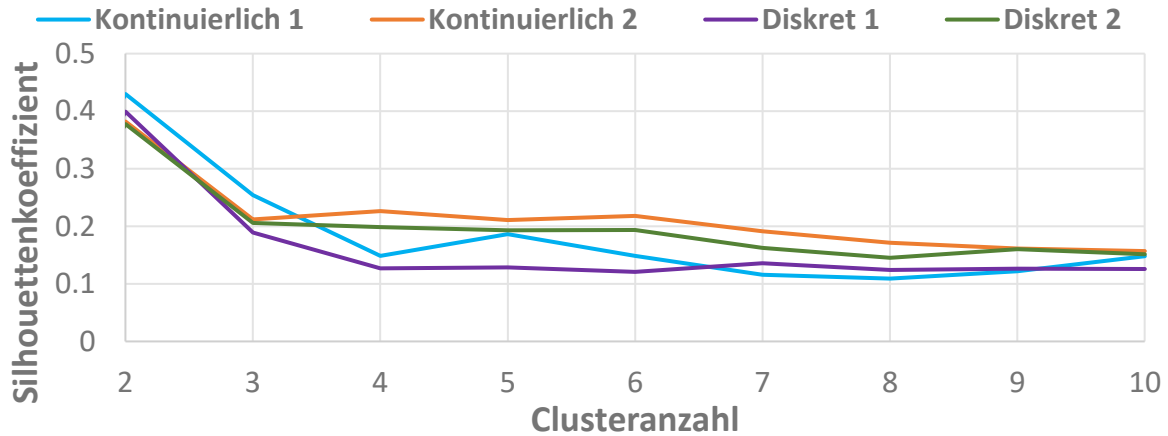


Abbildung 34: Identifikation der optimalen Anzahl an Clustern beim Distanzmatrix-Algorithmus

Wie aus Abbildung 34 zu entnehmen ist, zeigt der Silhouettenkoeffizient mit steigender Clusteranzahl eine sinkende Güte beim Clustering, was bedeutet, dass eine niedrigere Clusteranzahl zu bevorzugen ist. Durch diesen Sachverhalt wurde nachfolgend beim eigentlichen Clustering des Input-Datensatzes die Clusteranzahl jeweils zwischen 2-4 angenommen.

Auswahl der Eigenschaften aus dem Input-Datensatz

Für die Eigenschaften Jahresverlauf, Wochenverlauf und Sonnenverlauf existieren im Input-Datensatz zwei Darstellungsformen. Um herauszufinden, welche Darstellungsform am geeignetsten ist, wurden mehrere Kombinationen bzw. Zusammensetzungen der Eigenschaften untersucht und anschließend die Qualität der Clustering-Ergebnisse bewertet. Zusätzlich wurde auch bewertet, ob alle Eigenschaften einen Mehrwert für das Clustering-Ergebnis bringen, oder ob gewisse Eigenschaften entfallen können und der Input-Datensatz damit reduziert werden kann. Folgende Kombinationen von Eigenschaften in dem Input-Datensatz wurden untersucht:

Tabelle 16: Kombinationen der Eigenschaften im Input-Datensatz für die erste Clusteranalyse mittels EM-Clustering

		Kontinuierlich 1	Kontinuierlich 2	Diskret 1	Diskret 2
Jahresverlauf	Jahrestag	X	X		
	Monat			X	X
Wochenverlauf	Wochenstunde	X	X		
	Wochentag			X	X
Tagesverlauf	Uhrzeit	X	X	X	X
Temperaturverlauf	Temperatur	X	X	X	X
Sonnenverlauf	Globalstrahlung	X		X	
	Rel. Helligkeit		X		X
Belastungsverlauf	Transformatorauslastung	X	X	X	X

Wie schon in Punkt 3.3 erwähnt wurde, sind nie zwei Darstellungsformen einer Eigenschaft in einer Kombination vorhanden, weshalb man die 4 Kombinationsvarianten in kontinuierlich und diskret unterteilen kann. Die Eigenschaften Monat (= diskret) und Wochenstunde (= kontinuierlich) sind somit immer in unterschiedlichen Varianten zu finden. Die Eigenschaften Uhrzeit, Temperatur und Sonnenverlauf können keiner der beiden Varianten zugeordnet werden, weshalb diese in den kontinuierlichen und diskreten Varianten zu finden sind. Die Eigenschaft Transformatorauslastung ist verständlicherweise ebenfalls in allen Varianten vertreten.

EM-Algorithmus

Abbildung 35 zeigt anhand der Ellbogen-Methode die identifizierte Clusteranzahl bei den 4 Kombinationsvarianten.

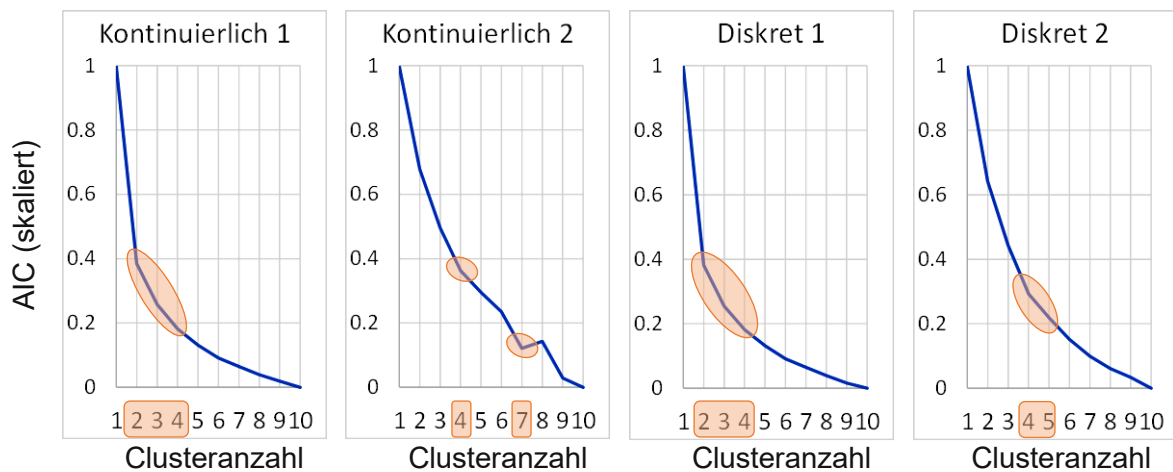


Abbildung 35: Ellbogen-Methode bei der Auswahl der Eigenschaften bei der ersten Clusteranalyse mittels EM-Algorithmus

Da kein eindeutiger Knick im Verlauf des AICs existiert, wurden ein oder mehrere Bereiche ausgewählt, bei denen die optimale Clusteranzahl angenommen wurde (orange Markierung). Mit dieser Clusteranzahl wurde anschließend eine genauere Auswertung vorgenommen.

Variante Kontinuierlich 1: Jahrestag – Wochenstunde – Uhrzeit – Temperatur – Globalstrahlung – Transformatorauslastung

Die Variante „Kontinuierlich 1“ setzt sich aus den Eigenschaften Jahrestag, Wochenstunde, Uhrzeit, Temperatur, Globalstrahlung und Transformatorauslastung zusammen. Dieser 6-dimensionale Input-Datensatz wurde mittels EM-Algorithmus geclustert und laut Abbildung 35 eine optimale Clusteranzahl zwischen 2 und 4 identifiziert. Die Aufteilung der Datenpunkte ist aus Abbildung 36 zu entnehmen.

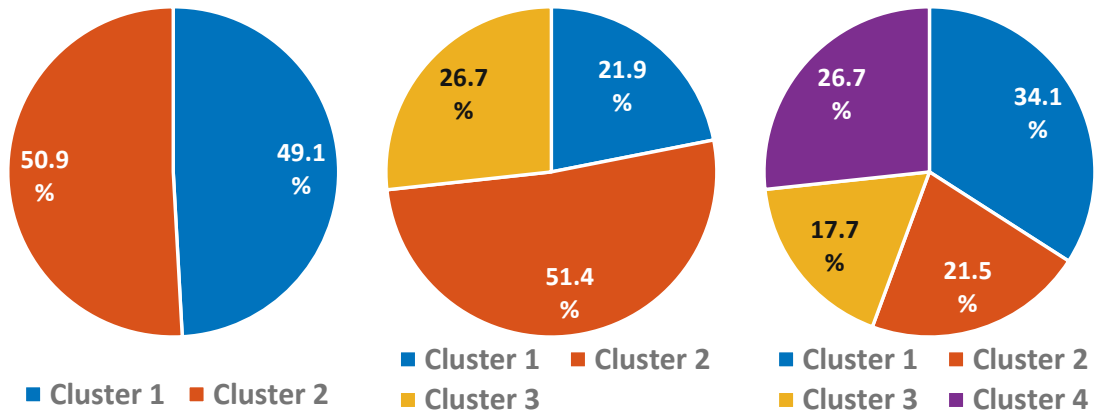


Abbildung 36: Aufteilung der Datenpunkte beim EM-Algorithmus „Kontinuierlich 1“

Bei allen Varianten existiert kein Cluster, der eine wesentlich geringere Anzahl an Datenpunkten besitzt als die anderen Cluster, sondern die Datenpunkte weisen eine brauchbare Aufteilung auf. Wie die Dichtefunktionen der Eigenschaften und Transformatorauslastung aussehen, ist in Abbildung 37 für eine Clusteranzahl von 2, in Abbildung 38 für eine Clusteranzahl von 3 und in Abbildung 39 für eine Clusteranzahl von 7 dargestellt. Die farbliche Zuordnung der Dichtefunktionen und Aufteilung der Datenpunkte ist aufeinander abgestimmt worden und kann daher in die Auswertungen miteinfließen.

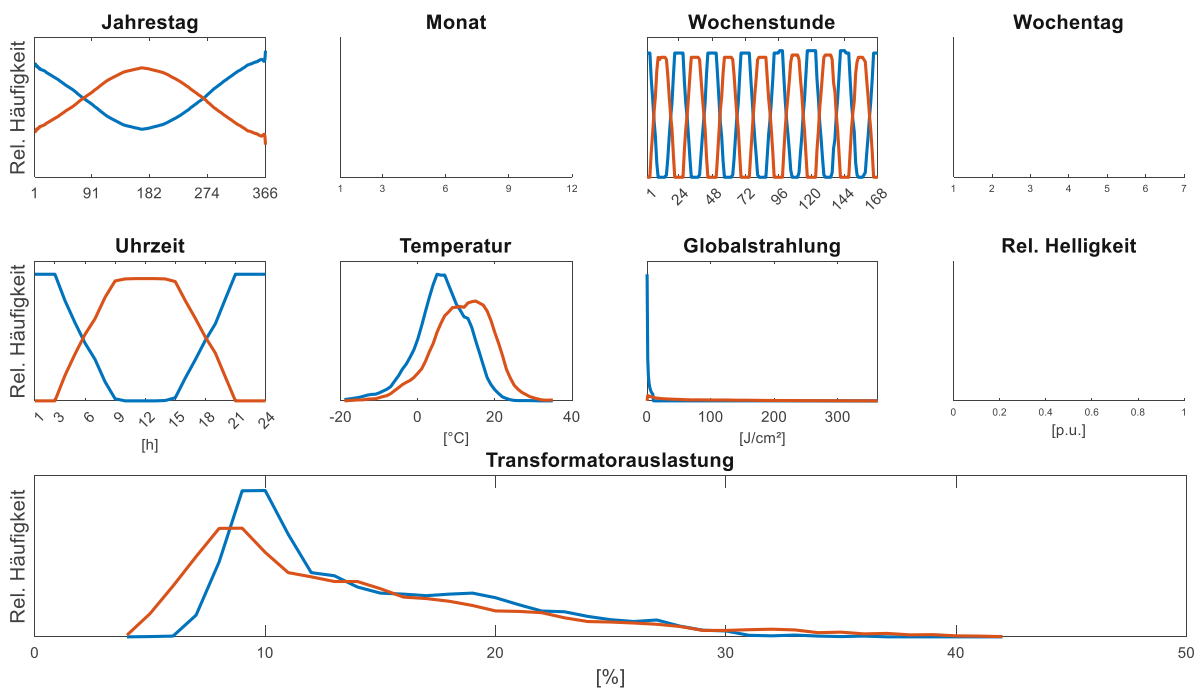


Abbildung 37: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 2

Bei der Clusteranzahl von 2 zeigt sich eine Differenzierung der Inputdaten in unterschiedliche Cluster, welche eine Charakteristik von Tag und Nacht aufweisen. Dieser Sachverhalt lässt sich anhand der Eigenschaft Wochenstunde, Uhrzeit und Globalstrahlung sehr einfach erkennen, da die Mittagsstunden und höheren Globalstrahlungswerte in den zweiten Cluster (Orange) fallen und die Nachtstunden ohne Globalstrahlung in den ersten Cluster (Blau). Auch die Eigenschaften Jahrestag und Temperatur zeigen diese Unterteilung. Der Cluster mit winterlichen Jahrestagen und tendenziell tieferen Temperaturen kann dem Nachtcluster (Blau) zugeschrieben werden und die sommerlichen Jahrestage mit tendenziell höheren Temperaturen dem Tagescluster (Orange). Die Cluster der Transformatorauslastung lassen sich bei dieser Clusteranzahl nicht eindeutig unterscheiden. Beide Cluster weisen eine höhere Wahrscheinlichkeit bei niedrigeren Auslastungswerten auf und besitzen aber auch Werte bis hin zur maximalen Auslastung.

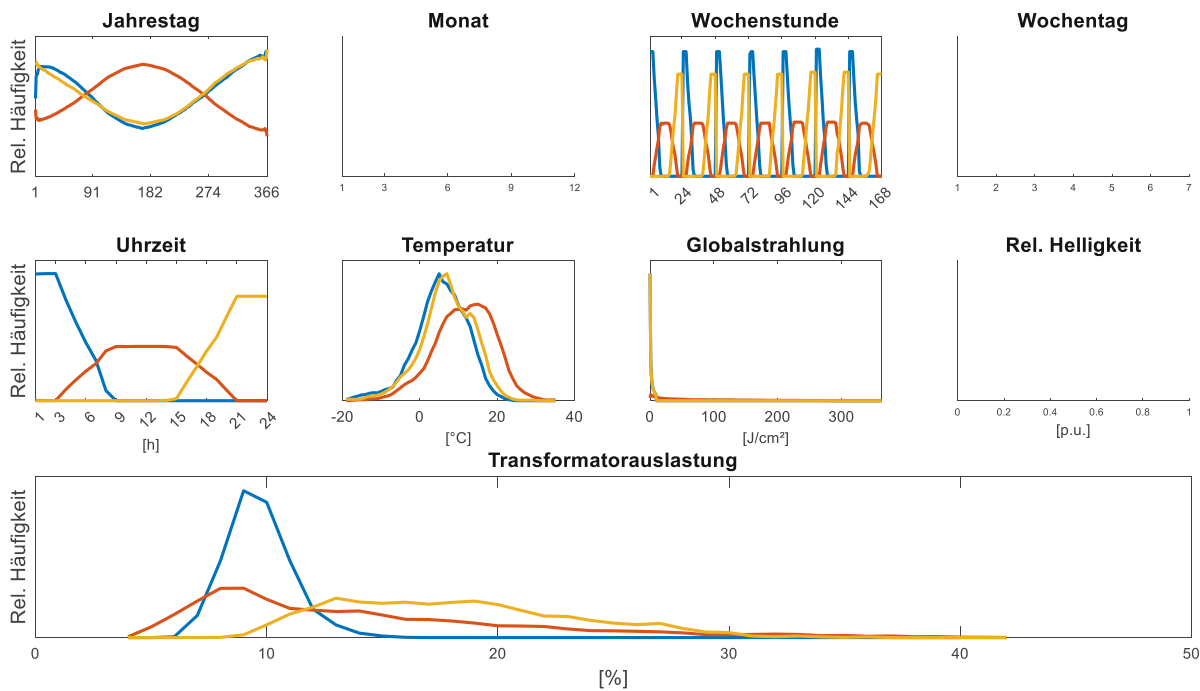


Abbildung 38: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 3

Bei der Clusteranzahl von 3 zeigt sich eine Differenzierung der Inputdaten in unterschiedliche Cluster, welche eine Charakteristik von Tag, Abend und Morgen aufweisen. Am besten lässt sich dieser Sachverhalt wieder anhand der Eigenschaft Wochenstunde, Uhrzeit und Globalstrahlung erkennen, da die Mittagsstunden und die höheren Globalstrahlungswerte in den orangenen Cluster fallen und die Abend- bzw. Morgenstunden ohne Globalstrahlung in die beiden anderen Cluster. Die Abend- und Morgen-Cluster (Blau und Gelb) zeigen beim Jahrestag und der Temperatur einen ähnlichen Verlauf, was auf die Verbindung zu Zeitpunkten ohne Sonneneinstrahlung (entspricht kälteren Temperaturen und Jahreszeiten) hinweist. Die Transformatorauslastung zeigt bei dieser Clusteranzahl eine bessere Unterteilung als bei der Clusteranzahl 2. Der Morgen-Cluster (Blau) zeigt niedrigere Auslastungswerte zwischen 6-15%, der Tag-Cluster (Orange) zeigt Auslastungswerte über den ganzen Bereich verteilt und der Abend-Cluster (Gelb) zeigt tendenziell höhere Werte ab 9%.

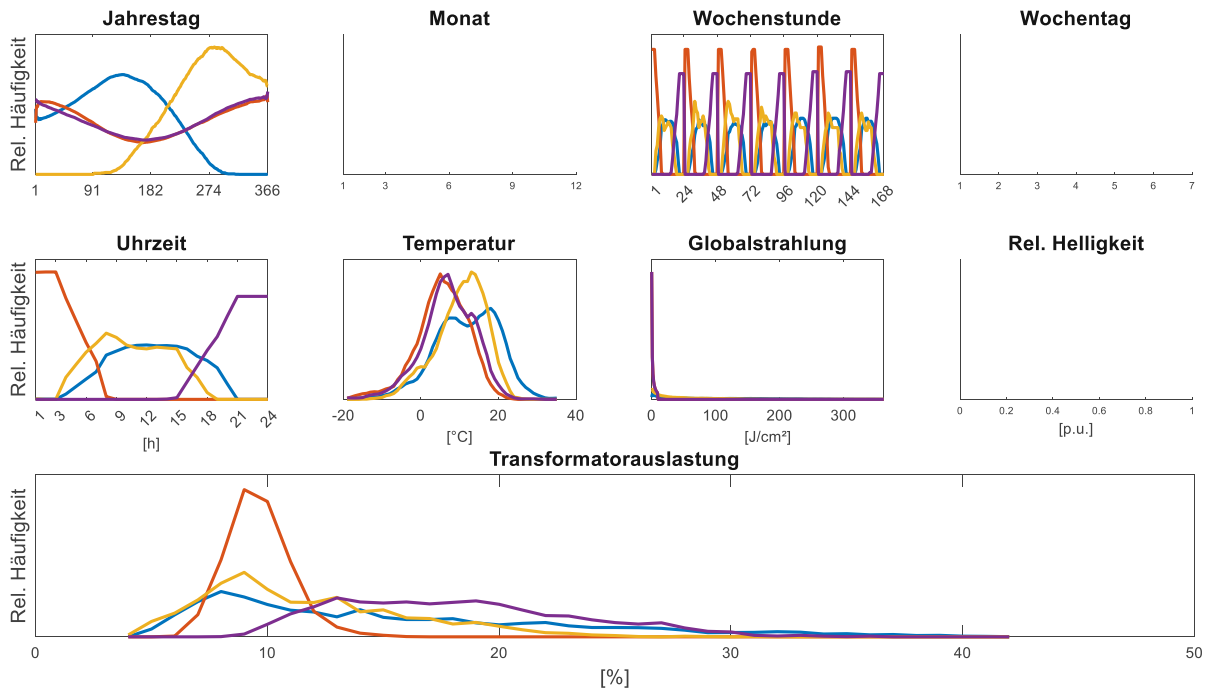


Abbildung 39: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 4

Bei der Clusteranzahl von 4 zeigt sich eine Differenzierung der Inputdaten in unterschiedliche Cluster, welche eine Charakteristik von Tag, Abend und Morgen aufweisen. Die Aufteilung ähnelt sehr dem Clustering mit einer Anzahl von 3 Clustern nur mit dem Unterschied, dass zwei Tag-Cluster (Blau und Gelb) existieren. Diese Tag-Cluster ähneln sich im Bereich Wochenstunde, Uhrzeit und Transformatorauslastung und unterscheiden sich lediglich beim Jahrestag und der Temperatur.

Variante Kontinuierlich 2: Jahrestag – Wochenstunde – Uhrzeit – Temperatur – Relative Helligkeit – Transformatorauslastung

Die Variante „Kontinuierlich 2“ setzt sich aus den Eigenschaften Jahrestag, Wochenstunde, Uhrzeit, Temperatur, relative Helligkeit und Transformatorauslastung zusammen. Die optimale Anzahl von Clustern wurde bei 4 bzw. 7 identifiziert. Die Aufteilung der Datenpunkte ist aus Abbildung 40 zu entnehmen.

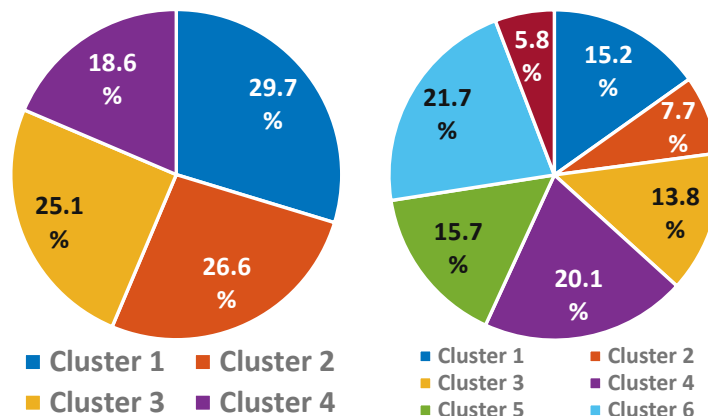


Abbildung 40: Aufteilung der Datenpunkte beim EM-Algorithmus „Kontinuierlich 2“

Beide Varianten zeigen eine brauchbare Aufteilung der Datenpunkte. Lediglich bei den Varianten mit einer Clusteranzahl von 7 existieren zwei Cluster (Orange und Rot), welche eine

geringere Anzahl an Datenpunkten aufweisen, als die restlichen Cluster. Wie die Dichtefunktionen der einzelnen Cluster und Eigenschaften aussehen, ist in Abbildung 41 für eine Clusteranzahl von 4 und in Abbildung 42 für eine Clusteranzahl von 7 dargestellt. Die farbliche Zuordnung der Dichtefunktionen und Aufteilung der Datenpunkte wurde hier wieder aufeinander abgestimmt.

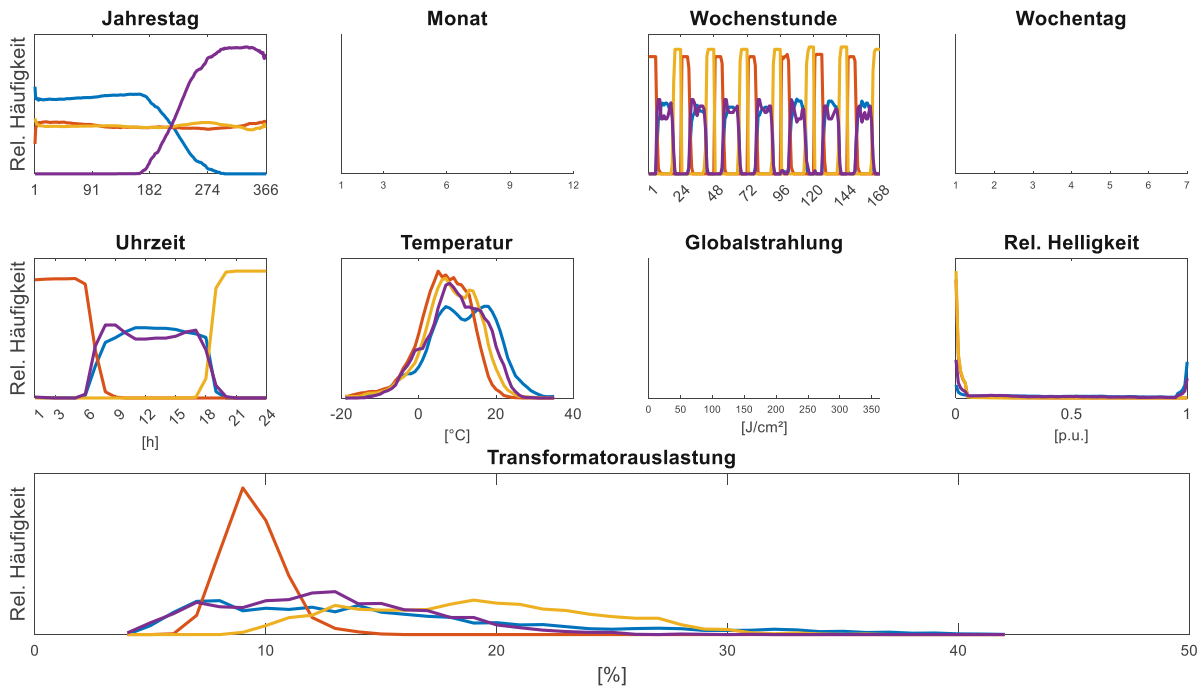


Abbildung 41: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 2“ mit einer Clusteranzahl 4

Bei der Clusteranzahl von 4 zeigt sich eine Differenzierung der Inputdaten in Cluster mit einer Charakteristik Tag, Abend und Morgen. Im Vergleich mit der Variante „Kontinuierlich 1“ mit 4 Clustern, zeigt sich eine ähnliche Aufteilung und Verteilung. Der Tag-Cluster ist wieder zweimal vorhanden und deckt sich bis auf die Eigenschaft Jahrestag in allen Bereichen. Es wird daher auf eine detailliertere Beschreibung verzichtet und erst am Ende dieses Unterkapitels beim Variantenvergleich darauf eingegangen.

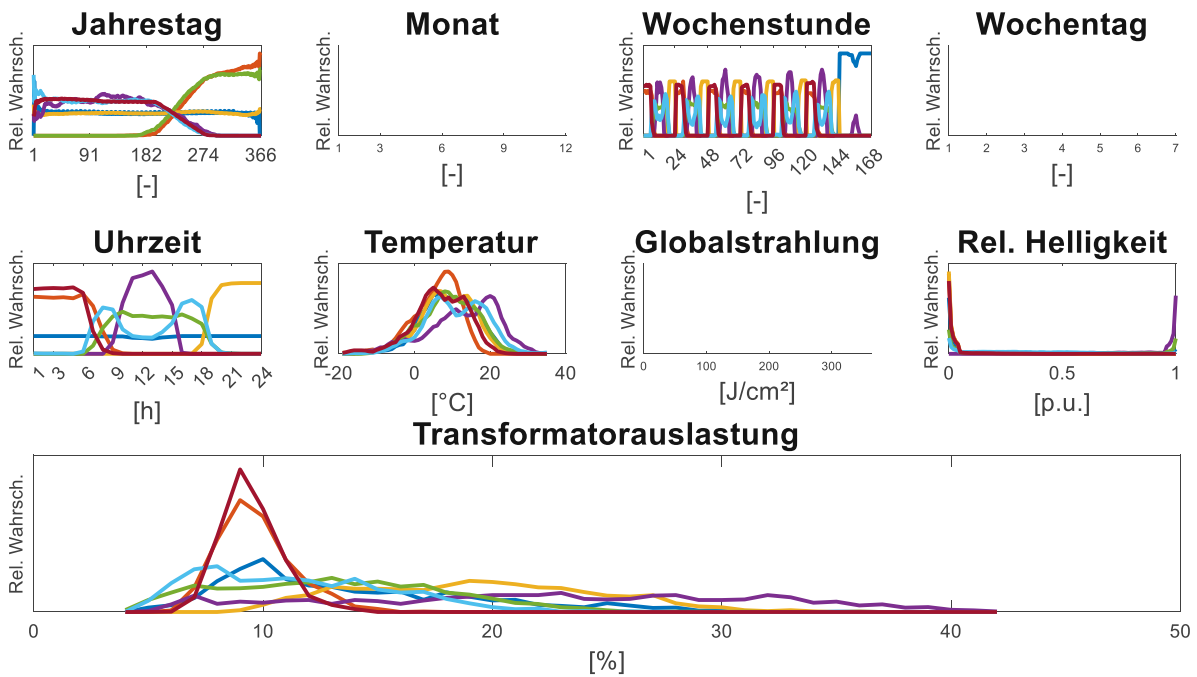


Abbildung 42: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 2“ mit einer Clusteranzahl 7

Bei der Clusteranzahl von 7 lässt sich eine Differenzierung zwischen den Clustern schwieriger durchführen, da meist keine eindeutige Unterscheidung vorhanden ist bzw. sich oft Cluster stark ähneln. Die deutlichste Unterscheidung ist wiederum durch die Eigenschaft Uhrzeit zu tätigen, da es Unterscheidungen zwischen Morgen, Vormittag, Mittag, Nachmittag und Abend gibt. In Bezug auf die Transformatorauslastung zeigen die beiden Morgen-Cluster (Orange und Rot) als einziges eine niedrigere Auslastungswahrscheinlichkeit im Bereich 7-12%, wohingegen die restlichen Cluster stets über einen wesentlich breiteren Bereich verteilt sind.

Variante Diskret 1: Monat – Wochentag – Uhrzeit – Temperatur – Globalstrahlung – Transformatorauslastung

Die Variante „Diskret 1“ setzt sich aus den Eigenschaften Monat, Wochentag, Uhrzeit, Temperatur, Globalstrahlung und Transformatorauslastung zusammen. Die optimale Anzahl von Clustern wurde zwischen 2 und 4 identifiziert. Die Aufteilung der Datenpunkte ist aus Abbildung 43 zu entnehmen.

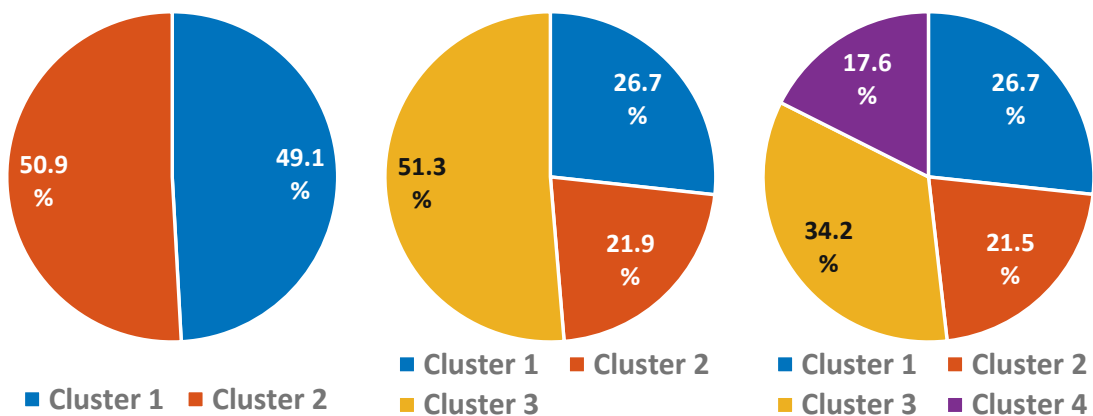


Abbildung 43: Aufteilung der Datenpunkte beim EM-Algorithmus „Diskret 1“

Ähnlich zu der Variante „Kontinuierlich 1“ existieren bei den drei Clusteranzahlen kein Cluster, die eine wesentlich geringere Anzahl an Datenpunkten besitzen. Die folgenden Abbildungen zeigen die Dichtefunktionen der Eigenschaften und Transformatorauslastung bei unterschiedlichen Clusteranzahlen. Da sich die Ergebnisse stark mit denen der Variante „Diskret 1“ decken, wird hier auf eine detailliertere Beschreibung verzichtet und erst am Ende dieses Unterkapitels beim Variantenvergleich darauf eingegangen.

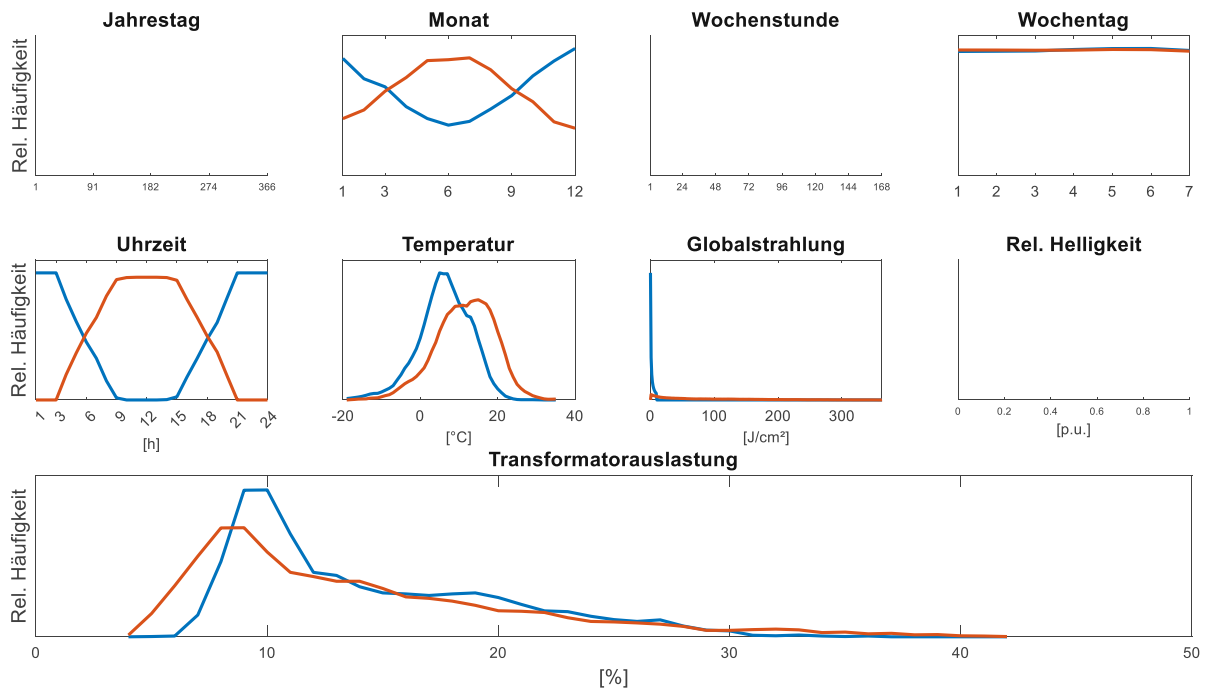


Abbildung 44: Dichtefunktionen beim EM-Algorithmus „Diskret 1“ mit einer Clusteranzahl 2

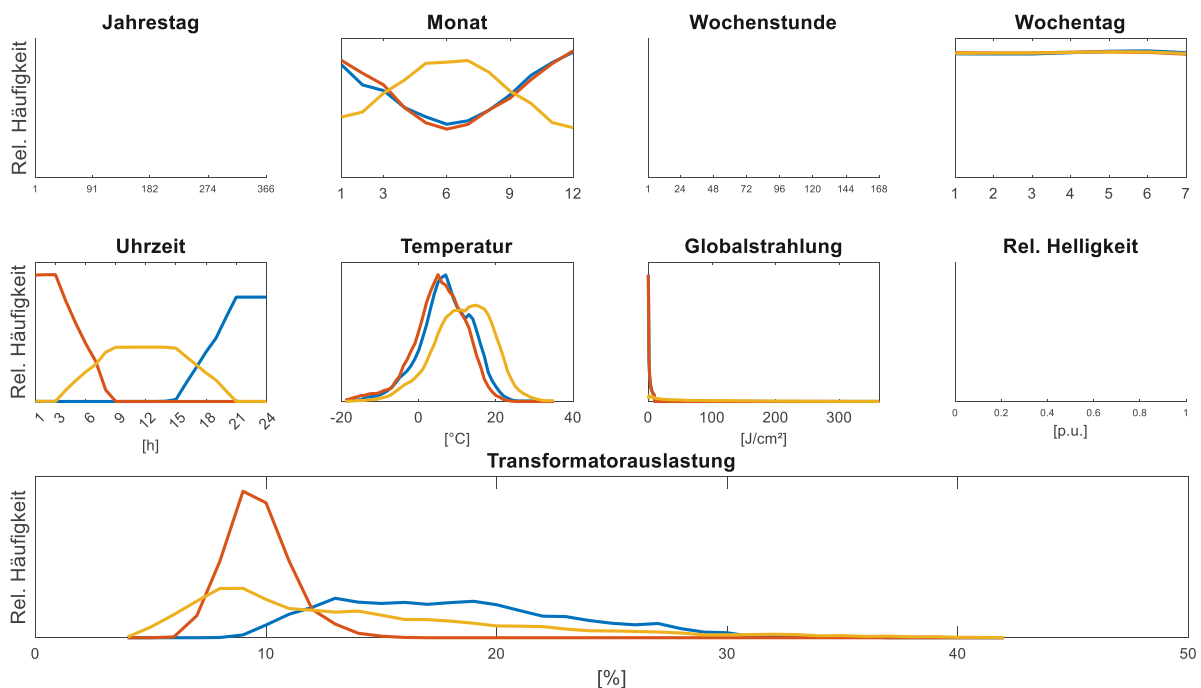


Abbildung 45: Dichtefunktionen beim EM-Algorithmus „Diskret 1“ mit einer Clusteranzahl 3

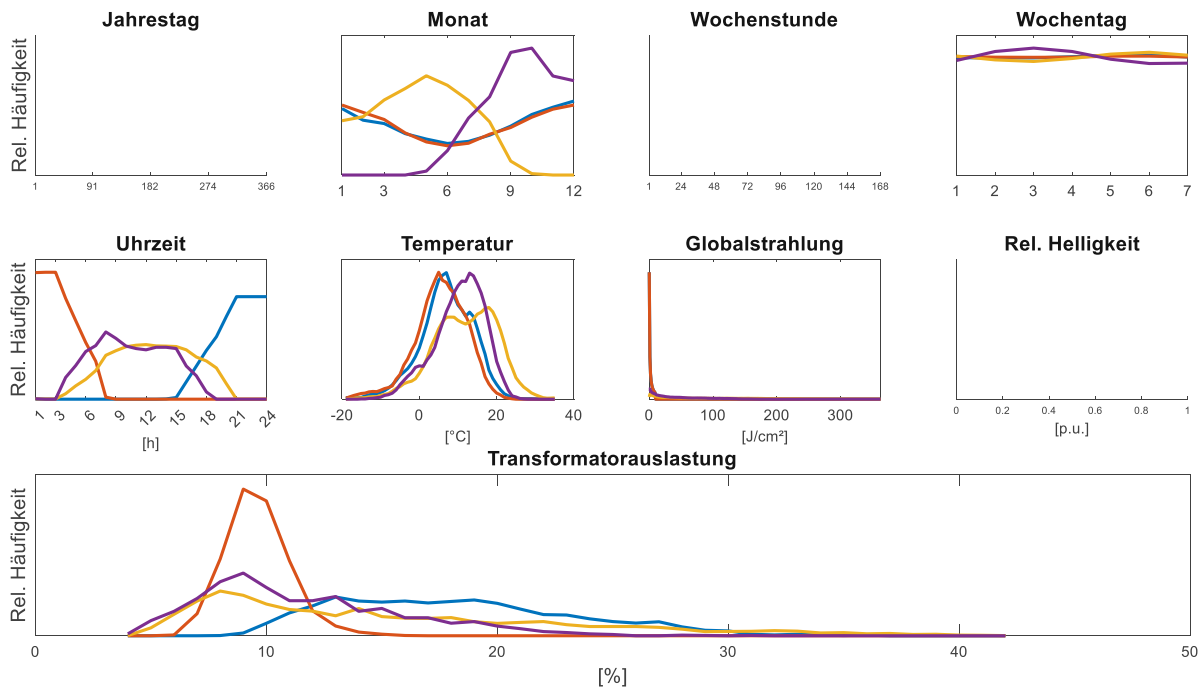


Abbildung 46: Dichtefunktionen beim EM-Algorithmus „Diskret 1“ mit einer Clusteranzahl 4

Variante Diskret 2: Monat – Wochentag – Uhrzeit – Temperatur – relative Helligkeit – Transformatorauslastung

Die Variante „Diskret 2“ setzt sich aus den Eigenschaften Monat, Wochentag, Uhrzeit, Temperatur, relative Helligkeit und Transformatorauslastung zusammen. Die optimale Anzahl von Clustern wurde bei 4 bzw. 5 identifiziert. Die Aufteilung der Datenpunkte ist aus Abbildung 47 zu entnehmen.

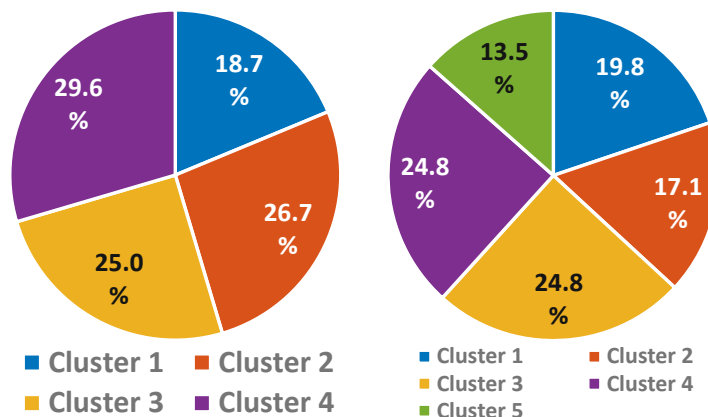


Abbildung 47: Aufteilung der Datenpunkte beim EM-Algorithmus „Diskret 2“

Wie bei den vorhergehenden Varianten, zeigen alle Cluster eine brauchbare Aufteilung der Datenpunkte. Die folgenden Abbildungen zeigen die Dichtefunktionen der Eigenschaften und Transformatorauslastung bei unterschiedlichen Clusteranzahlen. Da sich die Ergebnisse mit denen der Variante „Diskret 2“ decken, wird hier auf eine detailliertere Beschreibung verzichtet und erst am Ende dieses Unterkapitels beim Variantenvergleich darauf eingegangen.

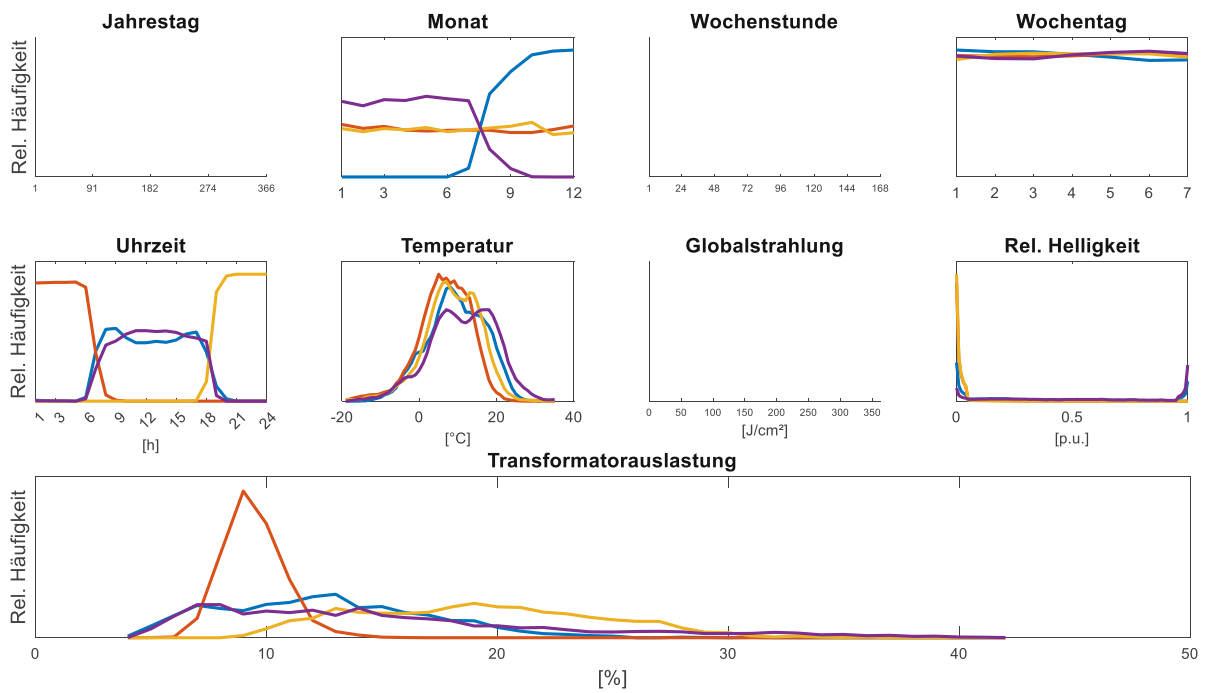


Abbildung 48: Dichtefunktionen beim EM-Algorithmus „Diskret 2“ mit einer Clusteranzahl 4

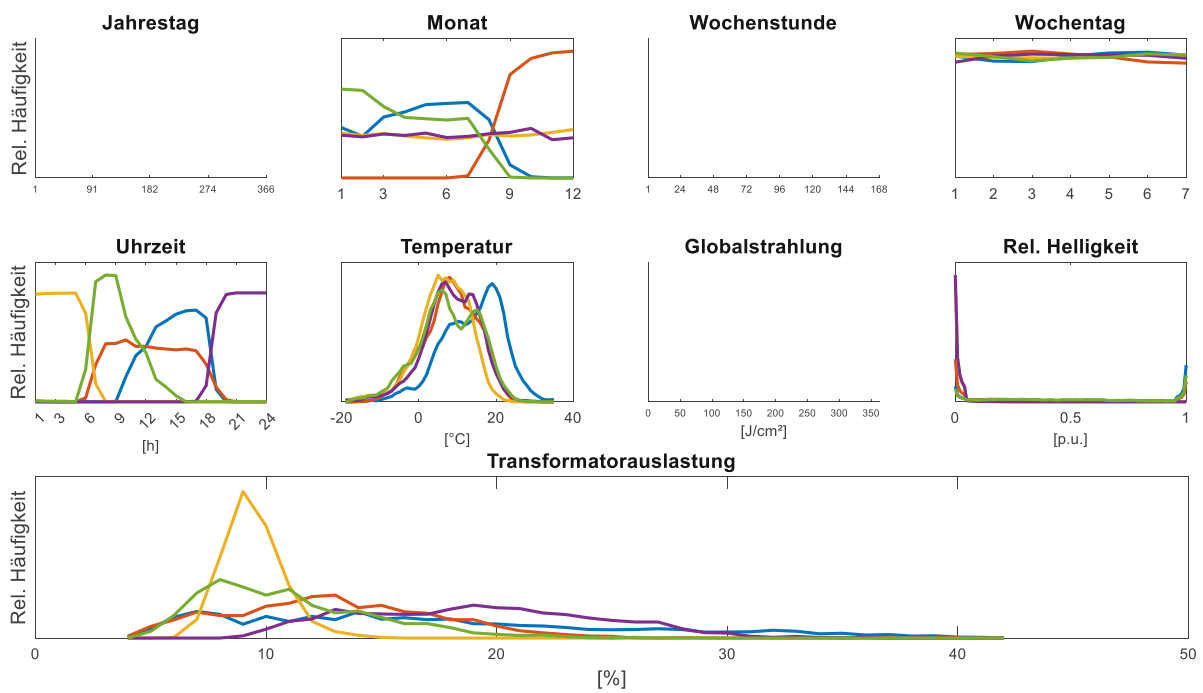


Abbildung 49: Dichtefunktionen beim EM-Algorithmus „Diskret 2“ mit einer Clusteranzahl 5

Distanzmatrix-Algorithmus

Ähnlich zu der Auswertung des EM-Algorithmus wurde auch beim Distanzmatrix-Algorithmus die Auswahl der Eigenschaften aus dem Input-Datensatz untersucht. Da sich die Clusterergebnisse der vier Varianten weitestgehend decken, wird nur die erste Variante „Kontinuierlich 1“ detaillierter beschrieben und die anderen Varianten erst am Ende dieses Unterkapitels bei der Gegenüberstellung miteinbezogen.

Variante Kontinuierlich 1: Jahrestag – Wochenstunde – Uhrzeit – Temperatur – Globalstrahlung – Transformatorauslastung

Die Variante „Kontinuierlich 1“ setzt sich aus den Eigenschaften Jahrestag, Wochenstunde, Uhrzeit, Temperatur, Globalstrahlung und Transformatorauslastung zusammen. Die optimale Clusteranzahl wurde zwischen von 2 und 3 evaluiert. Die Aufteilung der Datenpunkte ist aus Abbildung 50 zu entnehmen.

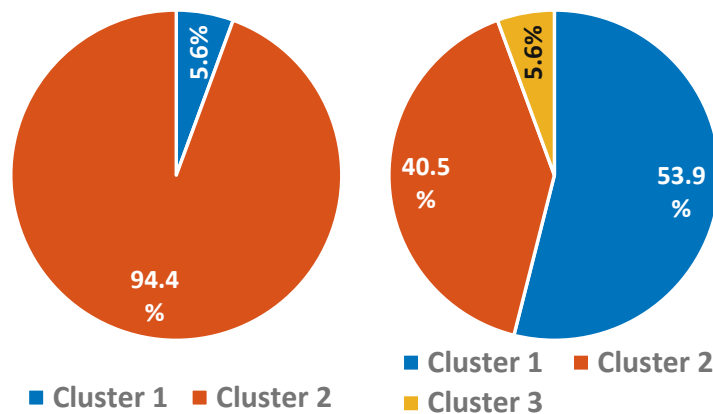


Abbildung 50: Aufteilung der Datenpunkte beim Distanzmatrix-Algorithmus „Kontinuierlich 1“

Bei beiden Clusteranzahlen zeigt sich, dass es ein oder mehrere dominante Cluster mit einer großen Anzahl an Datenpunkten gibt und einen schwachen Cluster. Die Dichtefunktionen der Eigenschaften und Transformatorauslastung sind in den nachfolgenden Abbildungen dargestellt. Die farbliche Zuordnung kann wieder zwischen der Datenpunktaufteilung und den Dichtefunktionen mitberücksichtigt werden.

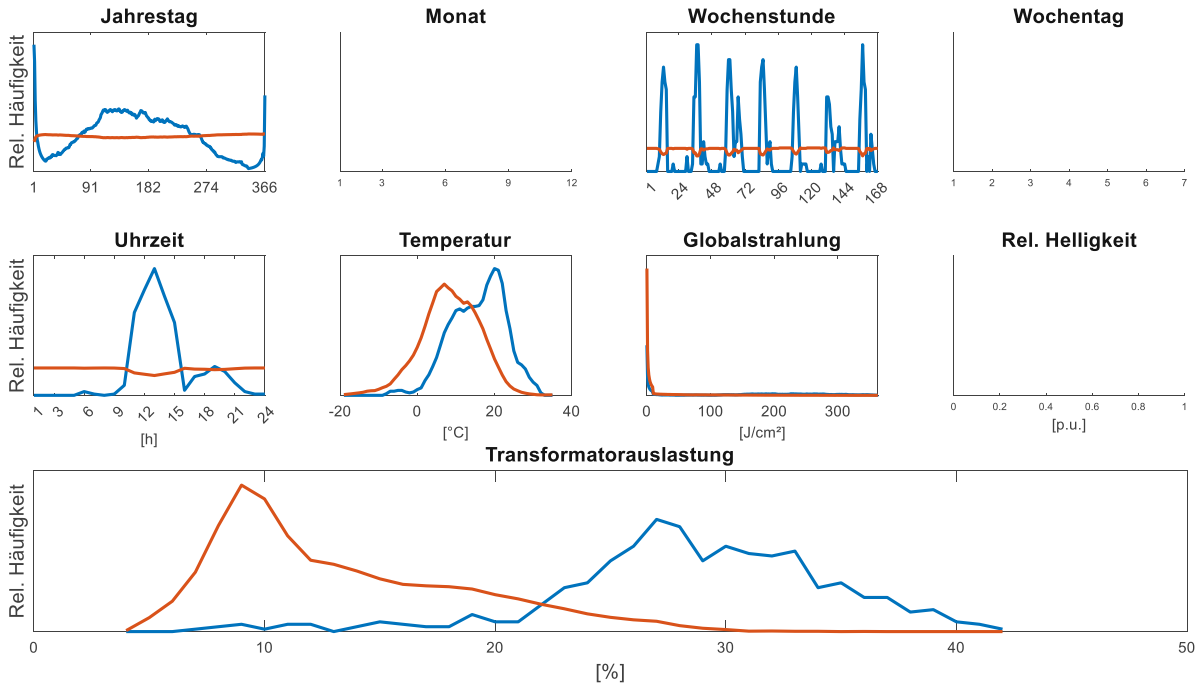


Abbildung 51: Dichtefunktionen beim Distanzmatrix-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 2

Bei der Clusteranzahl von 2 zeigt der dominante orange Cluster keine Unterscheidbarkeit innerhalb der Eigenschaften oder der Transformatorauslastung. Die Dichtefunktion ähnelt stark der Referenzverteilung, wenn alle Datenpunkte in einen Cluster gegeben werden (siehe Abbildung 28). Der schwache blaue Cluster besitzt hingegen vermehrt Datenpunkte, welche zur Mittagszeit und bei höheren Temperaturen auftreten. Bei diesen Belastungsszenarien treten durchwegs höhere Auslastungswerte am Transformator auf.

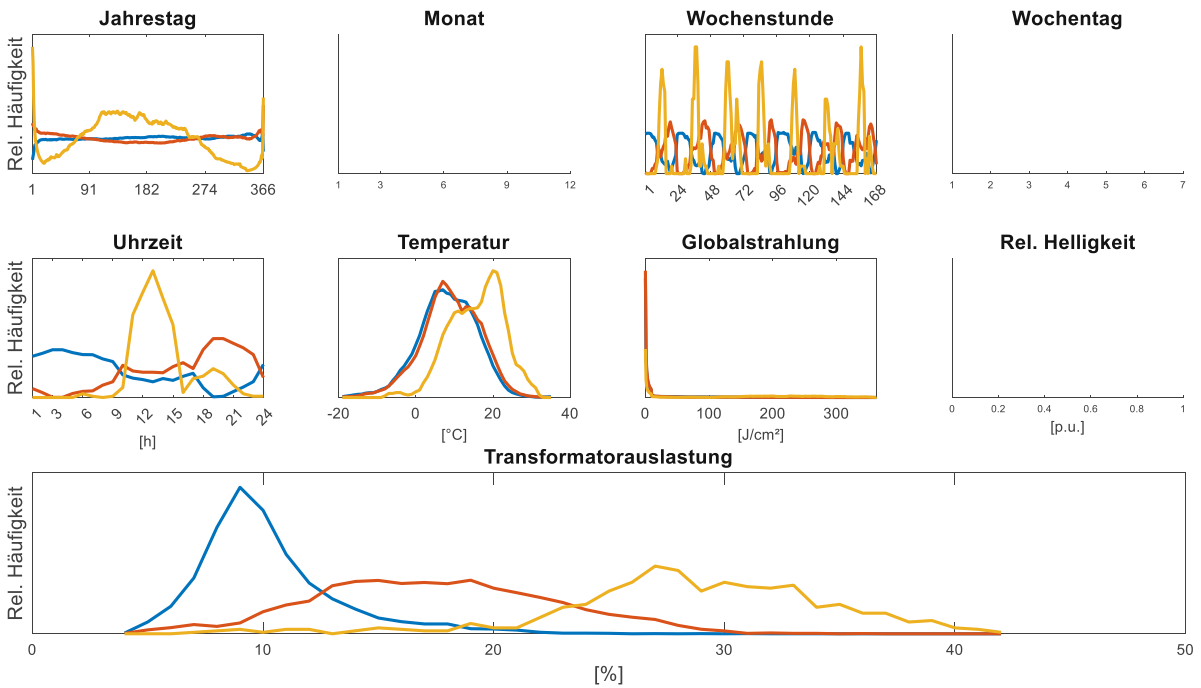


Abbildung 52: Dichtefunktionen beim Distanzmatrix-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 3

Bei der Clusteranzahl von 3 zeigt sich ein ähnliches Ergebnis wie bei der Clusteranzahl von 2. Die beiden dominanten Cluster (Blau und Orange) können nicht eindeutig einer Charakteristik zugeordnet werden. Tendenziell zeigt sich trotzdem ein Morgen-Cluster (Blau) und ein Abend-Cluster (Orange). Der schwache Cluster (hier in Gelb) besitzt wieder vermehrt Datenpunkte, welche zur Mittagszeit und bei höheren Temperaturen auftreten und eine höhere Transformatorauslastung hervorrufen.

Variante Kontinuierlich 2: Jahrestag – Wochenstunde – Uhrzeit – Temperatur – Relative Helligkeit – Transformatorauslastung

Die Variante „Kontinuierlich 2“ mit den Eigenschaften Jahrestag, Wochenstunde, Uhrzeit, Temperatur, relative Helligkeit und Transformatorauslastung besitzt folgende Aufteilung der Datenpunkte (Abbildung 53) und die dazugehörigen Dichtefunktionen (Abbildung 54 und Abbildung 55). Die Ergebnisse decken sich durchwegs mit denen der Variante „Kontinuierlich 1“, weshalb auf eine detailliertere Beschreibung verzichtet und erst am Ende des Unterkapitels beim Variantenvergleich darauf eingegangen wird.

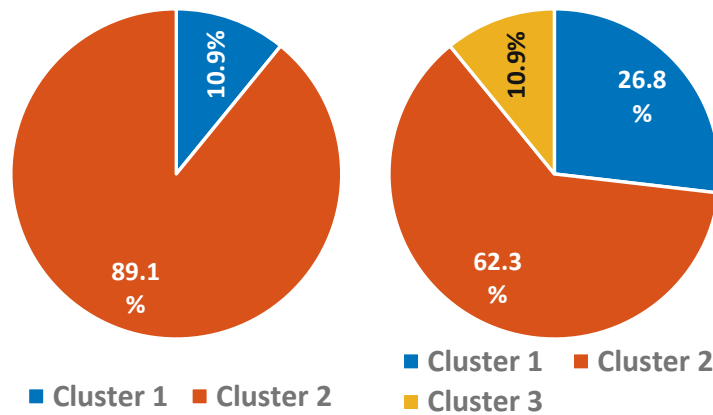


Abbildung 53: Aufteilung der Datenpunkte beim Distanzmatrix-Algorithmus „Kontinuierlich 2“

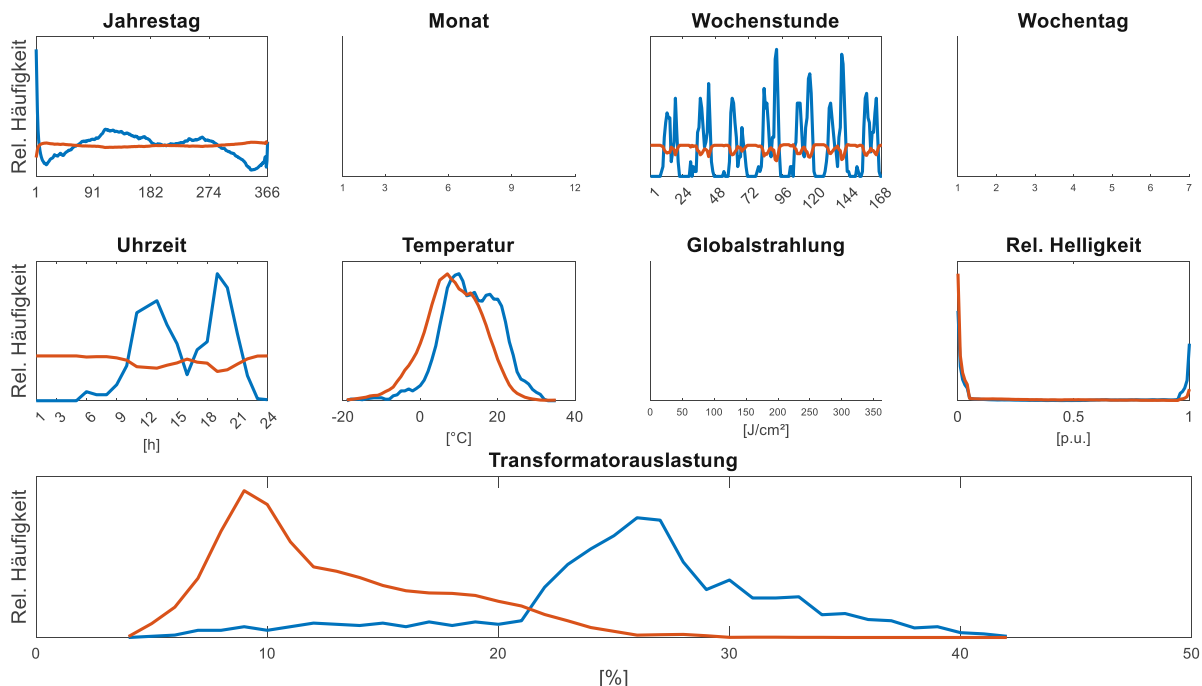


Abbildung 54: Dichtefunktionen beim Distanzmatrix-Algorithmus „Kontinuierlich 2“ mit einer Clusteranzahl 2

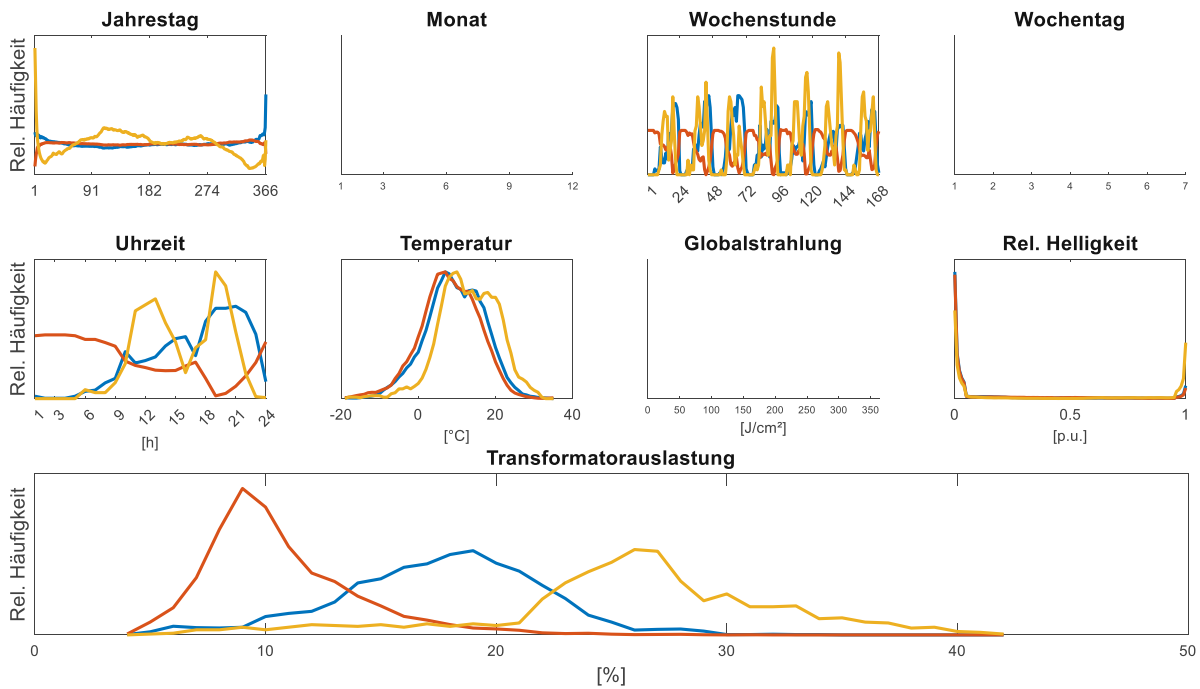


Abbildung 55: Dichtefunktionen beim Distanzmatrix-Algorithmus „Kontinuierlich 2“ mit einer Clusteranzahl 3

Variante Diskret 1: Monat – Wochentag – Uhrzeit – Temperatur – Globalstrahlung – Transformatorauslastung

Die Variante „Diskret 1“ mit den Eigenschaften Monat, Wochentag, Uhrzeit, Temperatur, Globalstrahlung und Transformatorauslastung besitzt folgende Aufteilung der Datenpunkte (Abbildung 56) und die dazugehörigen Dichtefunktionen (Abbildung 57 und Abbildung 58). Die Ergebnisse decken sich durchwegs mit denen der Variante „Kontinuierlich 1“, weshalb auf eine detailliertere Beschreibung verzichtet und erst am Ende des Unterkapitels beim Variantenvergleich darauf eingegangen wird.

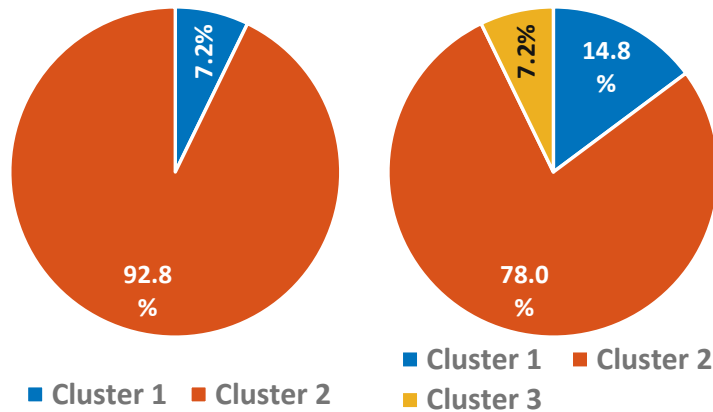


Abbildung 56: Aufteilung der Datenpunkte beim Distanzmatrix-Algorithmus „Diskret 1“

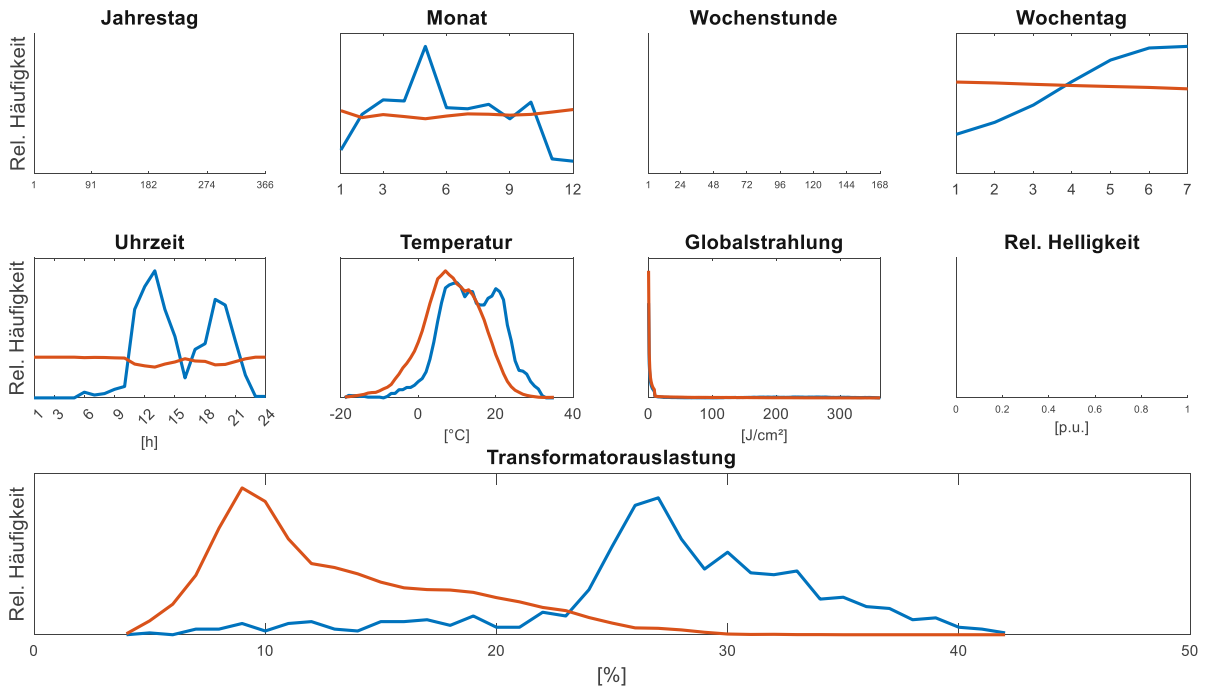


Abbildung 57: Dichtefunktionen beim Distanzmatrix-Algorithmus „Diskret 1“ mit einer Clusteranzahl 2

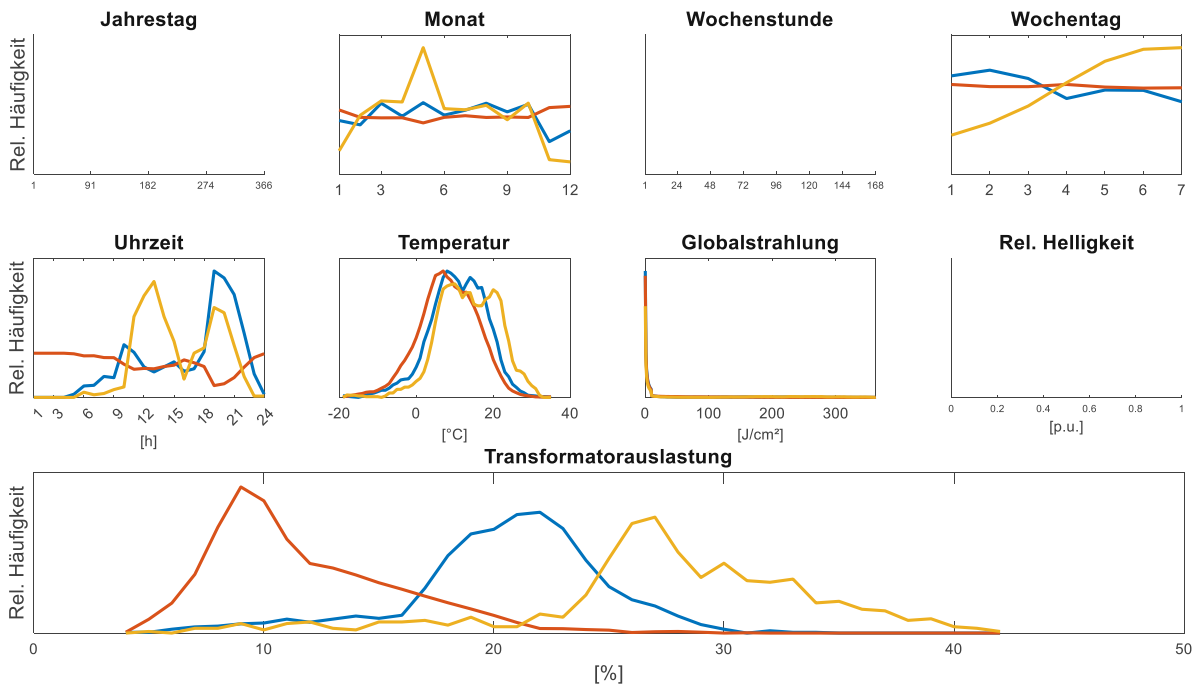


Abbildung 58: Dichtefunktionen beim Distanzmatrix-Algorithmus „Diskret 1“ mit einer Clusteranzahl 3

Variante Diskret 2: Monat – Wochentag – Uhrzeit – Temperatur – Relative Helligkeit – Transformatorauslastung

Die Variante „Diskret 2“ mit den Eigenschaften Monat, Wochentag, Uhrzeit, Temperatur, relative Helligkeit und Transformatorauslastung besitzt folgende Aufteilung der Datenpunkte (Abbildung 59) und die dazugehörigen Dichtefunktionen (Abbildung 60 und Abbildung 61). Die Ergebnisse decken sich durchwegs mit denen der Variante „Kontinuierlich 1“, weshalb auf eine detailliertere Beschreibung verzichtet und erst am Ende des Unterkapitels beim Variantenvergleich darauf eingegangen wird.

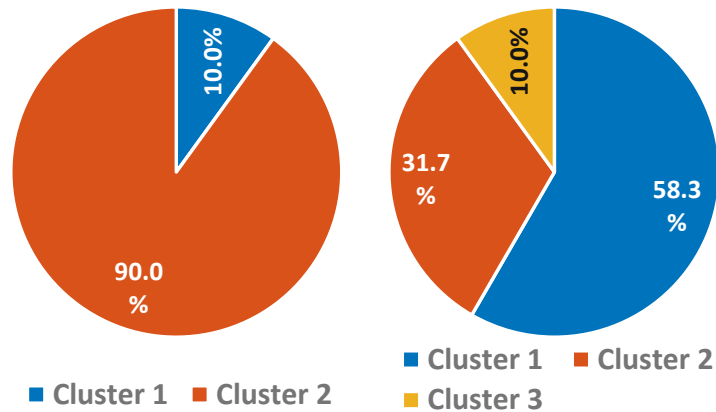


Abbildung 59: Aufteilung der Datenpunkte beim Distanzmatrix-Algorithmus „Diskret 2“

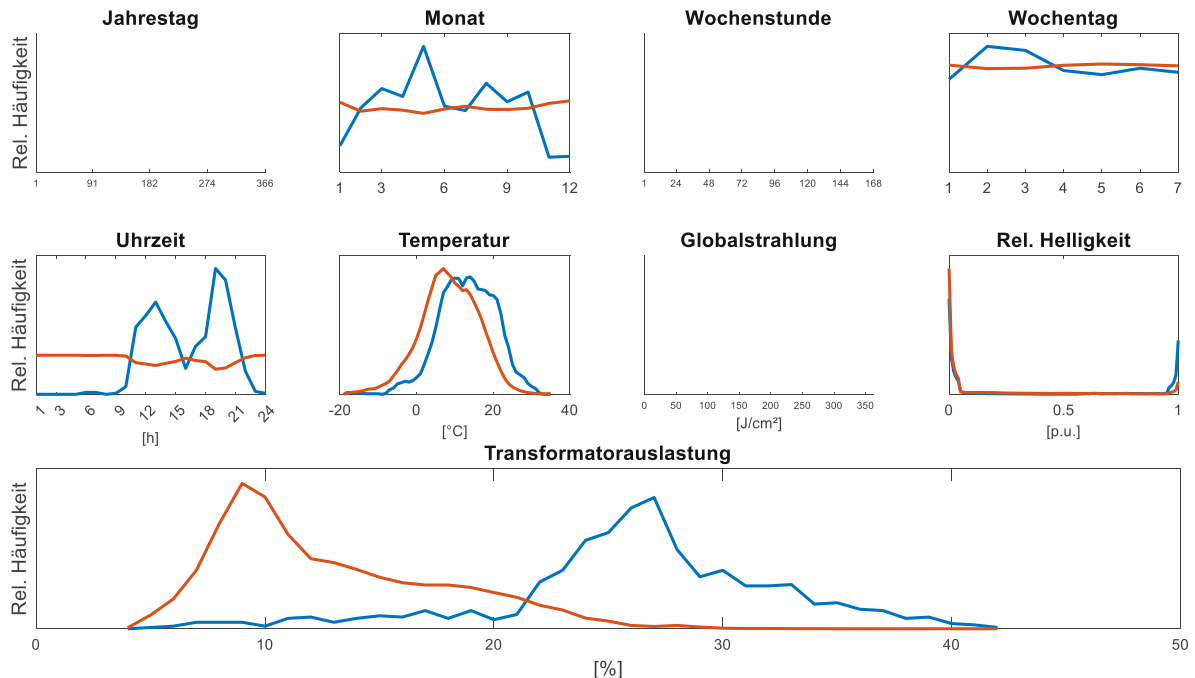


Abbildung 60: Dichtefunktionen beim Distanzmatrix-Algorithmus „Diskret 2“ mit einer Clusteranzahl 2

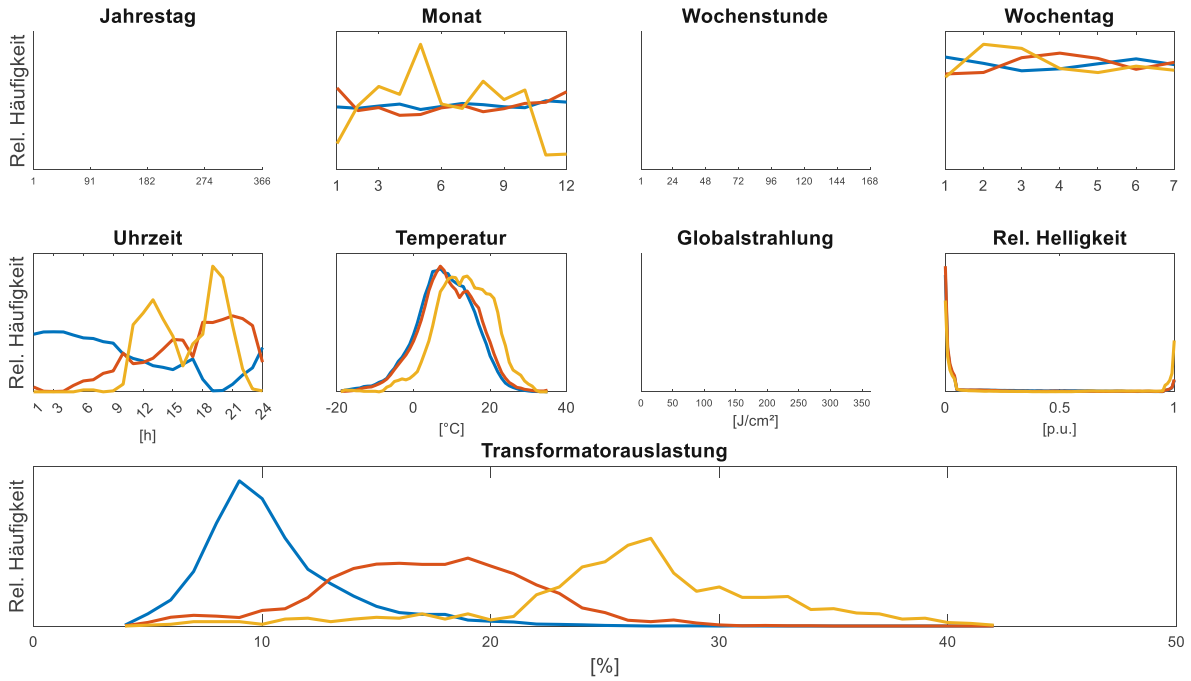


Abbildung 61: Dichtefunktionen beim Distanzmatrix-Algorithmus „Diskret 2“ mit einer Clusteranzahl 3

Gegenüberstellung und analytischer Vergleich der Algorithmen und Varianten

Nachdem die vier Varianten des Input-Datensatzes mit beiden Clusteralgorithmen geclustert wurden, wurden die Clusterresultate gegenübergestellt und Anpassungen vorgenommen. Folgende Erkenntnisse konnten gewonnen werden:

Aufteilung der Datenpunkte

Die Aufteilung der Datenpunkte in die unterschiedlichen Cluster ist dahingehend relevant, da Cluster mit einer zu geringen anteilmäßigen Anzahl an Datenpunkten nur eine bedingte Aussagekraft besitzen. Diese Datenpunkte zeigen zwar zu anderen Datenpunkten eine Unterscheidbarkeit, weshalb sie auch in unterschiedliche Cluster eingeteilt wurden, jedoch kann diese Minorität nicht gleichwertig bei der anschließenden Prognose von Belastungsszenarien verwendet werden. Abbildung 62 zeigt dazu einen Vergleich zwischen den Ergebnissen des EM-Algorithmus und des Distanzmatrix-Algorithmus. Die Balken stellen jeweils die maximale Abweichung der aufgeteilten Datenpunkte verglichen mit einer gleichmäßigen Aufteilung dar. Beispielsweise wäre eine gleichmäßige Aufteilung bei 4 Clustern jeweils 25% an Datenpunkten.

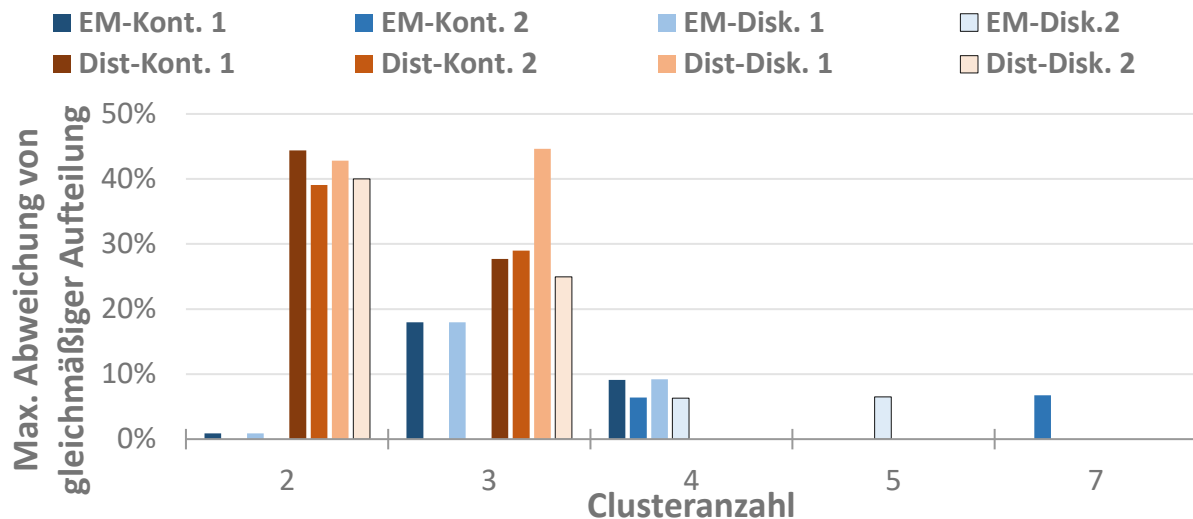


Abbildung 62: Vergleich zur Aufteilung der Datenpunkte

Die Ergebnisse des EM-Algorithmus zeigen eine geringe Abweichung zur gleichmäßigen Aufteilung der Datenpunkte und so auch keine zu kleinen Cluster. Beim Distanzmatrix-Algorithmus kommt es zu stärkeren Abweichungen und daher auch zu einem Ungleichgewicht zwischen den Clustern. Generell ist ein Ungleichgewicht bei der Clustergröße nicht zu verwerfen, jedoch können die so resultierenden Cluster schwieriger bis kaum in die anschließende Prognose miteinfließen, wodurch die Clusteranzahl reduziert werden müsste (zwei Cluster kombinieren) und so das eigentliche Clustering hinfällig macht.

Clusteranzahl

Da mit Hilfe der Informationskriterien bzw. Validierungskoeffizienten keine eindeutige zu bevorzugende Clusteranzahl gefunden wurde, wurden mehrere Bereiche definiert, in dem eine optimale Clusteranzahl vermutet wurde. Die resultierenden Cluster wurden anschließend analytisch auf dessen Unterscheidbarkeit bewertet und eine optimale Clusteranzahl ermittelt.

Beim EM-Algorithmus zeigte sich bei allen Durchläufen, dass die resultierenden Cluster eine starke Unterscheidbarkeit in Bezug auf die Uhrzeit aufweisen. Diese Unterscheidung kann sehr einfach auf die Eigenschaft Sonnenverlauf (Globalstrahlung und rel. Helligkeit) übertragen werden, da die Sonneneinstrahlung ebenfalls dem Tagesverlauf folgt. Betrachtet man das Clusterergebnis in Bezug auf die Transformatorauslastung, so kann man bis zu einer Clusteranzahl von 3 unterscheidbare Cluster erkennen. Diese Unterscheidung deckt sich mit der zeitlichen Abhängigkeit, da die geringe Auslastung vor allem in den Nachtstunden, eine gemischte Auslastung zur Mittagszeit und eine höhere Auslastung zur Abendzeit auftritt. Bei mehr als 3 Clustern, war die Unterscheidbarkeit nicht mehr so stark gegeben. Es kam dazu, dass sich Cluster überwiegend ähnlich waren und nur in einer Eigenschaft unterschieden.

Beim Distanzmatrix-Algorithmus zeigte sich eine starke Unterscheidbarkeit in Bezug auf die Transformatorauslastung, welche auch in der Eigenschaft Uhrzeit zu finden war. Verglichen mit dem EM-Algorithmus konnten jedoch keine eindeutigen Unterscheidungen wie beispielsweise Tag oder Nacht erkannt werden. Die Datenpunkte mit einer höheren Transformatorauslastung wurden Uhrzeiten um die Mittagszeit und Abendzeit zugeordnet. Es ist anzunehmen, dass die schwierige Interpretation des Clusterergebnisses auf die unausgeglichene Aufteilung der Datenpunkte zurückzuführen ist.

Variante des Input-Datensatzes

Die vier Varianten zur Zusammenstellung des Input-Datensatzes zeigten bei beiden Algorithmen übereinstimmende Resultate. Sowohl die kontinuierlichen Varianten als auch die diskreten Varianten wurden von den Algorithmen gleichwertig verarbeitet. Bezüglich der Eigenschaften Globalstrahlung und relative Helligkeit zeigten sich ebenfalls keine unterschiedlichen Ergebnisse.

Fazit

Da die Varianten bzw. Zusammensetzung des Input-Datensatzes keinen Einfluss auf das Clustering aufwies, wurde für die weiteren Auswertungen die Variante „Diskret 1“ angewendet. Diese weist die praktikabelste Darstellungsform auf, da die Eigenschaften Monat und Wochentag im Alltag gebräuchlicher sind und mit der Eigenschaft Globalstrahlung keine neue Variable erzeugt werden muss. Bei der Clusteranzahl wurden 3 Cluster als optimal definiert, da sich bis zu dieser Anzahl eine brauchbare Unterscheidbarkeit zeigte und eine größere Anzahl zu durchwegs redundanten Clustern führte. Als Algorithmus wurde der EM-Algorithmus weiterverfolgt, da dieser schlüssigere Ergebnisse produzierte. Der Distanzmatrix-Algorithmus wurde dennoch bei der Prognose der Belastungsszenarien in Punkt 4.2 als Vergleichsalgorithmus verwendet.

Für die nachfolgenden Auswertungen wurden folgende Anpassungen getroffen

- Clusteranzahl bei 3 festgelegt
- EM-Algorithmus mit Variante „Diskret 1“ weiterverwendet

4.1.2. Zweite Evaluierungsrunde

In der zweiten Evaluierungsrunde wurde nur mehr der EM-Algorithmus mit einer Clusteranzahl von 3 verwendet. Es wurde untersucht, ob der Wertebereich der Zeitvariablen einen Einfluss auf das Clustering hat und ob sich durch neue Anschlusspunkte in dem betrachteten Versorgungsgebiet im Entwicklungsszenario-2 das Clustering verändert.

Einfluss der Zeitvariablen

Die Zeitvariablen mussten für die Verwendung mit dem EM-Algorithmus auf einer Skala dargestellt werden, obwohl die Zeitvariablen in Wirklichkeit Kreisvariablen entsprechen (siehe auch Punkt 3.5.2 zur Umrechnung in Kreisvariablen beim Distanzmatrix-Algorithmus). Dadurch kommt es zu einer ungewollten Distanzierung beispielsweise von Morgen und Abend, da die zugewiesenen Werte weit voneinander entfernt liegen. Ob und wie stark dieser Einfluss auf das Clusterergebnis ist, wurde mit Hilfe einer Bias-Evaluierung ermittelt. Dazu wurden die Werte der Zeitvariablen manipuliert und der EM-Algorithmus erneut auf den Input-Datensatz angewendet. Je nachdem wie sehr das Ergebnis vom nicht manipulierten Datensatz abwich, desto empfindlicher ist der EM-Algorithmus und muss bei seiner Anwendung berücksichtigt werden.

Die Werte der Eigenschaft Uhrzeit bzw. Monat wurden um einen halben Tag bzw. ein halbes Jahr verschoben, sodass beispielsweise ein Tag mit dem Wert 13 startet, zur Mittagszeit den Wechsel vom Wert 24 auf 1 vornimmt und den Tag mit dem Wert 12 beendet. Abbildung 64 zeigt die Dichtefunktionen mit den manipulierten Werten der Uhrzeit. Im Vergleich dazu ist in Abbildung 63 die Ausgangslage ohne Manipulation dargestellt.

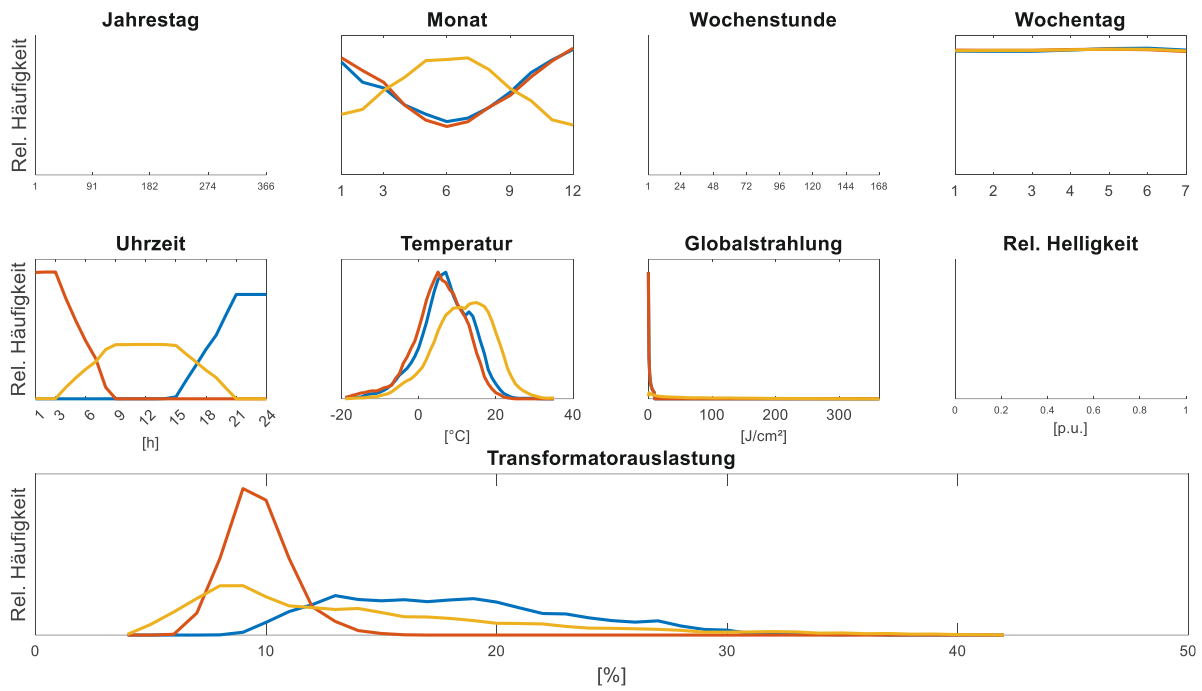


Abbildung 63: EM-Algorithmus mit "Diskret 1" Status Quo

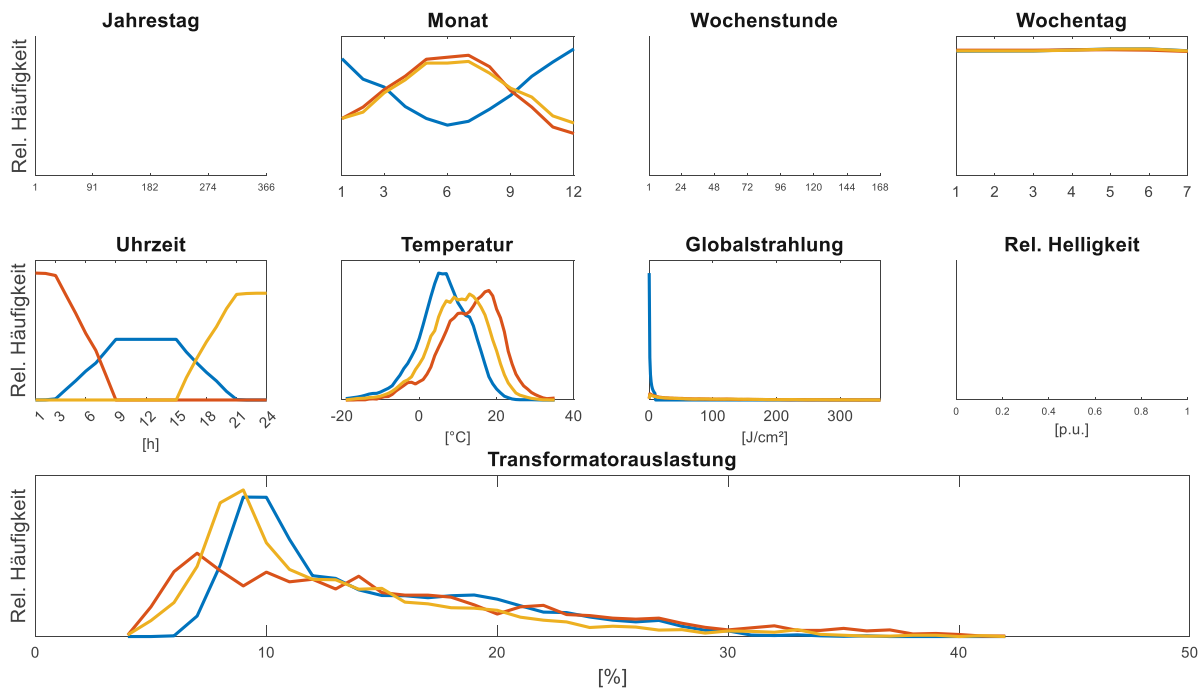


Abbildung 64: EM-Algorithmus mit "Diskret 1" und einem Bias der Uhrzeit

Aus dem Vergleich mit und ohne Bias der Uhrzeit erkennt man, dass die Wertezuteilung der Uhrzeit einen Einfluss auf das Clusterergebnis hat. Die Aufteilung der Cluster innerhalb der Eigenschaft Uhrzeit zeigt dasselbe Ergebnis mit und ohne Bias. Aufgrund der manipulierten Uhrzeit entspricht diese Aufteilung real einem Vormittags-, Nachmittags- und Nacht-Cluster. Als nächstes wurden die manipulierten Monatswerte miteinbezogen. Abbildung 65 zeigt die Dichtefunktionen mit manipulierten Werten von Monat und Uhrzeit. Im Vergleich dazu zeigt Abbildung 66 die Dichtefunktionen nur mit den manipulierten Werten vom Monat.

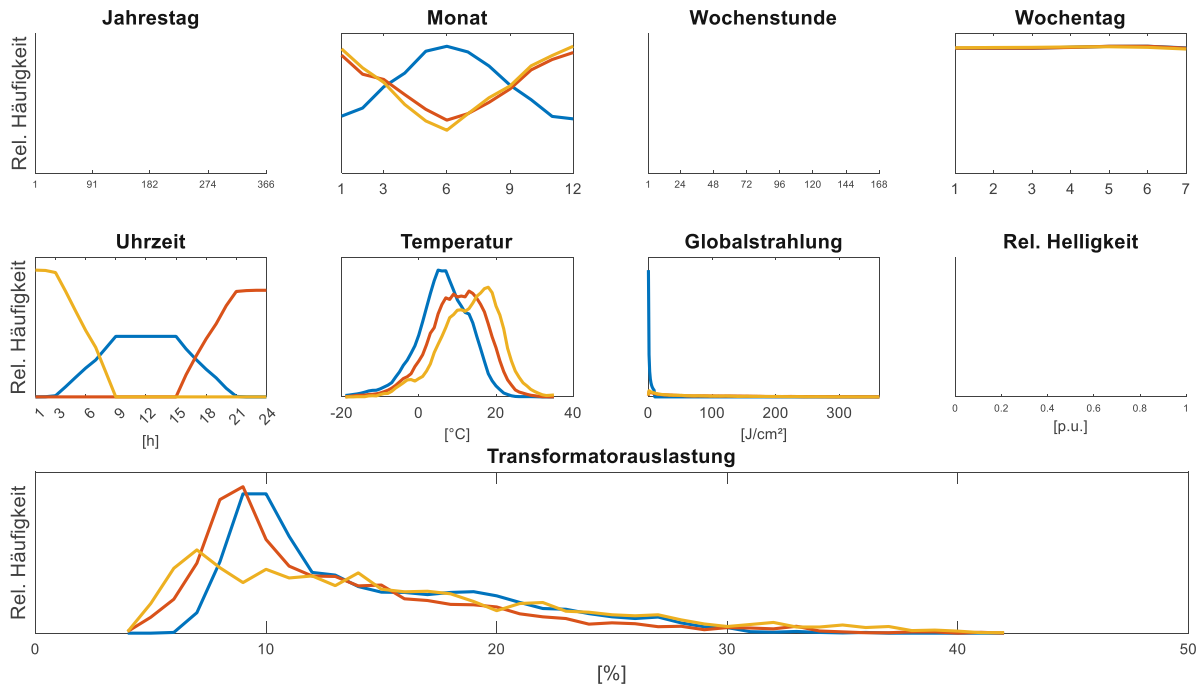


Abbildung 65: EM-Algorithmus mit "Diskret 1" und einem Bias bei Monat und Uhrzeit

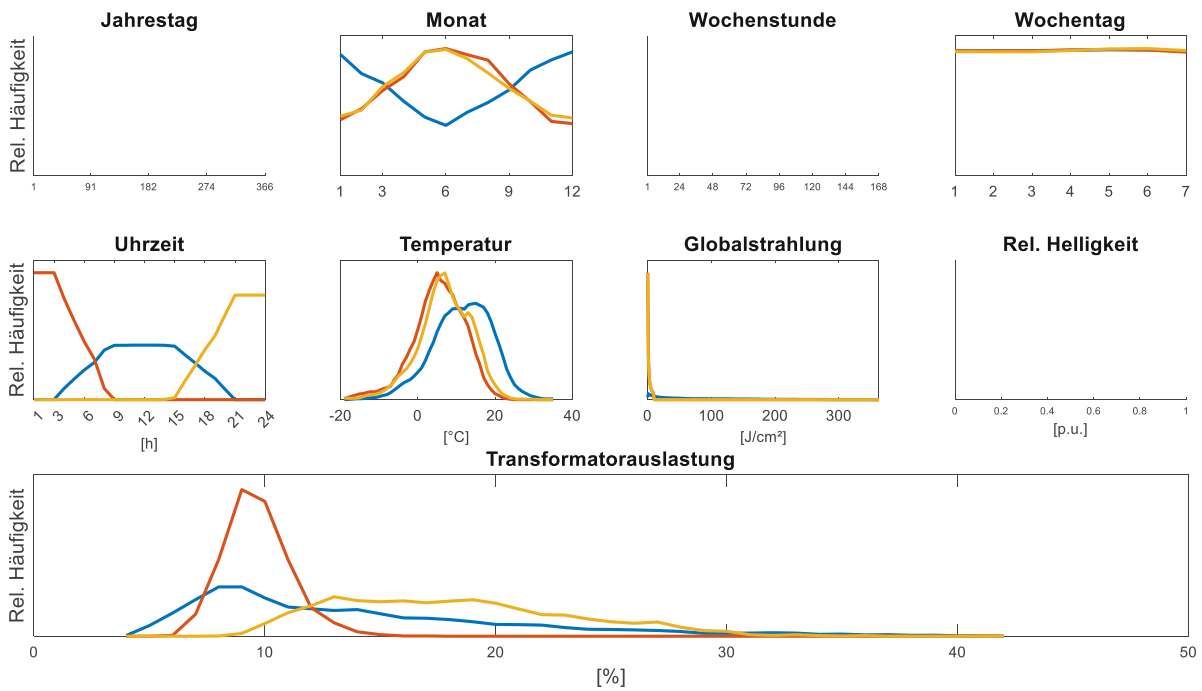


Abbildung 66: EM-Algorithmus mit "Diskret 1" und einem Bias beim Monat

Werden zusätzlich zur Uhrzeit auch noch die Werte vom Monat manipuliert, zeigt dies keinen Einfluss auf das Clusterergebnis, da sich Abbildung 64 und Abbildung 65 nicht unterscheiden. Werden nur die Werte vom Monat manipuliert und die Uhrzeit erhält wieder die ursprünglichen „richtigen“ Werte (Abbildung 66), zeigt sich wieder dasselbe Ergebnis wie bei der Ausgangslage in Abbildung 63. Es ist daher anzunehmen, dass die Eigenschaft Monat kaum einen Einfluss auf das Clustering besitzt.

Versorgungsnetz mit Entwicklungsszenario-2

Nachdem die Clusterergebnisse vor allem eine Abhängigkeit der Uhrzeit zeigten und weniger Abhängigkeiten zu anderen Eigenschaften, sollte mit einem Vergleich zwischen dem betrachteten Stromnetz im Basisszenario-0 und Entwicklungsszenario-2 abgeklärt werden, ob mit mehr erneuerbaren Erzeugereinheiten im Stromnetz (die bekanntlich stark wetterabhängig sind) auch die Abhängigkeit zu den Wettervariablen steigt. Es wurde für diese Auswertung auf die beiden Eigenschaften Jahresverlauf und Wochenverlauf verzichtet und der Input-Datensatz damit reduziert. Abbildung 67 zeigt dazu die Dichtefunktionen nach dem Clustering mit dem 4-dimensionalen Input-Datensatz.

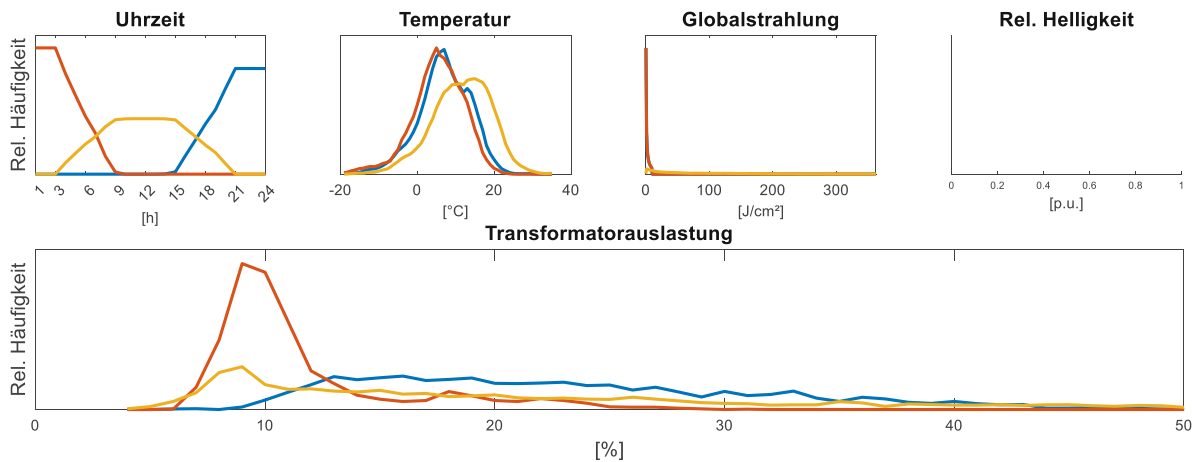


Abbildung 67: EM-Algorithmus im Entwicklungsszenario-2

Es zeigte sich, wie schon im Basisszenario-0, die stärkste Unterscheidbarkeit bei der Eigenschaft Uhrzeit und keine wesentliche Verstärkung der Abhängigkeit bei den beiden Wetter-Variablen.

Input-Datensatz mit mehreren Jahren

Zuletzt wurde noch untersucht, ob die Anzahl der Lerndaten einen Einfluss auf das Clusterergebnis hat bzw. wie es sich darauf auswirkt. Bei den vorangegangenen Analysen wurden immer Lerndaten aus einem Jahr verwendet, sprich ein Datensatz aus 8784 Datenpunkten. Für zusätzliche Lerndaten wurden die erzeugten Jahreszeitreihen aus Punkt 3.1 hergenommen und mit einer erneuten Lastflussberechnungen nach Punkt 3.2 mehrere Jahreszeitreihen der Transformatorauslastung erzeugt. Die restlichen Eigenschaften des Input-Datensatzes blieben unverändert.

Die Zusammensetzung des Input-Datensatzes wurde einmal mit „Diskret 1“ (6-dimensional) angewendet und einmal mit dem reduzierten 4-dimensionalen Datensatz aus dem vorhergehenden Punkt (Uhrzeit, Temperatur, Globalstrahlung und Transformatorauslastung). Bei einem Lerndatensatz, bestehend aus 5 Jahren (43.920 statt 8.784 Daten), zeigten sich bei beiden Varianten dieselben Ergebnisse wie aus Abbildung 45 bzw. Abbildung 67. Man kann daher annehmen, dass die Anzahl der Lerndaten keinen Einfluss auf das Clusterergebnis nimmt.

4.1.3. Fazit Clustering

Die Evaluierung der Clusteralgorithmen hat gezeigt, dass vor allem die Uhrzeit als Richtwert für die Auslastungswerte eines Transformators hergenommen werden kann. Als zweites zeigte auch die Temperatur eine brauchbare Unterscheidbarkeit, wodurch unterschiedliche Cluster getrennt wurden. Die restlichen Variablen zeigten teilweise keine Unterscheidbarkeit (Jahresverlauf und Wochenverlauf) bzw. eine Unterscheidbarkeit, die für die Prognose als ungeeignet empfunden wurde (Sonnenerlauf, da die Tag-/Nacht-Information besser anhand der Zeit verwendet werden kann). Von den beiden untersuchten Algorithmen hat sich der EM-Algorithmus als brauchbar erwiesen, da dieser die Input-Daten in unterscheidbare Cluster teilt und dabei eine brauchbare Aufteilung der Datenpunkte vornimmt. Der Distanzmatrix-Algorithmus wurde dennoch für die Prognose der Belastungsszenarien verwendet um einen Vergleichswerte zu erhalten. Als Einschränkung muss beim EM-Algorithmus die Anwendbarkeit bezüglich der Zeit-Variablen erwähnt werden, da dem Algorithmus keine Kreisvariablen übergeben werden können und es so zu einer unweigerlichen Trennung von Zeitpunkten an den Rändern der Wertebereiche kommt. Diese Einschränkung besitzt der Distanzmatrix-Algorithmus nicht, da die Zeit-Variablen als Kreisvariablen behandelt werden. Für die Prognose der Unsicherheiten wurden die beiden Algorithmen wie folgt auf den Input-Datensatz angewendet, um die Clustereinteilung zu erzeugen:

Tabelle 17: Anwendung der Algorithmen für die Prognose

	EM-Algorithmus	Distanz-Algorithmus
Konfiguration	Kovarianzmatrix voll + getrennt	mit Gewichtungsfaktoren
Clusteranzahl	3	
Input-Datensatz	5-dimensional Monat, Uhrzeit, Temperatur, Globalstrahlung, Trafo-Auslastung	
Verwendete Eigenschaften für die Prognose	Uhrzeit und Temperatur	

Als ideale Eigenschaften für die nachfolgende Prognose haben sich die Eigenschaften Uhrzeit und Temperatur erwiesen, da bei diesen beiden Eigenschaften die Clusterergebnisse eine Unterscheidbarkeit zwischen den Clustern zeigte, welche bei den Belastungsszenarien abgefragt werden konnte. Ein Belastungsszenario kann dennoch aus mehreren Eigenschaften definiert werden, jedoch dienen diese zusätzlichen Eigenschaften nur dem Anwender als leichtere Handhabung eines Belastungsszenarios.

4.2. Prognose der Unsicherheiten von Belastungsszenarien

Für die Evaluierung der Unsicherheiten von Belastungsszenarien wurden mehrere Szenarien definiert, für welche eine Prognose mit den zuvor geclusterten Daten durchgeführt wurde. Die Güte der Prognose wurde anschließend anhand eines Test-Datensatzes bewertet. Diese Test-Daten der Transformatorauslastung wurden aus Lastflussberechnungen gewonnen, welche mit Hilfe der zusätzlich generierten Verbraucher-Zeitreihen erzeugt wurden (siehe Generierung von Zeitreihen in Punkt 3.1). Es wurden auf diese Weise 10 neue Jahre mit den dazugehörigen Transformatorauslastungen erzeugt. Zusätzlich zu dem Vergleich mit den Test-Daten wurde die Prognose anhand der Jahresauslastung des Transformators bewertet (siehe Histogramm in Abbildung 26).

Die drei betrachteten Belastungsszenarien setzen sich wie folgt zusammen:

- Belastungsszenario 1: Jänner Früh 7 Uhr (± 1 h) -1 °C (± 2 °C)
- Belastungsszenario 2: März Mittag 12 Uhr (± 1 h) 15 °C (± 2 °C)
- Belastungsszenario 3: August Abend 20 Uhr (± 1 h) 20 °C (± 2 °C)

Ein Belastungsszenario besteht demnach aus einem Monat, einer Uhrzeit und Temperatur, wobei bei der Berechnung nur die Eigenschaften Uhrzeit und Temperatur verwendet wurden (siehe Punkt 4.1.3). Bei der Definition der Belastungsszenarien wurde auf die Eigenschaften Wochentag und Globalstrahlung verzichtet, da der Wochentag gemäß den Erkenntnissen aus Punkt 4.1 keinen Einfluss zeigte und sich die Globalstrahlung bei der Abfrage von Szenarien als ungeeignet erwies (Uhrzeit und Globalstrahlung müssten immer im Einklang stehen, da sonst widersprüchliche Abfragen entstehen würden).

Die Bereiche ± 1 h bei der Uhrzeit und ± 2 °C bei der Temperatur wurden verwendet, weil exakte Wertekombinationen aus Monat, Uhrzeit und Temperatur nur sehr selten bis gar nicht im Jahr auftraten, wodurch nicht genügend Test-Daten für eine Evaluierung verfügbar gewesen wären. Erst durch die Berücksichtigung der umliegenden Uhrzeiten und Temperaturen fielen genügend Test-Daten in die Belastungsszenarien. Durch den gewählten Wertebereich ergaben sich pro Jahr 35, 14 bzw. 26 Zeitpunkte in den Belastungsszenarien 1,2 und 3. Multipliziert mit den 10 Jahren konnte so eine ausreichend große Anzahl an Test-Daten generiert werden. Dieser Wertebereich wurde nur für die Ermittlung von Test-Daten verwendet. Bei der eigentlichen Prognose wurden die exakten Werte verwendet (siehe auch Methodik, Punkt 3.7)

Abbildung 68 zeigt die prognostizierten Dichtefunktionen der drei Belastungsszenarien. Als Referenz dazu dient die Dichtefunktion der Test-Daten und die der ganzen Jahresauslastung.

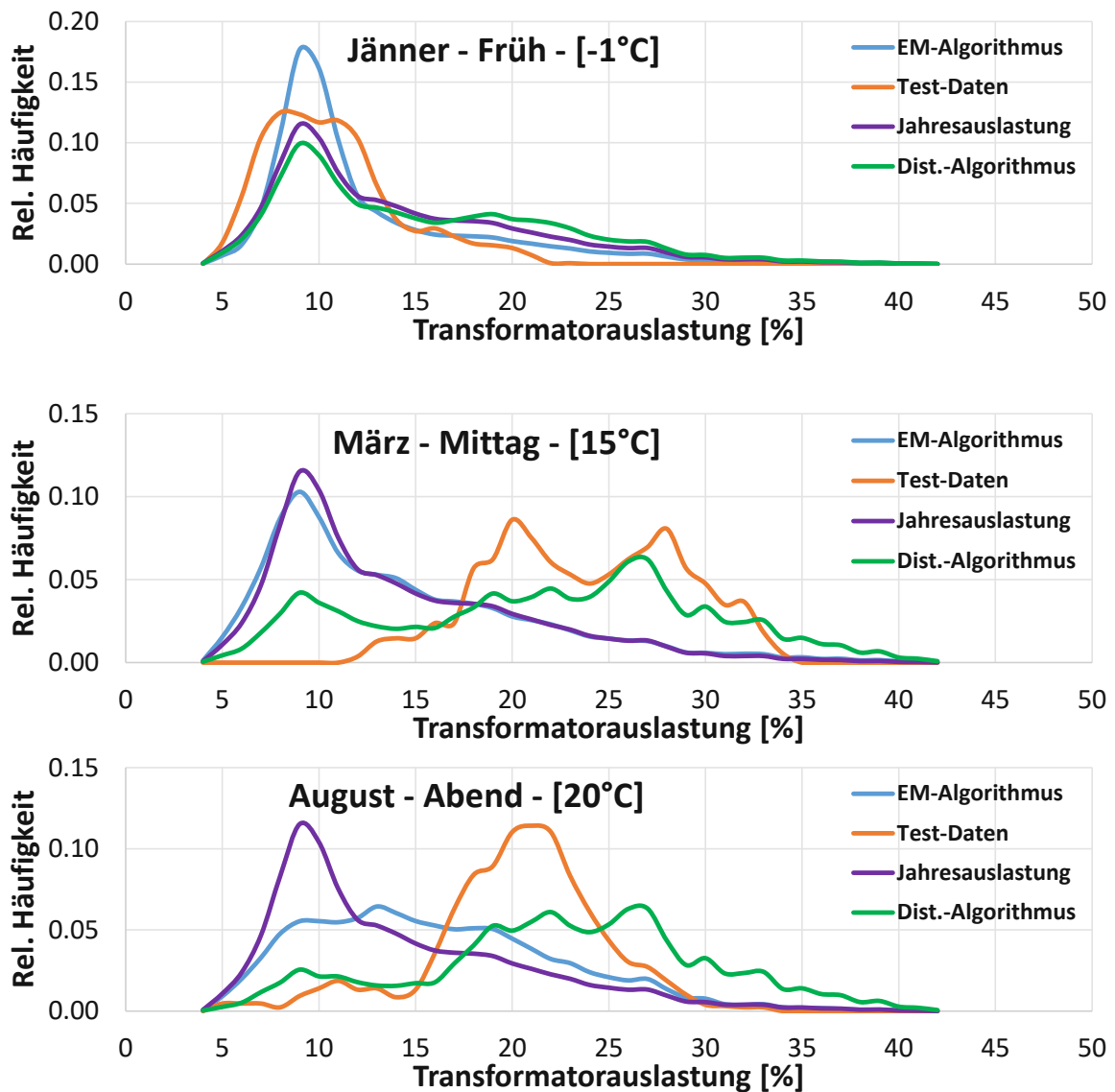


Abbildung 68: Prognose von drei Belastungsszenarien

Im Belastungsszenario Jänner zeigen die prognostizierten Dichtefunktionen beider Algorithmen ein gutes Ergebnis in Bezug auf die Test-Daten. Bei den Belastungsszenarien im März und August hingegen weicht die prognostizierte Verteilung des EM-Algorithmus stärker von den Test-Daten ab und deckt sich dadurch kaum mit ihnen. Der Distanzmatrix-Algorithmus hingegen zeigt eine bessere Übereinstimmung, wenn auch nicht im gleichen Maße wie beim Belastungsszenario im Jänner.

Mit Hilfe der Methode der kleinsten Quadrate wurde zusätzlich die Güte der prognostizierten Belastungsszenarien bewertet. Die herangezogenen Abstände zwischen den einzelnen Datenpunkten entsprechen den Abständen der relativen Dichtefunktion aus Abbildung 68.

Tabelle 18: Summe der Abweichungsquadrate der prognostizierten Belastungsszenarien zu den Test-Daten

	Szenario Jänner	Szenario März	Szenario August	Ø
Jahresauslastung	0,01	0,08	0,09	0,06
EM-Algorithmus	0,01	0,07	0,04	0,04
Distanzmatrix-Algorithmus	0,02	0,01	0,02	0,02

In Tabelle 18 dient die Jahresauslastung als Referenz in Bezug auf die Unsicherheiten dieser Belastungsszenarien. Der EM-Algorithmus erzielt somit eine Verbesserung der Prognose und kann somit die Unsicherheiten minimiert darstellen. Eine weitere Verbesserung erreicht der Distanzmatrix-Algorithmus, welcher die Unsicherheiten bezüglich des EM-Algorithmus nochmals halbiert.

Eine weitere Vergleichsmöglichkeit bietet die Darstellung als Kastendiagramm, welches in Abbildung 69 zu finden ist. Hier ist mit der durchgängigen Linie im Kasten der Medianwert abzulesen und mit dem Kreuz ist der jeweilige Erwartungswert markiert. Der Kasten beinhaltet jeweils die mittleren 50% aller Daten, sprich der Bereich vom ersten Quartil bis zum dritten Quartil (Interquartilsabstand). Die Punkte außerhalb der Spannweite wurden als Ausreißer betrachtet. Als Ausreißer wurden jene Punkte gewertet, welche 1,5-mal größer sind als das dritte Quartil bzw. 1,5-mal kleiner sind als das erste Quartil.

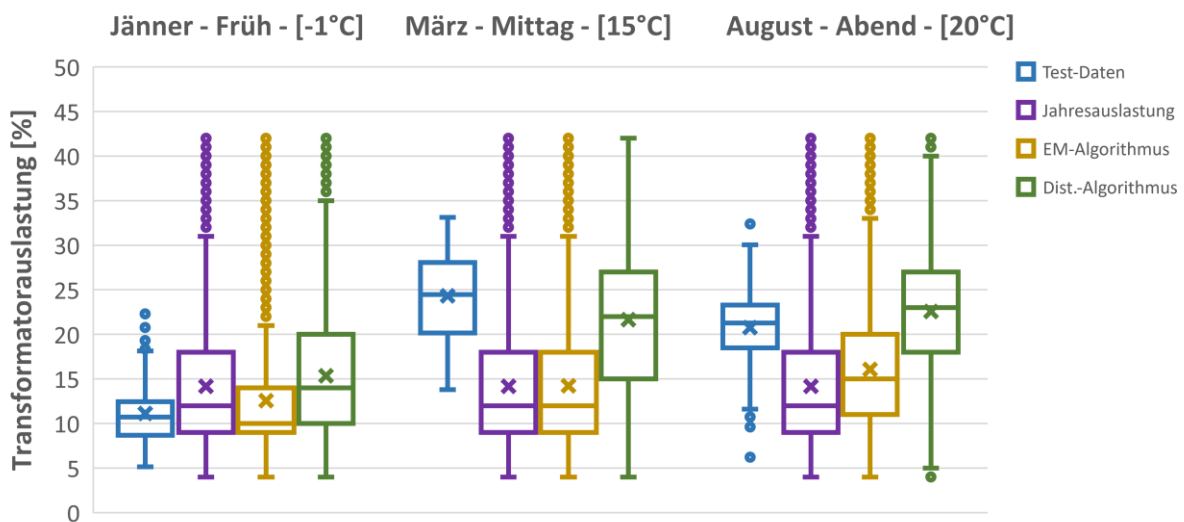


Abbildung 69: Boxplot der Belastungsszenarien

Mit Hilfe der Kastendiagramme ließen sich die prognostizierten Unsicherheiten der Belastungsszenarien sehr anschaulich vergleichen. Als Unsicherheiten wurde hier der Interquartilsabstand bzw. die komplette Spannweite betrachtet.

Der EM-Algorithmus konnte nur im Jänner-Szenario eine brauchbare Prognose tätigen, die einerseits den Test-Daten ähnelt und andererseits eine geringe Unsicherheit aufwies. In den beiden anderen Szenarien erwiesen sich die Prognosen nur bedingt als brauchbar, da die Aussage keine wesentliche Verbesserung im Vergleich zur Jahresauslastung zeigte. Im Vergleich dazu konnte der Distanzmatrix-Algorithmus bei den Interquartilsabständen brauchbare Ergebnisse erzielen, da diese besser an die Test-Daten herankamen. Bei der Spannweite zeigte er eine zu große Unsicherheit, da praktisch alle Auslastungswerte darin enthalten waren und kaum Ausreißer identifiziert werden konnten.

5. Schlussfolgerung und Ausblick

Die verwendeten Clusteralgorithmen konnten beide in den verarbeiteten Daten eines fiktiven Stromnetzes Strukturen erkennen, wodurch sich Bereiche unterscheiden ließen und damit eine Prognose der Unsicherheiten von Belastungsszenarien durchführen ließ. Bezugnehmend auf die eingangs gestellten Forschungsfragen, konnten mit Hilfe der durchgeführten Analysen folgende Erkenntnisse gewonnen werden:

- Wie können die Unsicherheiten von Belastungsszenarien einer Ortsnetzstation abgeschätzt werden?

Als Unsicherheiten von Belastungsszenarien einer Ortsnetzstation wurden in dieser Arbeit die Auslastungswerte der Transformatoren angesehen, die je nach Belastungsszenario unterschiedlich ausfallen und auch unterschiedlich verteilt sind. Als Unsicherheit wurde einerseits die Form der Wahrscheinlichkeitsverteilung (hier Dichtefunktion), welche die relativen Häufigkeiten aller Auslastungswerte beschreibt, definiert und andererseits das Streumaß (Quartile und Spannweite), welches sich aus den Daten der relativen Häufigkeiten ergab, als Bewertungsmaß herangezogen.

Die Unsicherheiten wurden dadurch ermittelt, indem mit Hilfe von geeigneten Clusteralgorithmen ein netzbeschreibender Datensatz, bestehend aus Zeit-, Wetter- und Betriebsdaten, in Gruppen (Cluster) unterteilt wurde. Die entstandenen Cluster dienten für die anschließende Berechnung einer resultierenden Dichtefunktion der Transformatorauslastung bei einem ausgewählten Belastungsszenario. Dazu wurde das Belastungsszenario jedem einzelnen Cluster in dem Ausmaß zugesprochen, welches der jeweiligen Zugehörigkeitsgewichtung des jeweiligen Clusters entsprach. Abschließend konnte mit dieser Gewichtung eine resultierende Dichtefunktion der Transformatorauslastung gemischt werden, welche das Belastungsszenario beschrieb.

- Wie ist die Güte der angewendeten Algorithmen in Bezug auf die Clusterergebnisse?

Der Distanzmatrix-Algorithmus wurde aus der Intension entwickelt, um Eigenschaften unterschiedlicher Strukturen (Kreisvariablen und skalare Variablen) richtig verarbeiten zu können und so eine uneingeschränkte Nutzung dieses Algorithmus möglich ist. Es zeigte sich aber, dass einerseits aufgrund der Berechnungsart der Clusterzugehörigkeit mit Hilfe der minimalen euklidischen Distanz und andererseits aufgrund der bereitgestellten eindimensionalen Distanzmatrix, die Resultate in unausgeglichene Cluster und nur bedingt interpretierbare Unterscheidungsmerkmale zwischen den Clustern resultierten. Bei der Zusammenführung der unterschiedlichen Dimensionen in eine gemeinsame Distanzmatrix gingen voraussichtlich die Informationen verloren, welche innerhalb einer Eigenschaft eine Unterscheidbarkeit zugelassen hätten, jedoch in der gemeinsamen eindimensionalen Distanzmatrix nicht mehr so stark ersichtlich waren. Des Weiteren muss bei der verwendeten Berechnungsart, rein mit der Minimierung von Distanzen, die Anfälligkeit auf Rauschen bzw. Anfälligkeit auf nicht kompakte Strukturen erwähnt werden. Dadurch kam es zum ungewollten Aufblähen der Cluster und wahrscheinlich auch zu der unausgeglichene Aufteilung der Datenpunkte. Eine Möglichkeit zur Identifikation dieser Ausreißer und besseren Verarbeitung war bei diesem Algorithmus nicht möglich.

Im Gegensatz dazu arbeitete der EM-Algorithmus mit den Eigenschaften als mehrdimensionaler Input-Datensatz, wodurch in jeder Eigenschaft Strukturen unterschieden werden konnten. Die Ergebnisse zeigten eine brauchbare Aufteilung der Datenpunkte und

zeigten eindeutige Unterscheidungsmerkmale zwischen den Clustern im Vergleich zum Distanzmatrix-Algorithmus. Die Berechnungsart mit Hilfe der Normalverteilungen eignete sich besser bei der Verwendung dieses Datensatzes, da Ausreißer nicht so schwer ins Gewicht fielen und dadurch die Cluster nicht unnötig aufgebläht wurden. Als Einschränkung des EM-Algorithmus wurde die Verarbeitung von Zeit-Variablen beobachtet, da alle Größen nach deren absoluten Werten verarbeitet wurden und die Struktur der Kreisvariablen verloren ging.

Die Ergebnisse beider Clusteralgorithmen lieferten trotz unterschiedlicher Berechnungsart dieselben Erkenntnisse in Bezug auf die Wahl der Eigenschaften aus dem Input-Datensatz. Der Jahresverlauf (Jahrestag und Monat) trug keinen wesentlichen Beitrag zur Unterscheidbarkeit der Cluster bei und zeigte durchwegs idente Dichtefunktionen aller Cluster. Ein ähnliches Verhalten zeigte der Wochenverlauf (Wochenstunde und Wochentag), wobei eine Unterscheidung der Dichtefunktionen zwischen den Clustern möglich war, jedoch der Beitrag zum Clusterergebnis als gering erachtet wurde. Den stärksten Einfluss auf das Clustering trug die Uhrzeit bei, da alle Unterscheidungsmerkmale auf eine Tagesstruktur zurückzuführen waren. Dadurch wurde die Uhrzeit für die Prognose der Belastungsszenarien als Abfrageeigenschaft verwendet. Die Temperatur zeigte ebenfalls unterschiedliche Dichtefunktionen zwischen den Clustern und wurde ebenfalls bei der Prognose verwendet. Der Sonnenverlauf zeigte vor allem Unterscheidungsmerkmale zwischen Tag und Nacht. Es zeigte sich aber keine differenziertere Unterscheidbarkeit, wodurch ein Mehrwert zur Uhrzeit, welche ebenfalls eine Tag- und Nacht-Unterscheidbarkeit aufwies, nicht gegeben war und auch diese Eigenschaft nicht für die Prognose herangezogen wurde.

Die Clusterergebnisse, im speziellen die starke Unterscheidbarkeit mit Hilfe der Uhrzeit, sind durchwegs auf die Tatsache zurückzuführen, dass die angeschlossenen Erzeuger- und Verbrauchereinheiten alle einen typischen Tagesverlauf aufweisen, welcher einen nächtlichen Verbrauch mit Grundlast zeigt und über den Tag gesehen eine oder mehrere Lastspitzen besitzt. Die Eigenschaften Temperatur und Sonnenverlauf können ebenfalls genau auf diese Struktur geführt werden. Bezüglich der Abhängigkeit zum Sonnenverlauf (Globalstrahlung und relative Helligkeit) wurde besonders beim Entwicklungsszenario-2 eine stärkere Abhängigkeit vermutet, welche sich aber bei der Auswertung nicht zeigte. Der Grund dafür könnte in der Zusammensetzung des Input-Datensatzes, in Kombination mit dem verwendeten fiktiven Stromnetz, liegen. Obwohl die Eigenschaften mit bestmöglicher Sorgfalt ausgewählt wurden und mit dem Stromnetz abgestimmt waren, stammen diese Teile nicht aus derselben Quelle. Es kann dadurch die exakte Korrelation zwischen der Sonneneinstrahlung und den PV-Anlagen nicht zweifelsfrei bestätigt werden. Diese Korrelation ist aber eine zentrale Voraussetzung bei der Anwendung der Clusteralgorithmen und könnte daher eine Ursache für die schwachen Unterscheidungsmerkmale beim Sonnenverlauf sein.

Der Einfluss von der Anzahl der Input-Jahre, sprich die Größe des Lern-Datensatzes, wurde als untergeordnet eingestuft. Es zeigten sich idente Ergebnisse der geclusterten Daten mit einem Input-Datensatz bestehend aus einem Jahr und bestehend aus fünf Jahren. Demzufolge sind alle Strukturen in dem Datensatz bereits bei einem einjährigen Datensatz enthalten und konnten von dem EM-Algorithmus identifiziert werden.

- Welchen Mehrwert bietet die getätigte Abschätzung der Unsicherheiten von Belastungsszenarien?

Die Abschätzungen der Unsicherheiten bei den untersuchten Belastungsszenarien zeigten keine gleichbleibende Güte im Vergleich mit den Test-Daten. Würde man laut Jänner-Szenario der Prognose eine hohe Güte zuschreiben, so müsste man beim März- und August-Szenario eine vergleichsweise niedrige Güte ableiten. Diese schwankende Aussagekraft der

abgeschätzten Unsicherheiten resultierte durch die Abfrage der Belastungsszenarien anhand der Dichtefunktionen. Da die Unterscheidbarkeit der geclusterten Daten nicht immer eindeutig ist, kommt es nach der Zusammensetzung des Belastungsszenarios bei der resultierenden Dichtefunktion annähernd wieder zur Form der ursprünglichen Jahresauslastung (entspricht einem gemeinsamen Cluster). Die Abschätzungen zeigen somit eine hohe Unsicherheit der Belastungsszenarien, welche nicht wesentlich durch die Clusteranalyse verfeinert werden konnte.

- Welche Verbesserungen oder Weiterentwicklungen müssen für eine brauchbare Abschätzung gemacht werden?

Mit der Durchführung dieser Forschungsarbeit konnte gezeigt werden, wie Belastungsszenarien mit Hilfe von maschinellem Lernen abgeschätzt werden können. Dabei zeigte sich besonders bei den Ergebnissen der Clusteranalyse, dass eine Einteilung der Daten in unterschiedliche Cluster nur bedingt möglich war. Es stellte sich hier die Frage, ob der verwendete SimBench-Datensatz in Kombination mit den Wetterdaten des Deutschen Wetterdienstes allgemein für diese Art von Datenanalyse geeignet ist. Daher wäre diese Art von Analyse anhand eines realen Stromnetzes zu wiederholen, um optimal korrelierende Eigenschaften als Input-Datensatz verwenden zu können.

Die angewendeten Algorithmen zeigten beide Vor- und Nachteile bei der Abschätzung der Unsicherheiten. Der Vorteil des Distanzmatrix-Algorithmus bestand in der Verarbeitung der Zeit-Variablen als Kreisvariablen und der Vorteil des EM-Algorithmus war die gute Aufteilung der Datenpunkte bzw. die Handhabung von Ausreißern. Es würde sich daher anbieten, diese Vorteile zu kombinieren und erneut mit Daten aus einem realen Stromnetz zu verwenden.

Literaturverzeichnis

- [1] H. Schmid, Die Elektrifizierung Niederösterreichs in der Zwischenkriegszeit: Baugeschichte und Netzentwicklung, Wien: Universität Wien, 2008.
- [2] E-Control Austria, „Versorgungssicherheit,“ 2021. [Online]. Available: <https://www.e-control.at/industrie/strom/versorgungssicherheit>. [Zugriff am 21 November 2021].
- [3] P. Wiest, Probabilistische Verteilnetzplanung zur optimierten Integration flexibler dezentraler Erzeuger und Verbraucher, Göttingen: sierke Verlag, 2018.
- [4] B. Özalay, H. Schuster, J. Kellermann, J. Priebe und A. Moser, „Wirtschaftlicher Vorteil der netzdienlichen Nutzung von Flexibilität in Verteilnetzen,“ E-Bridge Consulting GmbH, Bonn, 2019.
- [5] Austrian Power Grid AG, „N-1 Sicherheit,“ 2021. [Online]. Available: <https://www.apg.at/de/Energiezukunft/Glossar/N-1-Sicherheit>. [Zugriff am 21 November 2021].
- [6] Next Kraftwerke GmbH, „Warum ist das (N-1)-Kriterium im Stromnetz?,“ 2021. [Online]. Available: <https://www.next-kraftwerke.de/wissen/n-1-kriterium>. [Zugriff am 21 November 2021].
- [7] Bundeskanzleramt Österreich, „Fit for 55"-Paket - EU-Kommission geht Herausforderungen zum Klimaschutz an,“ 16 Juli 2021. [Online]. Available: <https://www.bundeskanzleramt.gv.at/themen/europa-aktuell/fit-for-55-paket-eu-kommission-geht-herausforderungen-zum-klimaschutz-an.html>. [Zugriff am 21 November 2021].
- [8] BMK, „Erfolgreiche Einigung bei Erneuerbaren-Ausbau-Gesetz,“ 7 Juli 2021. [Online]. Available: https://www.bmk.gv.at/service/presse/gewessler/20210706_eag.html. [Zugriff am 21 November 2021].
- [9] A. Nolde, R. Frings und P. Zimmer, „Smart Area Aachen,“ 2014.
- [10] M. Fiedeldey, Analyse und Prognose elektrischer Lastgangszeitreihen, Göttingen: Cuvillier Verlag, 2010.
- [11] DIHK Berlin & Brüssel; EFET Deutschland, Strombeschaffung und Stromhandel, Berlin: DIHK - Deutscher Industrie- und Handelskammertag Berlin & Brüssel; EFET Deutschland - Verband Deutscher Energiehändler e.V., 2020.
- [12] OTH Regensburg/FENES, „Engpassmanagement,“ 2022. [Online]. Available: <https://www.fenes.net/forschung/energienetze/themengebiete/engpassmanagement/>. [Zugriff am 11 Jänner 2022].
- [13] Austrian Power Grid AG, „Redispatch-Kosten in der APG im Jahresvergleich 2011-2021,“ 2021. [Online]. Available: <https://www.apg.at/de/Media-Center/Infografiken>. [Zugriff am 14 Jänner 2022].
- [14] Brandenburgische Technische Universität Cottbus, Anwendung der Repräsentativen VDEW-Lastprofile step - by - step, Frankfurt: VDEW-Frankfurt, 2000.

- [15] R. Vilgut und U. Sommersguter, „Kleine Zeitung,“ 19 Juli 2021. [Online]. Available: https://www.kleinezeitung.at/wirtschaft/6009962/Zeitplan-gescheitert_Smart-Meter_Einfuehrung-wird-auf-die-lange#:~:text=Die%20%22Intelligente%20Messger%C3%A4te%2DEinf%C3%BChrungsv erordnung%22,wirtschaftlich%20und%20technisch%20machbar%20ist.. [Zugriff am 23 Jänner 2022].
- [16] Enerdata, „SECTORAL PROFILE - HOUSEHOLDS,“ 2021. [Online]. Available: <https://www.odyssee-mure.eu/publications/efficiency-by-sector/households/declining-share-space-heating-eu.html>. [Zugriff am 5 Mai 2021].
- [17] H. Saele, E. Rosenberg und N. Feilberg, State-of-the-art Projects for estimating the electricity end-use demand, Trondheim: SINTEF Energy Research, 2010.
- [18] C. Yi-Tui, „The Factors Affecting Electricity Consumption and the Consumption Characteristics in the Residential Sector-A Case Example of Taiwan,“ National Taipei University of Nursing and Health Sciences, Taiwan, 2017.
- [19] U.S. Energy Information Administration, „Use of energy explained - Energy use in homes,“ 23 Juni 2021. [Online]. Available: <https://www.eia.gov/energyexplained/use-of-energy/homes.php>. [Zugriff am 23 Oktober 2021].
- [20] J. Guo, N. Zheng Khanna und X. Zheng, Electricity Demand in Chinese Households: Findings from China Residential Energy Consumption Survey, Pacific Grove: American Council for an Energy-Efficient Economy, 2016.
- [21] H. L. Willis, „CHAPTER 24 CHARACTERISTICS OF DISTRIBUTION LOADS,“ 1997.
- [22] S. Meinecke, N. Bornhorst, L.-P. Lauven, J.-H. Menke, M. Braun, S. Drauz, C. Spalthoff, D. Cronbach, T. Kneiske, A. Klettke, J. Sprey, T. van Leeuwen, A. Moser, D. Sarajlic, C. Kittl und C. Rehtanz, SimBench - Dokumentation, DE-1.0.0 Hrsg., 2019.
- [23] T. Lange und K. Mosler, Normalverteilung und zentraler Grenzwertsatz, Berlin: Springer Berlin Heidelberg, 2017.
- [24] T. Wieland, Eine neuartige probabilistische Methode zur Betriebsmitteldimensionierung in aktiven urbanen Niederspannungsnetzen, Graz, Steiermark: TU Graz, 2016.
- [25] P. Kayal und C. Chanda, Optimal mix of solar and wind distributed generations considering performance improvement of electrical distribution network, Renewable Energy, 2014.
- [26] F. Skarke und C. Kluge, „Freie Universität Berlin,“ 28 August 2019. [Online]. Available: <https://wikis.fu-berlin.de/pages/viewpage.action?pagelId=736857812>. [Zugriff am 23 Jänner 2022].
- [27] L. Hammerstein, S. Ellßel, V. S. A. Tröbs, M. Zach und P. Hupp, „Clusteranalyse: K-Means,“ 2009. [Online]. Available: <https://www-m9.ma.tum.de/material/felix-klein/clustering/Methoden/K-Means.php>. [Zugriff am 16 Jänner 2022].
- [28] D. Liebl, „Computational Statistics: 1.2 Der EM Algorithmus zur ML-Schätzung Gaußscher Mischverteilungen,“ 2020. [Online]. Available: http://www.dliebl.com/Computational_Statistics_Script/1-2-der-em-algorithmus-zur-ml-

sch%C3%A4tzung-gau%C3%9Fscher-mischverteilungen.html#ch:EM1. [Zugriff am 16 Jänner 2022].

- [29] The MathWorks, Inc., „Cluster Using Gaussian Mixture Model,“ 2022. [Online]. Available: <https://de.mathworks.com/help/stats/clustering-using-gaussian-mixture-models.html>. [Zugriff am 2 Jänner 2022].
- [30] W. Garland, Vizualizing and Clustering Data that Includes Circular Variables, Bozeman: Montana State University, 2016.
- [31] P. Plöhn, Kategorisierung von Indizes zur Clustervalidierung, Darmstadt: Technische Universität Darmstadt, 2014.

Abbildungsverzeichnis

Abbildung 1: Potential netzdienlicher Flexibilitäten [4]	2
Abbildung 2: Neue Anforderungen an das Stromnetz von morgen [9]	3
Abbildung 3: Redispatch-Kosten in der APG [13]	7
Abbildung 4: Endverbrauchskategorien der elektrischen Energie in Haushalten	8
Abbildung 5: Analyse von aggregierten Lastprofilen [21]	9
Abbildung 6: Einzellast zu aggregierter Last [21]	10
Abbildung 7: Darstellung der MS-Netze: ländlich (links oben), vorstädtisch (rechts oben), städtisch (links unten) und gewerblich (rechts unten)	11
Abbildung 8: Schematische Darstellung der NS-Netze (von oben links nach unten rechts): 3x ländlich, 2x vorstädtisch und 1x städtisch	12
Abbildung 9: Gesamtübersicht aller zur Verfügung gestellter Profile für Sammelverbraucher	13
Abbildung 10: Dichte- und Verteilungsfunktion einer Normalverteilung	14
Abbildung 11: Beispiel Monte-Carlo-Simulation	15
Abbildung 12: Dendrogramm beim hierarchischen Clustering	18
Abbildung 13: Flussdiagramm Methodik	20
Abbildung 14: Netztopologie vom verwendeten Netzmodell im Basisszenario-0	23
Abbildung 15: Lastprofilevergleich – Jahresdauerlinie	27
Abbildung 16: Lastprofilevergleich – Jahreslastgang und Wochenlastgänge	27
Abbildung 17: Boxplot der maximalen Transformatorauslastungen	28
Abbildung 18: Matrix des Input-Datensatzes	29
Abbildung 19: Jahresverlauf der Globalstrahlung und der erzeugten mittleren Tageshöchstwert-Kurve	31
Abbildung 20: Flussdiagramm EM-Algorithmus	34
Abbildung 21: Arten der Konfiguration der Kovarianzmatrix [29]	35
Abbildung 22: Flussdiagramm Distanzmatrix-Algorithmus	36
Abbildung 23: Beispiel für Dichtefunktion der eingeteilten Cluster	42
Abbildung 24: Beispiel eines Belastungsszenarios	43
Abbildung 25: Resultierende Dichtefunktion des Beispielszenarios	43
Abbildung 26: Histogramm der Transformatorauslastung über ein Jahr	45
Abbildung 27: Einliniendiagramm des Versorgungsgebietes LV1.203	46
Abbildung 28: Ausgangslage der Dichtefunktionen der Eigenschaften aus dem Input- Datensatz	47
Abbildung 29: AIC und BIC der Kombination Uhrzeit-Transformatorauslastung	48
Abbildung 30: Verhältnis der bevorzugten Kovarianzmatrizen laut AIC und BIC bei einer Clusteranzahl von 1-10	49
Abbildung 31: Gewichtungsfaktoren der Transformatorstation MV2.101-LV1.203-Trafo 1 ...	49
Abbildung 32: Vergleich des Clusterings mit und ohne Gewichtungsfaktor beim Distanzmatrix-Algorithmus „Kontinuierlich 1“	50
Abbildung 33: Identifikation der optimalen Anzahl an Clustern beim EM-Algorithmus	51
Abbildung 34: Identifikation der optimalen Anzahl an Clustern beim Distanzmatrix- Algorithmus	52
Abbildung 35: Ellbogen-Methode bei der Auswahl der Eigenschaften bei der ersten Clusteranalyse mittels EM-Algorithmus	53
Abbildung 36: Aufteilung der Datenpunkte beim EM-Algorithmus „Kontinuierlich 1“	54
Abbildung 37: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 2	54
Abbildung 38: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 3	55

Abbildung 39: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 4	56
Abbildung 40: Aufteilung der Datenpunkte beim EM-Algorithmus „Kontinuierlich 2“	56
Abbildung 41: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 2“ mit einer Clusteranzahl 4	57
Abbildung 42: Dichtefunktionen beim EM-Algorithmus „Kontinuierlich 2“ mit einer Clusteranzahl 7	58
Abbildung 43: Aufteilung der Datenpunkte beim EM-Algorithmus „Diskret 1“	58
Abbildung 44: Dichtefunktionen beim EM-Algorithmus „Diskret 1“ mit einer Clusteranzahl 2	59
Abbildung 45: Dichtefunktionen beim EM-Algorithmus „Diskret 1“ mit einer Clusteranzahl 3	59
Abbildung 46: Dichtefunktionen beim EM-Algorithmus „Diskret 1“ mit einer Clusteranzahl 4	60
Abbildung 47: Aufteilung der Datenpunkte beim EM-Algorithmus „Diskret 2“	60
Abbildung 48: Dichtefunktionen beim EM-Algorithmus „Diskret 2“ mit einer Clusteranzahl 4	61
Abbildung 49: Dichtefunktionen beim EM-Algorithmus „Diskret 2“ mit einer Clusteranzahl 5	61
Abbildung 50: Aufteilung der Datenpunkte beim Distanzmatrix-Algorithmus „Kontinuierlich 1“	62
Abbildung 51: Dichtefunktionen beim Distanzmatrix-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 2	63
Abbildung 52: Dichtefunktionen beim Distanzmatrix-Algorithmus „Kontinuierlich 1“ mit einer Clusteranzahl 3	63
Abbildung 53: Aufteilung der Datenpunkte beim Distanzmatrix-Algorithmus „Kontinuierlich 2“	64
Abbildung 54: Dichtefunktionen beim Distanzmatrix-Algorithmus „Kontinuierlich 2“ mit einer Clusteranzahl 2	64
Abbildung 55: Dichtefunktionen beim Distanzmatrix-Algorithmus „Kontinuierlich 2“ mit einer Clusteranzahl 3	65
Abbildung 56: Aufteilung der Datenpunkte beim Distanzmatrix-Algorithmus „Diskret 1“	65
Abbildung 57: Dichtefunktionen beim Distanzmatrix-Algorithmus „Diskret 1“ mit einer Clusteranzahl 2	66
Abbildung 58: Dichtefunktionen beim Distanzmatrix-Algorithmus „Diskret 1“ mit einer Clusteranzahl 3	66
Abbildung 59: Aufteilung der Datenpunkte beim Distanzmatrix-Algorithmus „Diskret 2“	67
Abbildung 60: Dichtefunktionen beim Distanzmatrix-Algorithmus „Diskret 2“ mit einer Clusteranzahl 2	67
Abbildung 61: Dichtefunktionen beim Distanzmatrix-Algorithmus „Diskret 2“ mit einer Clusteranzahl 3	68
Abbildung 62: Vergleich zur Aufteilung der Datenpunkte	69
Abbildung 63: EM-Algorithmus mit "Diskret 1" Status Quo	71
Abbildung 64: EM-Algorithmus mit "Diskret 1" und einem Bias der Uhrzeit.....	71
Abbildung 65: EM-Algorithmus mit "Diskret 1" und einem Bias bei Monat und Uhrzeit	72
Abbildung 66: EM-Algorithmus mit "Diskret 1" und einem Bias beim Monat	72
Abbildung 67: EM-Algorithmus im Entwicklungsszenario-2	73
Abbildung 68: Prognose von drei Belastungsszenarien	76
Abbildung 69: Boxplot der Belastungsszenarien	77

Tabellenverzeichnis

Tabelle 1: Netzbeschreibende Parameter des Mittelspannungsnetzes	21
Tabelle 2: Netzbeschreibende Parameter der Niederspannungsnetze	22
Tabelle 3: Identifizierte Lastzeitreihen im verwendeten SimBench-Datensatz	24
Tabelle 4: Einteilung Tagestypen	25
Tabelle 5: Lastprofilevergleich - Kumulierter Energieverbrauch	26
Tabelle 6: Zeit-Variablen	30
Tabelle 7: Wetter-Variablen	31
Tabelle 8: Zuordnung der Wetterdaten zu Zeitreihen der Photovoltaikanlagen	33
Tabelle 9: Einflussfaktoren auf Erzeugung und Verbrauch	37
Tabelle 10: Paarweiser Vergleich - Haushalt	37
Tabelle 11: Paarweiser Vergleich - Gewerbe und Landwirtschaft	38
Tabelle 12: Einflussfaktor-Matrix E	38
Tabelle 13: Distanzmatrix mit Zeit-Variable Monat.....	40
Tabelle 14: Zusammensetzung des Versorgungsgebietes LV1.203	46
Tabelle 15: Maximalwerte des Silhouettenkoeffizient mit und ohne Gewichtungsfaktoren beim Distanzmatrix-Algorithmus	50
Tabelle 16: Kombinationen der Eigenschaften im Input-Datensatz für die erste Clusteranalyse mittels EM-Clustering	52
Tabelle 17: Anwendung der Algorithmen für die Prognose	74
Tabelle 18: Summe der Abweichungsquadrate der prognostizierten Belastungsszenarien zu den Test-Daten.....	76