

Masterarbeit

im Studiengang Master Wirtschaftswissenschaften

Entwicklung eines Software-Tools zur Analyse, Prognose und Bewertung der Prognosegüte von Zeitreihen im Strommarkt

Betreuer : Christian Puchert, M. Sc.
Externer Betreuer: Dr. Kianhwa Djie

vorgelegt am Lehrstuhl für Operations Research
der Rheinisch-Westfälischen Technischen Hochschule Aachen

von: Dipl. Inform. (FH) Nina Deeg
Dennewartstraße 5
52068 Aachen
Matr.-Nr.: 316310
Abgabetermin: 01.06.2015

Inhaltsverzeichnis

1. Einleitung	9
1.1. Ausgangssituation	9
1.2. Erkenntnisinteresse des Verfassers	10
1.2.1. Stromprodukte und Handel	10
1.2.2. Motivation	13
1.3. Ziele	14
1.4. Aufbau der Arbeit	15
1.5. Zusammenfassung	15
 2. Allgemeiner Teil	 17
2.1. Theoretischer Hintergrund, Aktueller Forschungsstand	17
2.1.1. Traditionelle Prognoseverfahren	17
2.1.2. Verfahren der künstlichen Intelligenz	19
2.2. Zeitreihenprognose in FirstSale	19
2.3. Anforderungsanalyse	20
2.4. Fazit	21
2.5. Vorgehensweise	21
2.6. Zusammenfassung	22
 3. Daten-Analysatoren	 23
3.1. Einflussfaktoren auf den Energieverbrauch	23
3.2. Analysatoren	24
3.2.1. Streuungs-Analysator	24
3.2.2. Korrelations-Analysator	27
3.2.3. Periodizitäts-Analysator	28
3.3. Zusammenfassung	29

4. Modellierung von Optimierungsproblem und Algorithmus	31
4.1. Optimierungsproblem	31
4.1.1. Zielfunktion	31
4.1.2. Definition der Prognosefunktion	34
4.2. Algorithmus	37
4.2.1. Struktur des Algorithmus	38
4.2.2. Ermittlung der Tagesgruppierung	38
4.2.3. Festlegung des Fensters	40
4.2.4. Überprüfung der Tagesgruppe	42
4.2.5. Berechnung der Prognosezeitreihe	43
4.3. Zusammenfassung	43
5. Analyse von Datenbestand und Algorithmus	45
5.1. Deskriptive Datenanalyse	45
5.1.1. Kundenprofil	45
5.1.2. Datenarchiv	46
5.1.3. Korrelationsanalysen	47
5.2. Methodik Auswertung	49
5.2.1. Konfiguration der Tagesgruppierungen	49
5.2.2. Konfiguration der Originalprognose	51
5.2.3. Auswertungsverfahren der Prognosegüte	52
5.3. Zusammenfassung	55
6. Ergebnisse der Algorithmusanalyse	56
6.1. Auswertung der Prognosegüte	56
6.2. Finanzielle Bewertung	57
6.3. Diskussion	58
6.3.1. Zeitreihen mit Trendanteilen	59
6.3.2. Zeitreihen mit abweichendem Verbrauchsprofil zu unterschiedlichen Tageszeiten	60
6.3.3. Zeitreihen mit zufälligem Verbrauch	61
6.4. Zusammenfassung	61
7. Überarbeitung Algorithmus	63
7.1. Anpassung Modell	63
7.2. Auswertung und Diskussion	64

7.3. Zusammenfassung	65
8. Prognostizierbarkeit von Zeitreihen	67
8.1. Idee	67
8.2. Herangehensweise	68
8.3. Umsetzung	70
8.4. Auswertung	70
8.4.1. Parameter des Prognosealgorithmus	71
8.4.2. Einfluss der Korrelationskoeffizienten	74
8.5. Diskussion	74
8.6. Prognose mit neuer Zielfunktion	76
8.7. Zusammenfassung	78
9. Zusammenfassung und Ausblick	80
9.1. Zusammenfassung	80
9.2. Limitationen der Arbeit und zukünftige Forschung	81
A. Anhänge	83
B. Literatur	92
C. Versicherung	94

Tabellenverzeichnis

1.1. Übersicht über Stromhandelsplätze und Produkte	11
5.1. Übersicht über Tageseigenschaften	50
5.2. Beispiel der Tageseigenschaften des 06.01.2015 anhand von unterschiedli- chen Kalendern	50
5.3. Konfiguration der Originalprognose	51
6.1. Übersicht über Prognosefehler der Originalprognose und der verschiedenen Algorithmusstufen	57
8.1. Datenbeschreibung: Quantile der unabhängigen Variablen	69
8.2. Ergebnisse der Regressionsanalyse	71
A.1. Übersicht über alle Tagesgruppen Schemata nach Kapitel 5.2.1	91

Abbildungsverzeichnis

1.1. Wochen-Lastgang einer Schule	9
1.2. Preisentwicklung eines Peak-Produkts für Februar 2014 über 5 Monate . .	12
1.3. Preisentwicklung von Spot-Produkten im Februar 2014	13
1.4. Preisentwicklung der Ausgleichsenergie im Februar 2014	14
3.1. Histogramm von Tagesenergie und Maximalleistung einer Schule an Wochenenden	25
3.2. Histogramm von Tagesenergie und Maximalleistung einer Schule an Werktagen	25
3.3. Kurve der Streuung für eine Schule	26
3.4. Streudiagramm der Tagesenergie zweier Zeitreihen mit verschiedenen Einflussfaktoren	27
3.5. Autokorrelationsdiagramm einer Schule	29
4.1. Energieverbrauch eines Schwimmbades an Wochentagen zu unterschiedlichen Kalenderwochen im Jahr 2013	36
5.1. Kategorisierung der Zeitreihen nach Branchen	46
5.2. Übersicht über das Datenarchiv	48
5.3. Ergebnisse der Korrelationsanalysen über das Datenarchiv	49
5.4. Schematische Ansicht der finanziellen Bewertung einer Differenzzeitreihe .	54
6.1. Relative Prognosegüte im Vergleich neuer Algorithmus und alte Prognose .	58
6.2. Finanzielle Bewertung der Fehlmengen mit Spotmarktpreisen	59
6.3. Zeitreihe mit Trendanteil und ihre Prognose	60
6.4. Heatmap der Prognosequalität nach Uhrzeit und Zeitfenstergröße einer Schule für einen Prognosetag	61
6.5. Zeitreihe ohne erkennbare Zyklen und tägliche Verbrauchsmuster	62
7.1. Vergleich relativer Prognosegüte zwischen bester Algorithmusstufe und dem Algorithmus mit viertelstündlichen Prognoseparametern	65

7.2. Finanzielle Bewertung der Fehlmengen mit Spotmarktpreisen nach Berechnung von viertelstündlichen Prognoseparametern	66
8.1. Grafik von Prognosegüte und Betrag der Autokorrelationswerte bei einem Versatz von einer Woche	68
8.2. Darstellung der Regressionskurven, abgeleitet aus Tabelle 8.2	72
8.2. Darstellung der Regressionskurven, abgeleitet aus Tabelle 8.2	73
8.3. Darstellung des Zusammenhangs von Zeitfensterbreite x , Streuung y und Archivtiefe $z = 1$ mit dem Prognosefehler	73
8.4. Darstellung der Regressionskurven der Korrelationskoeffizienten, abgeleitet aus Tabelle 8.2	75
8.5. Vergleich relativer Prognosegute zwischen Algorithmus mit viertelstündlichen Prognoseparametern und der Optimierung nach neuer Zielfunktion . .	78
8.6. Finanzielle Bewertung der Ergebnisse im Vergleich	79
A.1. Darstellung des Zusammenhangs von Zeitfensterbreite x , Streuung y und Archivtiefe $z \in \{1, 2\}$ mit dem Prognosefehler	84

1. Einleitung

Die Masterarbeit widmet sich der Prognose von Energieverbräuchen auf Basis historischer Verbrauchsdaten. Diese wurden regelmäßigen Zeitintervallen gemessen und liegen in Form von diskreten Zeitreihen vor. Abbildung 1.1 zeigt exemplarisch den zeitlichen Verlauf der abgenommenen elektrischen Leistung an einer Schule über eine Woche, in der Energiewirtschaft auch Lastgang genannt.

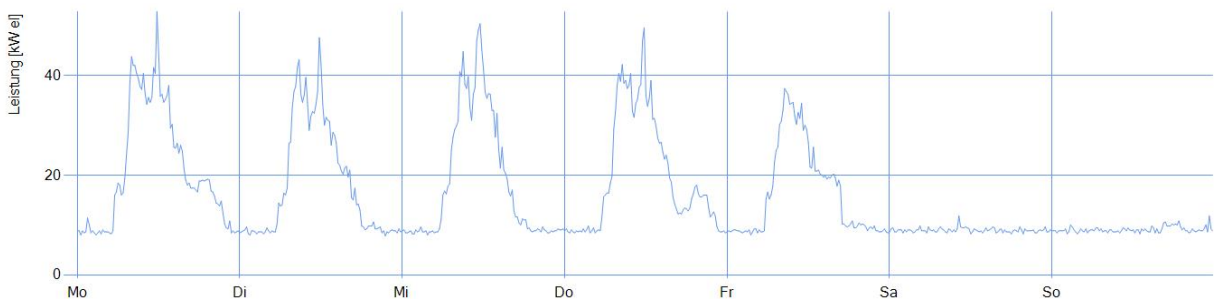


Abbildung 1.1.: Wochen-Lastgang einer Schule

Bereits mit bloßem Auge ist eine gewisse Regelmäßigkeit in der Kurve zu erkennen. Nachts und am Wochenende ist der Verbrauch gleichbleibend niedrig. Dieser ständig vorhandene Verbrauch ist die Grundlast. Die Verbräuche an Wochentagen sind recht ähnlich, wobei die Spitzenlast am Freitag geringfügig nach unten abweicht.

Das automatisierte Erkennen dieser und anderer Charakteristika von Zeitreihen sowie die Prognose von Lastgängen sind Inhalt der Arbeit.

1.1. Ausgangssituation

Diese Arbeit wird im Rahmen des Prognosemoduls der Klafka & Hinz Energie-Informationssysteme GmbH verfasst. Klafka & Hinz ist ein mittelständisches Unternehmen mit Sitz in Aachen, das Softwarelösungen für den Energiemarkt erstellt und vertreibt. Dabei liegt der Fokus nicht auf Energieerzeugung und Netzbetrieb, sondern auf Abbildung von Geschäftsprozessen im Energiemarkt, Datenaustausch zwischen den verschiedenen Marktteilnehmern, Abrechnungen, Prognosen und Optimierungen. Solche Systeme werden auch als

Energiedaten-Management-Systeme, kurz EDM, bezeichnet. Das hauseigene EDM System für Energievertriebe heißt *FirstSale*.

Lastgangdaten fallen für Energieverbraucher an, die einen Zähler mit registrierender Leistungsmessung besitzen. Dazu ist ein minimaler Verbrauch von 100 MWh Strom oder 1,5 GWh Gas notwendig (Deutscher Bundestag 2013, Deutscher Bundestag 2010). Im Abstand von 15 Minuten wird die zu diesem Zeitpunkt abgenommene Leistung des Verbrauchers gemessen. Dies ergibt 96 Werte pro Tag. Der Verbrauch wird täglich am Zähler ausgelesen und an *FirstSale* übermittelt. Diese Daten bilden die Basis für die Prognose zukünftiger Lastgänge.

1.2. Erkenntnisinteresse des Verfassers

Die Erstellung von Prognosezeitreihen ist für Energie-Beschaffungsplanung und eine Angebotslegung seitens eines Energievertriebs notwendig. Energie wird über verschiedene Wege beschafft. Deren Preise orientieren sich weitestgehend an den großen Energiebörsen, weswegen an dieser Stelle der Handel an Termin- und Spotmarkt beschrieben ist. Im Folgenden wird der Handel mit Energieprodukten exemplarisch am Stromhandel der EEX (European Energy Exchange) erklärt. Anschließend wird die Motivation zur Erstellung dieser Arbeit näher erläutert.

1.2.1. Stromprodukte und Handel

Strom als Handelsgut unterscheidet sich in seinen Eigenschaften stark von anderen handelbaren Waren: Die Speicherung von Elektrizität ist schwierig, Störungen werden mit nahezu Lichtgeschwindigkeit im Netz übertragen und Abweichungen in Netzspannung oder Frequenz werden nur eingeschränkt toleriert (Shively und Ferrare 2008). Hieraus ergibt sich die Notwendigkeit, dass Stromerzeugung und Verbrauch zu jedem Zeitpunkt übereinstimmen müssen, um die Stabilität der Stromnetze zu garantieren. Jedoch hängt der Verbrauch von vielen Einflussfaktoren ab, die selbst teils besser (z.B. Kalendereigenschaften, Sonnenaufgang und -untergang), teils schlechter (Wetter, menschliches Verhalten) prognostizierbar sind (siehe Kapitel 3.1).

Das bedeutet aber auch, dass je kurzfristiger eine Prognose ist, desto kleiner ist auch der Bereich, in dem mögliche Einflussfaktoren schwanken. Dies spiegelt sich auch im Stromhandel wieder: An der EEX gibt es getrennte Handelsplätze für lang- und kurzfristige Strombeschaffung, an denen unterschiedliche Stromprodukte gehandelt werden. Jedes

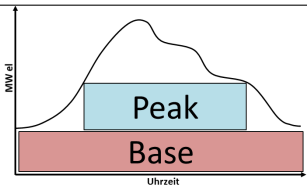
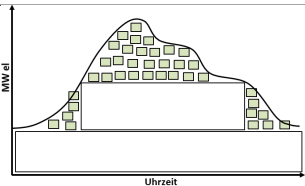
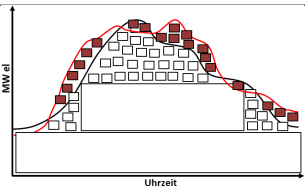
Ort	Terminmarkt	Spotmarkt	Ausgleichsenergie- markt
Produkte	Base, Peak, Offpeak	Spot	Ausgleichsenergie
Zeitintervall	Produkte werden für folgende Zeiträume gehandelt: ein Tag, Wochenende, Monat, Quartal, Jahr	15-minütige bis mehrstündliche Zeiträume werden gehandelt	15 Min
Granularität	1 MW	0,1 MW	0,1 MW
Lieferzeitpunkt	6 Jahre bis 1 Tag nach Kauf	45min bis 24h	sofort
Volatilität	schwach	mittel	stark
Beispiel			

Tabelle 1.1.: Übersicht über Stromhandelsplätze und Produkte

Stromprodukt liefert eine konstante Menge Strom. Sie unterscheiden sich in Zeitraum, Tageszeit und Fälligkeit der Lieferung.

Base-Produkt Ein Base-Produkt entspricht einer Stromlieferung für mindestens einen Tag.

Peak-Produkt Ein Peak-Produkt entspricht einer Lieferung im Zeitraum von 08:00 Uhr bis 20:00 Uhr an mindestens einem Tag.

Offpeak-Produkt Ein Offpeak-Produkt entspricht einer Stromlieferung von 0:00 Uhr bis 8:00 Uhr und von 20:00 Uhr bis 24:00 Uhr an mindestens einem Tag.

Spot-Produkt Ein Spot-Produkt entspricht der Stromlieferung innerhalb eines Zeitintervalls von 15 Minuten bis mehreren Stunden an einem bestimmten Tag.

Ausgleichsenergie Die Ausgleichsenergie schließt die Lücke zwischen dem prognostizierten und dem tatsächlichen Verbrauch.

Wie bereits erwähnt, gibt es an der EEX unterschiedliche Handelsplätze für kurz- und langfristigen Stromhandel: den Termin- und den Spotmarkt. Tabelle 1.1 grenzt diese voneinander ab und ordnet die eingeführten Stromprodukte den Handelsplätzen zu. Am Terminmarkt werden Produkte bis zu sechs Jahre in die Zukunft gehandelt, Fälligkeit und

Erfüllung werden auf einen späteren Zeitpunkt vereinbart. Hier sichern sich die Händler langfristig gegen Preisrisiken ab (European Energy Exchange AG 2014a).

Abbildung 1.2 zeigt den Preisverlauf eines Monats-Peak-Produktes. Man kann sehen, dass die Schwankungen des Strompreises größer werden, je näher der Liefertermin rückt.

Am Spotmarkt wird Strom für denselben oder den nächsten Tag gehandelt. Er ist gut geeignet, um anhand von Kurzfristprognosen das Produktportfolio durch den Kauf und Verkauf von Spotprodukten optimieren, da Strom für kleinere Zeitintervalle und in kleineren Mengen gehandelt werden darf. Dadurch ermöglicht er die Reaktion auf kurzfristige Ereignisse wie Wetterschwankung oder Kraftwerksausfälle. Handelsgeschäfte werden innerhalb der nächsten 24 Stunden realisiert, frühestens aber 45 Minuten nach Kauf (European Energy Exchange AG 2014a).

Abbildung 1.3 zeigt den Verlauf der Preise am EEX-Spotmarkt im Februar 2014. Der niedrigste und höchste Preis, der für eine MWh bezahlt wurde, sind jeweils als Kurve aufgetragen. Im Vergleich zum Terminmarkt (Abbildung 1.2) kann man gut die höheren Preisschwankungen erkennen. Folglich beinhaltet die Beschaffung am Spotmarkt ein deutlich höheres Preisrisiko.

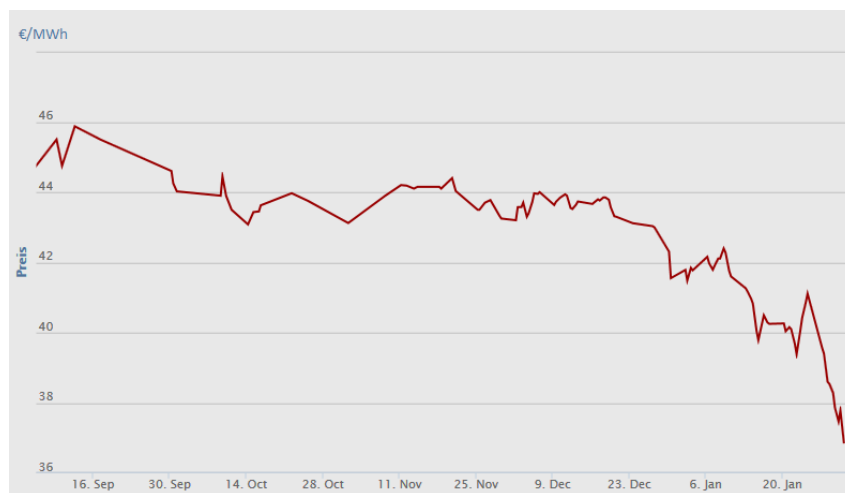


Abbildung 1.2.: Preisentwicklung eines Peak-Produkts für Februar 2014 über 5 Monate (European Energy Exchange AG 2014b)

Die extremsten Preise für Strom werden in der Ausgleichsenergie erreicht. Abbildung 1.4 zeigt eine Entwicklung der Tagesmaximal- und Minimalpreise innerhalb eines Monats. Ausgleichsenergie wird immer dann benötigt, wenn der gemessene Verbrauch von der Prognose abweicht. Preissetzung, Ermittlung der Ausgleichsmengen und Abrechnung finden am Folgetag statt.

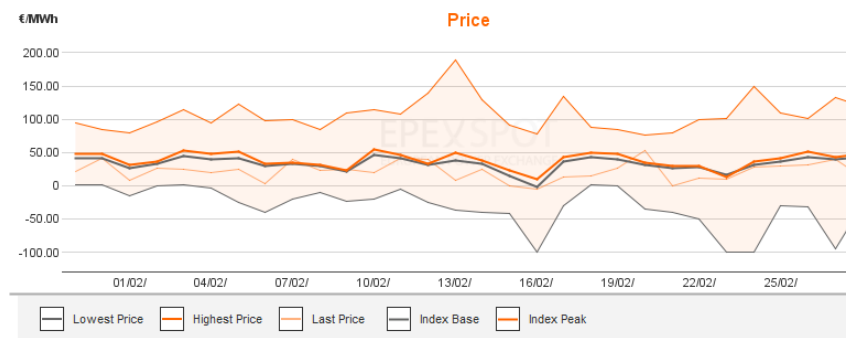


Abbildung 1.3.: Preisentwicklung von Spot-Produkten im Februar 2014
(EPEX Spot SE 2014)

Nicht nur aus wirtschaftlichen Gründen sollte die Nutzung von Ausgleichsenergie unter allen Umständen vermieden werden. Zur Wahrung der Netzstabilität ist jeder Lieferant verpflichtet, Beschaffung und Verbrauch in Einklang zu bringen.

1.2.2. Motivation

Die Kunden von Klafka & Hinz beschaffen ihren Strom, wie im vorherigen Kapitel beschrieben, an verschiedenen Energiehandelsplätzen. Dazu benötigen sie möglichst genaue Prognose des Energieverbrauches ihrer Stromkunden, die sie mit dem Prognosemodul von *FirstSale* erstellen.

Die Software bietet verschiedene Verfahren zur Zeitreihenprognose an mit vielfältigen Optionen zur Konfiguration. Die große Anzahl der vorhandenen Einstellmöglichkeiten erschweren allerdings die effiziente Nutzung von *FirstSale*. Daraus resultierend wird die Konfiguration seitens der Energievertriebe vernachlässigt und schlechtere Prognosen werden in Kauf genommen. Den Vorgang der Prognoseerstellung von Zeitreihen bei den Kunden soll in Zukunft besser unterstützt und vereinfacht werden.

In Abschnitt 1.2.1 wurde gezeigt, dass das Preisrisiko zunimmt, je näher der Liefertermin rückt und dass eine ungenaue Prognose durch Ausgleichsenergie zusätzlich hohe Kosten verursachen kann. Folglich wird sich eine genauere Zeitreihenprognose an mehreren Stellen positiv bemerkbar machen.

Ein weiterer Anwendungsfall für einen Energievertrieb ist die Risikobewertung für Neukunden. In der Regel liegen zu diesem Zeitpunkt die Lastgänge mehrerer Monate vor, auf deren Basis ein Angebot für einen Stromtarif erstellt werden soll. Es ist bekannt, dass es gewisse Faktoren den Stromverbrauch beeinflussen. Sind diese sowie der Anteil zufälliger Schwankungen für eine Zeitreihe bekannt, kann eine Aussage darüber getroffen werden,

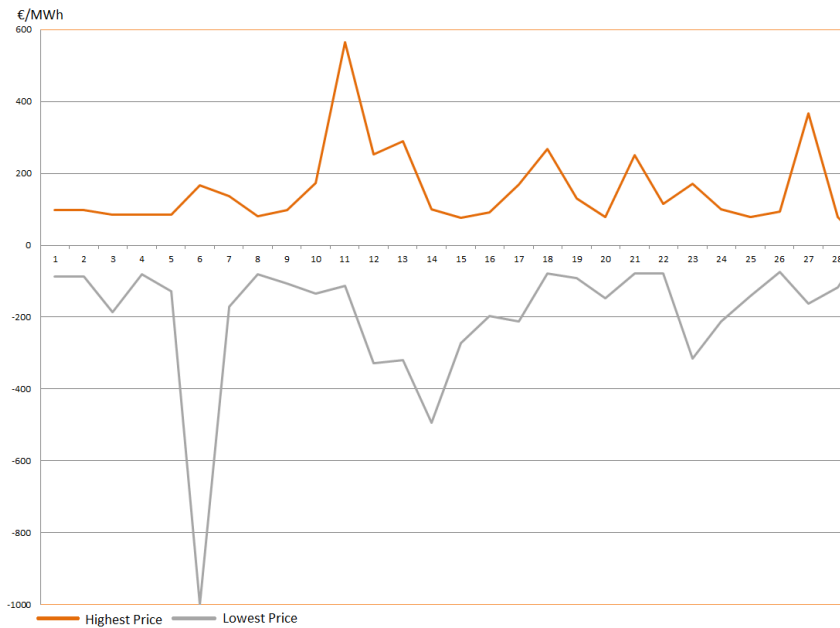


Abbildung 1.4.: Preisentwicklung der Ausgleichsenergie im Februar 2014
(50Hertz Transmission GmbH 2014)

wie gut ein Lastgang prognostizierbar ist. Dieser Vorgang wird vom Prognosemodul nur unzureichend unterstützt.

Eine realistischere Prognose sollte die Beschaffungsplanung beeinflussen und damit Kosten im Einkauf senken. Dies wiederum ermöglicht, Stromtarife individuell auf den Kunden zuzuschneiden. Wird der Verbrauch eines Einzelnen genauer prognostiziert, ist es möglich, den daraus resultierenden Kostenvorteil in Form von niedrigeren Tarifen weiterzugeben. Ein positiver Einfluss auf Kundengewinnung und Kundenbindung ist denkbar. Ist eine Verbrauchsstelle schlecht prognostizierbar, kann das so entstandene Preisrisiko in Form einer Risikoprämie an den Kunden weitergegeben werden.

1.3. Ziele

Das Ziel dieser Arbeit ist es, die Prognose von Zeitreihen zu verbessern und das Preisrisiko einer Zeitreihe zu bestimmen. Eine Verbesserung der Prognose tritt dann ein, wenn die Abweichung zum tatsächlichen Verbrauch verringert wird und dadurch Kosten bei der Energiebeschaffung eingespart werden.

Zur Prognose von Zeitreihen sind in der eingesetzten Software bereits verschiedene Verfahren vorhanden, deren Konfiguration Schwierigkeiten bereitet (siehe Kapitel 2.2). Das

Ziel dieser Arbeit ist es, ein Verfahren zu entwickeln, welches gleichzeitig die Konfiguration der Prognose und die Prognose selbst beinhaltet.

Weiterhin sollen Eigenschaften einer Zeitreihe analysiert und quantifiziert werden, die eine Aussage über ihre Prognostizierbarkeit treffen. Darauf aufbauend soll ein Verfahren zur Risikobewertung von Zeitreihen erarbeitet werden, damit die so gewonnenen Informationen bei der Angebotslegung berücksichtigt werden können. Dadurch würde ein Energievertrieb in die Lage versetzt, diese Preisrisiken an seinen Kunden weiterzureichen.

1.4. Aufbau der Arbeit

Im nächsten Kapitel wird die Ausgangssituation noch einmal näher erläutert und die Anforderungen an eine neue Prognose definiert. Anschließend werden verschiedene Verfahren zur Zeitreihenprognose vorgestellt und diskutiert. Die in der Arbeit gewählte Vorgehensweise und die dabei eingesetzten Tools werden anschließend näher erläutert.

Kapitel 3 beschäftigt sich mit den Einflussfaktoren auf Zeitreihen und deren Analyse. Aufbauend darauf wird in Kapitel 4 das der Prognose zugrundeliegende Optimierungsproblem aufgestellt, welches es zu lösen gilt. Anschließend wird ein neu gefundener Algorithmus vorgestellt, der Zeitreihenanalyse und Prognose umfasst.

Bevor der vorgestellte Algorithmus getestet wird, werden zunächst die Datengrundlage und die Methoden, nach denen das Ergebnis evaluiert wird, beschrieben (Kapitel 5).

In Kapitel 6 wird anschließend die Performance des neuen Algorithmus ausgewertet und dessen Schwächen werden diskutiert. Daraufhin wurde der Algorithmus verändert und neu ausgewertet (siehe Kapitel 7).

Das darauffolgende Kapitel 8 beschäftigt sich mit der Prognostizierbarkeit einer Zeitreihe und der Vorhersage ihrer Prognosequalität. Zuletzt werden in Kapitel 9 die Ergebnisse noch einmal zusammengefasst, evaluiert und Impulse für weitere Forschung an diesem Thema gegeben.

1.5. Zusammenfassung

Das vorangegangene Kapitel gibt einen Einstieg in das Thema der Zeitreihenprognose. Im Rahmen dieser Arbeit soll das Prognosemodul von *FirstSale*, welches von vielen Energievertrieben zu Zeitreihenprognose genutzt wird, verbessert werden.

Energie wird an verschiedenen Handelsplätzen mit unterschiedlichen Zeitperspektiven gehandelt. Eine gute Zeitreihenprognose ist dabei die Grundlage der Energiebeschaffung.

Differenzen zwischen Prognose und tatsächlichen Verbräuchen müssen zu hohen Preisen ausgeglichen werden. Ungenaue Energieprognosen gefährden die Netzstabilität und damit die Versorgungssicherheit und bedeuten gleichzeitig ein finanzielles Risiko in der Energiebeschaffung.

Das Prognosemodul von *FirstSale* besitzt umfangreiche Prognosefunktionalitäten, wird aber von der Energievertrieben wegen des hohen Konfigurationsaufwands nicht optimal eingesetzt. Eine Risikobewertung von Neukunden ist mit dem aktuellen Funktionsumfang nur eingeschränkt möglich. Ziel dieser Arbeit ist es, einen selbstkonfigurierenden Prognosealgorithmus zu finden, der bessere Ergebnisse liefert, und eine automatisierte Risikobewertung auf Basis einer Zeitreihe zu ermöglichen.

2. Allgemeiner Teil

Im folgenden Kapitel wird zunächst ein Einblick in bekannte Prognoseverfahren von Energiezeitreihen geworfen und die in *FirstSale* eingesetzten Verfahren benannt. Daraufhin werden Anforderungen an ein neues Prognoseverfahren zusammengefasst und Ideen für ein eigenes Verfahren entwickelt. Abschließend wird die gewählte Vorgehensweise bei der Erstellung des neuen Verfahrens erläutert.

2.1. Theoretischer Hintergrund, Aktueller Forschungsstand

Dieses Kapitel gibt eine Einführung in die Prognose von Energiezeitreihen. In ihrem Review über verwendete Prognoseverfahren unterscheiden Hahn u. a. (2009) zwischen traditionellen Prognoseverfahren und Verfahren der künstlichen Intelligenz.

2.1.1. Traditionelle Prognoseverfahren

Unter den traditionellen Prognoseverfahren sind zumeist statistische Verfahren zu finden. Ihnen gemeinsam ist, dass sie ein explizites Modell auf Basis von verschiedenen Einflussfaktoren formulieren nach dem eine Zeitreihe prognostiziert wird. Es werden hier regressionsbasierte Modelle und zeitreihenbasierte Methoden unterschieden (Hahn u. a. 2009).

Regressionsbasierte Modelle

Die Leistungsabnahme von Energieverbrauchern ist von verschiedenen Faktoren wie Wiedereinflüssen und kalendarischen Ereignissen abhängig (siehe Kapitel 3.1). Zusammenhänge zwischen einer Zeitreihe und mehreren Variablen werden mittels mehrdimensionaler Regressionsanalyse geschätzt. Dazu werden zunächst eine Reihe von Einflussfaktoren identifiziert und ein Modell in Form einer Regressionsfunktion formuliert:

$$E = \beta_0 + \beta_1 f(x_1) + \beta_2 f(x_2) + \cdots + \beta_n f(x_n) \quad (2.1)$$

Mit E wird der Energieverbrauch beschrieben, x_1, \dots, x_n sind dabei die verschiedenen Einflussfaktoren. Die Funktion f definiert dabei die Art des Zusammenhangs.

Häufig wird die Abhängigkeit von Stromzeitreihen als lineares Modell aus mehreren Faktoren abgebildet, da sie einfach verständlich sind, und Zusammenhänge korrekt erkennen und modellieren können. Hor u. a. (2005) bilden beispielsweise die Wetterabhängigkeit von Zeitreihen in England aus Temperatur, Windstärke, Sonnenstunden und zwei Maßen für die Luftfeuchtigkeit als lineares Modell ab.

$$E = \beta_0 + \beta_1 x_{\text{Temp}} + \beta_2 x_{\text{Wind}} + \beta_3 x_{\text{Sonn}} + \beta_4 x_{\text{Feu1}} + \beta_5 x_{\text{Feu2}} \quad (2.2)$$

Als Funktion f wird in diesem Modell die Identität verwendet, sie bildet den Wert auf sich selbst ab. Auch andere Abhängigkeiten sind denkbar: beispielsweise wird eine Abhängigkeit von der Außentemperatur oft quadratisch Modelliert (Hahn u. a. 2009); Bruhns u. a. (2005) modellieren den Einfluss der Jahreszeiten als Fourierreihe abhängig von verschiedenen Tagestypen.

Die Koeffizienten in derartigen Modellen lassen sich auf der vorhandenen Datenbasis mit der Methode der kleinsten Quadrate (siehe Auer 2011) vergleichsweise einfach und schnell schätzen. Anschließend werden die Faktoren in die Zukunft fortgeschrieben und der Energieverbrauch mittels der im Modell festgelegten Koeffizienten berechnet.

Zeitreihenbasierte Methoden

Bei zeitreihenbasierten Verfahren ist hingegen die Zeitreihe selbst die Datengrundlage für eine Prognosefunktion. Ein beliebtes Verfahren zur Prognose ist das *autoregressive moving average* Verfahren, ARMA genannt (Hahn u. a. 2009). Es betrachtet stationäre Zeitreihen als Realisation einer Folge von identisch verteilten Zufallsvariablen. Dabei ist ϵ_t ein Störterm mit einem Erwartungswert von 0, auch weißes Rauschen genannt.

$$y_t = \sum_{i=1}^n a_i y_{t-i} + \sum_{j=1}^m b_j \epsilon_{t-j} + \epsilon_t \quad (2.3)$$

Der erste Summand beschreibt die Annahme, dass die Werte einer Zeitreihe y_t von ihren vergangenen Werten y_{t-i}, \dots, y_{t-n} abhängen. Damit setzen sich Störungen in der Zeitreihe weiter fort. Der zweite Summand beschreibt die Messfehler, die dabei auftreten.

können. Die Koeffizienten a_i und b_i können durch eine Regressionsanalyse bestimmt werden (Auer 2011; Bastian 1985; Hahn u. a. 2009)

2.1.2. Verfahren der künstlichen Intelligenz

Seit einigen Jahren werden erfolgreich Verfahren der künstlichen Intelligenz zur Prognose von Zeitreihen eingesetzt, allen voran die neuronalen Netze (NN). Hier wird ein Netzwerk nach Vorbild des menschlichen Gehirns aufgebaut. Die Knoten des Netzwerks sind den Neuronen im Gehirn nachempfunden. Sie sind untereinander durch gewichtete Kanten verknüpft und besitzen jeweils eine Aktivierungsfunktion.

Ein Knoten besitzt in der Regel mehrere eingehende Kanten, über die numerische Informationen eingehen. Diese werden anhand ihrer Gewichte addiert und als Argument an die Aktivierungsfunktion übergeben. Diese ist entweder eine Schwellwert- oder eine Sigmoidfunktion und steuert die Ausgabe eines Knoten (Russell u. a. 1995).

Neuronale Netze sind üblicherweise in Schichten organisiert, wobei Knoten in der selben Schicht nicht miteinander kommunizieren. Sie bekommen ihre Information aus der vorherigen Schicht und geben sie nur an die nächste weiter.

Um eine Zeitreihe mit einem solchen Netzwerk zu prognostizieren muss es zunächst trainiert werden. Vor Beginn des Trainings werden wie zuvor Faktoren identifiziert werden, welche die zu prognostizierende Zeitreihe beeinflussen. Diese müssen ebenfalls als Zeitreihe vorliegen. Beim Training werden historische Daten dieser Faktoren als Vektor in die oberste Schicht des Netzwerks hereingegeben. Als Ausgabe in der untersten Schicht werden Archivdaten der Zeitreihe angelegt (Hippert u. a. 2001).

Das Training besteht darin, dass Werte für die Gewichte der Kanten so festgelegt werden, dass der Fehler zwischen erwarteter und tatsächlicher Ausgabe minimiert wird. Mit einem so konfigurierten Netzwerk kann dann eine Prognose in die Zukunft erfolgen, sofern die Werte der Einflussfaktoren festgelegt sind (Hahn u. a. 2009; Hippert u. a. 2001; Russell u. a. 1995).

2.2. Zeitreihenprognose in FirstSale

Das Prognosemodul von *FirstSale* bietet verschiedene Verfahren zur Zeitreihenprognose an, die hier kurz erläutert werden.

Typische Tage Die Prognosezeitreihe eines Tages wird aus der Durchschnittszeitreihe historischer, ähnlicher Tage gebildet. Die Ähnlichkeit wird z.B. über Tagestypen,

Jahreszeiten oder Temperatur festgelegt. Die zugrundeliegenden Tage können je nach zeitlicher Distanz exponentiell gewichtet werden. Die Gewichtungsfunktion ist parametrierbar.

Regressionsverfahren Die Prognose wird über multiple lineare Regression berechnet. Es können bis zu drei externe Faktoren zur Regression berücksichtigt werden.

Neuronale Netze Zur Prognose wird ein mehrschichtiges Backpropagation Netzwerk mit sigmoider Anregungsfunktion verwendet (siehe Russell u. a. 1995). Externe Einflussfaktoren, die Anzahl der Trainingszyklen und die Lerngenauigkeit muss dazu festgelegt werden.

Jedes der Verfahren kann pro zu prognostizierender Zeitreihe individuell konfiguriert werden. Gerade bei Neuronalen Netzen aber auch bei Nutzung des Regressionsverfahrens ist dazu ein tieferes Wissen um Einflussfaktoren und die Funktionsweise des Verfahrens notwendig.

Aus diesem Grund ist bei den meisten Kunden das Verfahren *typische Tage* im Einsatz. Die Konfiguration (siehe auch Kapitel 5.2.1) ist nicht auf die einzelne Zeitreihe abgestimmt, sondern für alle identisch.

2.3. Anforderungsanalyse

Aus der Problematik der umfangreichen Parametrisierung und der Abneigung der Kunden, dies manuell durchzuführen, ergibt sich die Anforderung an ein neues Prognoseverfahren, dass basierend auf dem zur Verfügung gestellten Datenarchiv die Prognose eigenständig konfiguriert.

Im Gegensatz zu der bereits implementierten Vorgehensweise der neuronalen Netze, welches sich für den Benutzer als Blackbox darstellt, sollte ein neues Prognoseverfahren möglichst leicht Verständlich und dessen Entscheidungen für den Benutzer nachvollziehbar sein. Ebenfalls sollte es die Möglichkeit anbieten, händisch in die Parameter der Prognose einzugreifen. Dies erhöht die Akzeptanz beim Kunden.

Die Anforderungen an eine Risikobewertung sind sehr ähnlich: Einfachheit des Verfahrens und Nachvollziehbarkeit der Entscheidung. Auf diese Weise kann dem Energiekunden eine Preissetzung transparent gemacht und die Notwendigkeit eines etwaigen Risikoabschlages untermauert werden.

2.4. Fazit

Es stellt sich nun die Frage, welche Methode zur Zeitreihenprognose als neues Prognoseverfahren für die *FirstSale* den Anforderungen gerecht wird.

Obwohl die Methode der neuronalen Netze gute Ergebnisse liefert und für Lang- und Kurzfristprognosen gut geeignet ist (Hippert u. a. 2001), wird sie bereits jetzt von den Kunden abgelehnt. Eine Nachvollziehbarkeit der Entscheidungen, wie sie ein solches Netzwerk trifft, ist nicht gegeben.

Eine weitere Möglichkeit wäre es, das bereits angebotene Regressionsverfahren zu erweitern. Um den Verbrauch einzelner Zeitreihen vorherzusagen, müssten mehrere Faktoren und damit auch eine Kalenderkomponente einbezogen werden. Dadurch würde das Prognosemodul sehr schnell an Komplexität gewinnen, da zur korrekten Abbildung der Abhängigkeiten nichtlineare Regressionsfunktionen herangezogen werden müssten (siehe auch Bruhns u. a. 2005).

Auch ein ARMA Modell verletzt die Anforderung der Einfachheit seitens der Kunden. Zusätzlich ist es eher geeignet aggregierte Zeitreihen auf makroökonomischem Niveau, wie den Stromverbrauch einer ganzen Region zu prognostizieren (Bruhns u. a. 2005).

Aus diesen Gründen wird im Folgenden das Verfahren der *typischen Tage* erweitert. Es ist ein einfaches *lookup* Verfahren, welches den historischen Verlauf einer Zeitreihe berücksichtigt. Saisonalität und Wetterabhängigkeit können hier abgebildet werden, indem bestimmte Tage von der Prognose ausgeschlossen werden.

2.5. Vorgehensweise

Um die in Kapitel 1.3 gesetzten Ziele zu erreichen und den Anforderungen gerecht zu werden, wird wie folgt vorgegangen.

Zunächst sollen Analysatoren programmiert werden, die jene Abhängigkeiten von Einflussfaktoren, die mit dem menschlichen Auge erkennbar sind, quantifizieren können. Es wird erwartet, dass eine Auswertung mit einfachen statistischen Methoden wie Streuungs- und Regressionsanalysen bereits gute Ergebnisse liefert in Bezug auf die Ähnlichkeit von Tageslastgängen, Einflüsse externer Größen und Periodizität einer Zeitreihe.

Im nächsten Schritt müssen die Ergebnisse der Analyse in ein Modell übersetzt und ein Algorithmus zur Implementierung entwickelt werden. Dieser ergibt dann das neue Prognoseverfahren.

Um die Ergebnisse dieser Arbeit mit anderen vergleichbar zu machen, wird das Datenarchiv beschrieben, mit Hilfe dessen der Algorithmus evaluiert wird und die Auswertungsmethoden vorgestellt. Mit Hilfe dieses Archivs werden Prognosen mit neuer und alter Konfiguration für beliebige Zeitreihen erstellt und mit den gemessenen Lastgängen verglichen.

Das Ergebnis der Auswertung sollte anhand von Kosten, die durch eine Fehlprognose entstehen, bewertet werden, um den Nutzen des neuen Verfahrens zu untermauern. Die beiden Verfahren werden dann gegenübergestellt und diskutiert.

Die Prognostizierbarkeit einer Zeitreihe kann anschließend bewertet werden, indem gefundene Einflussfaktoren der erreichten Prognosegüte des neuen Verfahrens gegenübergestellt werden. Um hier Zusammenhänge feststellen und evaluieren zu können, bietet sich eine Regressionsanalyse der Daten an.

Auf diese Weise kann eine Bewertungsfunktion für unbekannte Zeitreihen erstellt werden, die Auskunft über die zu erwartende Prognosequalität anhand von verschiedenen Kennzahlen gibt, welche aus der betreffenden Zeitreihe ermittelt werden können.

2.6. Zusammenfassung

Im vorangegangenen Kapitel werden verschiedene in der Literatur gängige sowie in *First-Sale* eingesetzte Prognoseverfahren vorgestellt. Den Forderungen nach einer einfachen und nachvollziehbaren Prognose, die an ein neues Modul seitens der Kunden gestellt werden, kann mit den aufgezählten Verfahren nicht nachgekommen werden. Aus diesem Grund wird der Entschluss gefasst, ein neues Verfahren zu implementieren.

3. Daten-Analysatoren

Energieverbräuche werden von verschiedenen externen Faktoren beeinflusst. Im folgenden Kapitel werden verschiedene Analysatoren vorgestellt, die Charakteristika von Zeitreihen analysieren und den Einfluss der vorgestellten Faktoren sichtbar machen.

3.1. Einflussfaktoren auf den Energieverbrauch

Zur Prognose von Energiezeitreihen ist es zunächst unerlässlich zu wissen, welche Faktoren den Verbrauch beeinflussen. In der Literatur werden immer wieder bestimmte Faktoren angeführt: historische Verbräuche, Temperatur (Wetter), kalendarische Ereignisse, Konjunktur und sozioökonomische Ereignisse (Hahn u. a. 2009; Hippert u. a. 2001; Taylor u. a. 2006).

In den historischen Verbrauchsdaten finden sich zumeist drei Arten von Zyklen: einen Tages-, einen Wochen- und einen Jahreszyklus, die bei der Prognose fortgeschrieben werden. Während der Wochenzyklus zumeist Werktage und Wochenenden unterscheidet, bildet der Jahreszyklus das Fortschreiten der Jahreszeiten ab (Hahn u. a. 2009).

Abhängigkeiten vom Wetter, insbesondere der Temperatur, werden in vielen Prognosemodellen berücksichtigt, wobei die Intensität des Einflusses auf den Stromverbrauch regional unterschiedlich ist (Fidalgo und Matos 2007). Der Zusammenhang wird oft als U-förmige Kurve beschrieben. An kalten Tagen steigt der Stromverbrauch beispielsweise wegen Heiztätigkeit, während an besonders heißen Tagen der Stromverbrauch für Klimaanlage und Kühlungen zu Buche schlägt (Hippert u. a. 2001; Hor u. a. 2005).

Unter kalendarischen Ereignissen, die einen Einfluss auf den Stromverbrauch haben, werden oftmals Feiertage, Brückentage und Ferientage angeführt. Diese haben häufig einen anderen Tageslastgang. Nicht nur sind die Verbräuche oftmals geringer, auch stellen sie teilweise ein Problem bei der Prognose dar, wenn zu wenig repräsentative Tageslastgänge im Datenarchiv existieren (Chen u. a. 2004; Fidalgo und Matos 2007; Taylor u. a. 2006).

Ebenfalls unterschieden wird der Zeitraum, für den eine Prognose erstellt werden soll. Die vorgestellten Faktoren beeinflussen vornehmlich Prognosen für Zeiträume bis zu ei-

nem Jahr. Insbesondere für Kurzfristprognosen von bis zu 6 Stunden ist eine genaue Wetterprognose unerlässlich. Für längerfristige Prognosen von mehreren Jahren treten andere Faktoren, wie Konjunktur und Bevölkerungswachstum, in den Vordergrund Hor u. a. 2005. Das Wetter ist für größere Zeiträume nur spärlich vorhersagbar, so dass dessen Einfluss nur in Form von saisonalen Schwankungen enthalten ist (Feinberg und Genethliou 2005; Fidalgo und Matos 2007).

Für den hier vorgestellten Algorithmus wird vor allem die Abhängigkeit der Zeitreihen von ihren historischen Verbrauchsdaten genutzt. Weiterhin wird der Einfluss von Kalendereignissen wie Feiertagen und Ferien berücksichtigt. Auch die Zyklizität von Zeitreihen wird zur Bewertung der Prognostizierbarkeit von Zeitreihen herangezogen.

3.2. Analysatoren

Aufbauend auf dem Wissen um Faktoren, die den Energieverbrauch beeinflussen können, werden im folgenden Analysatoren vorgestellt, die Einflüsse sichtbar machen und quantifizieren können. Auf diese Weise werden Kennzahlen errechnet, welche anschließend zur Bewertung des Datenarchivs herangezogen werden. Im folgenden sind die verschiedenen Analysatoren aufgeführt und deren Analysemethoden beschrieben.

3.2.1. Streuungs-Analysator

Der erste Analysator, der vorgestellt wird, untersucht die abgenommene Maximalleistung und die verbrauchte Energie eines Archivtages. Er soll eine Aussage über die Ähnlichkeit der Tage innerhalb eines Datenarchivs treffen. Das Prinzip des Streuungs-Analysators soll im Folgenden an zwei Beispielen vorgestellt werden.

Abbildung 3.1 zeigt ein Histogramm von gemessener Tagesmaximalleistung und Tagesenergie einer Schule an Wochenenden über einen Zeitraum von zweieinhalb Jahren. Auf der horizontalen Achse ist die prozentuale Abweichung vom Durchschnitt über alle Werte aufgetragen. Die vertikale Achse zeigt die Anzahl der Tage, an denen die gemessenen Werte um einen bestimmten Prozentsatz vom Durchschnittswert abweichen. Zur besseren Übersicht wurden Quantile eingezeichnet und farblich hervorgehoben.

Erkennbar ist eine starke Konzentration um den Durchschnittswert. Der überwiegende Teil aller Werte weicht nur gering vom Durchschnitt ab. Folglich kann der Stromverbrauch einer Schule am Wochenende als ähnlich angesehen werden. Somit stellt eine Unterteilung der Archivtage nach Werktagen und Wochenenden eine vielversprechende Einteilung dar.

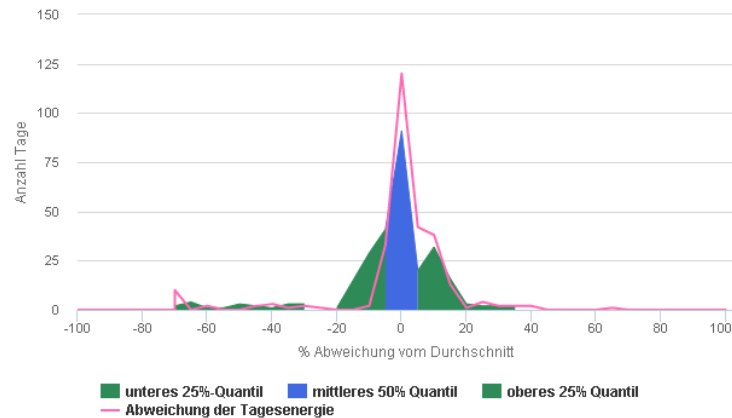


Abbildung 3.1.: Histogramm von Tagesenergie und Maximalleistung einer Schule an Wochenenden

Dem steht jedoch die Auswertung über alle Werktage in Abbildung 3.2 entgegen. An zwei Stellen konzentrieren sich Werte auf der horizontalen Achse, die jeweils weit von dem Mittelpunkt entfernt sind. Bei dem betrachteten Objekt handelt es sich ebenfalls um eine Schule. Die Werktage fallen zu einem nicht unerheblichen Teil in die Ferienzeit, was offensichtlich einen stark verringerten Stromverbrauch an manchen Tagen zur Folge hat.

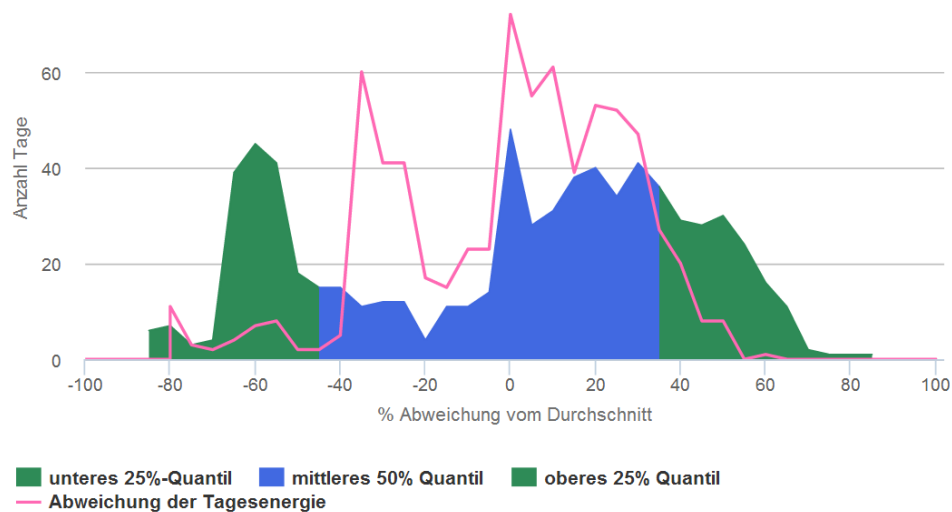


Abbildung 3.2.: Histogramm von Tagesenergie und Maximalleistung einer Schule an Werktagen

Aus dem oben angeführten Beispiel ergibt die folgende Annahme: Eine Menge von Tagen kann als ähnlich bezeichnet werden, wenn Energieverbrauch und Maximalleistung nur gering vom Durchschnittstag abweichen.

Gesucht wird eine Kennzahl, die diese Beobachtung quantifiziert und eine Aussage darüber trifft, wie stark die Werte vom Mittelwert abweichen. Die deskriptive Statistik bietet zu diesem Zweck verschiedene Streuungsmaße an. Die prominentesten Vertreter, auf die an dieser Stelle eingegangen wird, sind Varianz und Streuung.

Die Varianz wird berechnet als Wurzel der Summe der quadratischen Abweichungen vom Mittelwert über alle ausgewählten Archivtage. Durch das Quadrieren der Abweichungen wird einer großen Abweichung stärkeres Gewicht beigemessen als mehreren, betragsmäßig identischen geringen Abweichungen vom Mittel. Das Bestrafen von großen, falsch prognostizierten Mengen findet keine Entsprechung am Strommarkt. Nicht zu erwarten ist, dass die leicht erhöhte Nachfrage eines Marktteilnehmers den Strompreis im selben Maße beeinflusst.

Die mittlere absolute Abweichung hingegen gewichtet die Ausreißer nicht stärker und wird hier zur Bewertung eines Datenarchivs herangezogen. Der im Rahmen der Arbeit eingesetzte Streuungs-Analysator berechnet mit Formel 3.1 für jedes der 96 Zeitintervalle t die absolute Abweichung δ_t vom Durchschnitt \bar{x}_t aller Archivwerte x_{ti} . Damit trifft er eine Aussage darüber, wie gut die Werte x_t durch den Mittelwert über alle Archivtage n prognostizierbar sind.

$$\delta_t = \frac{1}{n} \sum_{i=1}^n |x_{ti} - \bar{x}_t| \quad (3.1)$$

Abbildung 3.3 zeigt das Ergebnis einer solchen Analyse für eine Schule an Werktagen. Daran kann man ablesen, dass die Durchschnittszeitreihe für Werte zwischen 21:00 und 5:00 Uhr eine recht gute Prognose des Verbrauchs darstellen wird. Die schlechtesten Ergebnisse sind zur Mittagszeit zu erwarten.

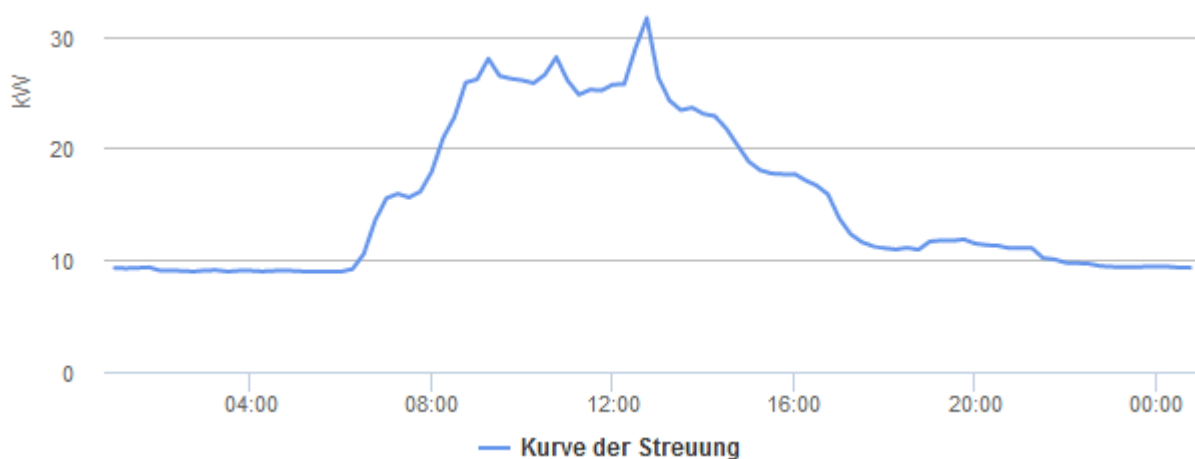


Abbildung 3.3.: Kurve der Streuung für eine Schule

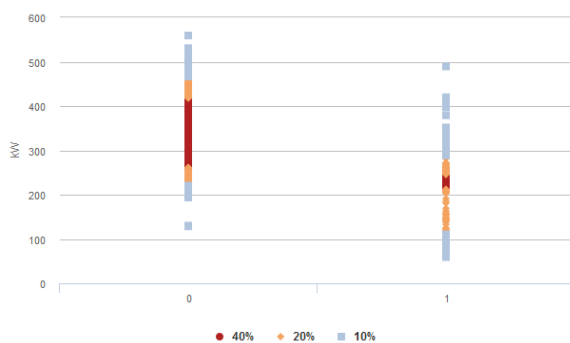
3.2.2. Korrelations-Analysator

In Kapitel 3.1 wurden bereits mehrere Faktoren erläutert, die den Stromverbrauch beeinflussen. Es stellt sich nun die Frage, ob und in welchem Maß eine Zeitreihe von diesen abhängig ist. Manche der Einflussfaktoren liegen selbst als Zeitreihen vor, andere wiederum müssen in solche umgewandelt werden. Im Folgenden werden Methoden vorgestellt, um Zusammenhänge mit den jeweiligen Faktoren sichtbar zu machen.

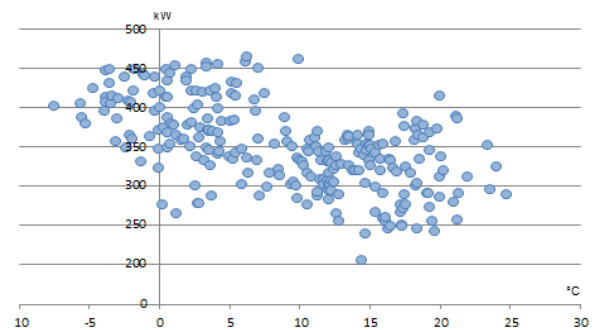
Um Abhängigkeiten des Stromverbrauches von zeitlich wiederkehrenden Ereignissen, die mittels Kalender festgehalten sind, zu analysieren, muss aus einem Kalender K für einen festgelegten Zeitraum eine Zeitreihe erzeugt werden. Zur Untersuchung einer Abhängigkeit von Schulferien kann eine solche Zeitreihe wie folgt erzeugt werden: Für jeden Tag d wird abhängig von der Tagesart ein Wert erzeugt.

$$f(d, K) = \begin{cases} 1, & \text{wenn } d \text{ Ferientag in } K, \\ 0, & \text{wenn } d \text{ kein Ferientag in } K. \end{cases} \quad (3.2)$$

In Abbildung 3.4a ist die Tagesenergie einer Schule über einem Zeitraum von zwei Jahren anhand einer generierten Ferienzeitreihe als Streudiagramm aufgetragen. Zur besseren Anschaulichkeit sind Werte nach ihrem Anteil an Gesamtwerten farblich markiert: 40% der Nicht-Ferientage haben einen Energieverbrauch zwischen 280 und 420 kW. Wie zu erwarten ist, liegen die Verbräuche an Ferientagen deutlich unter denen der Werkstage. Die Werte der generierten Ferienzeitreihe und der Verbrauchszeitreihe sind also gegenläufig, wie bereits in Grafik 3.2 zu erkennen.



(a) Schule mit Schulferien



(b) Tankstelle mit Temperatur

Abbildung 3.4.: Streudiagramm der Tagesenergie zweier Zeitreihen mit verschiedenen Einflussfaktoren

Die meisten Einflussfaktoren auf den Energieverbrauch sind natürlichen Ursprungs und nehmen damit mehr als zwei Werte an. Ist dies der Fall, ergibt ein Streudiagramm eine

Punktwolke, wie sie in Abbildung 3.4b dargestellt ist. Sie zeigt den Energieverbrauch durch eine Tankstelle in Abhängigkeit von der Tagesdurchschnittstemperatur. Auch hier ist ein Zusammenhang der beiden Größen sichtbar: Höhere Außentemperaturen führen zu einem geringeren Energieverbrauch.

In beiden Beispielen ähnelt die Gestalt der Punktwolke einer fallenden Gerade, auch Regressionsgerade genannt, und lässt einen linearen Zusammenhang der jeweiligen Größen vermuten. Ein derartiger Zusammenhang lässt sich mittels Korrelationskoeffizient von Bravais Pearson quantifizieren:

$$\text{Kor}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.3)$$

Er wird berechnet aus der Summe über das Produkt aller Abweichungen der beiden Größen x_i , y_i von ihren jeweiligen Mittelwerten \bar{x} und \bar{y} . Durch die Division durch die jeweilige Standardabweichung wird der Wert auf das Intervall $[-1 \dots 1]$ normiert. Damit wird der Effekt, den die Skalierung der Ausgangsgrößen auf den Quotienten hat, eliminiert.

Ein Korrelationskoeffizient von 1 beschreibt eine exakte Gleichläufigkeit der untersuchten Größen, während -1 Gegenläufigkeit ausdrückt. Dies bedeutet jedoch nicht, dass die Größen stochastisch unabhängig sind. Ein quadratischer oder anders gearteter Zusammenhang ist durchaus denkbar.

Der Korrelationskoeffizient beschreibt also, wie nahe die Werte der Regressionsgeraden folgen. Er trifft keine Aussage über die Steigung der Gerade, also darüber, wie stark der Einfluss auf die abhängige Variable ist. Dennoch wird die Annahme getroffen, dass ein größerer Betrag des Korrelationskoeffizienten eine positive Wirkung auf die Vorhersagbarkeit einer Zeitreihe durch den Einflussfaktor hat. Eine lineare Abhängigkeit wird an dieser Stelle vorausgesetzt. Die Prognosegüte einer Zeitreihe müsste in diesem Zusammenhang eine Anhängigkeit von dem Korrelationskoeffizienten der Zeitreihe mit ihrer Einflussgröße aufweisen.

3.2.3. Periodizitäts-Analysator

Wie in Abschnitt 3.1 beschrieben, sind bei Stromzeitreihen mehrere Zyklen beobachtbar: tägliche, wöchentliche und jährliche. Vergangenes Verbrauchsverhalten beeinflusst das Zukünftige.

Im vorangegangenen Kapitel wurde bereits eine Methode vorgestellt, um Zusammenhänge mit anderen Zeitreihen sichtbar zu machen. Diese wird nun auch eingesetzt, um Zyklen innerhalb einer Zeitreihe aufzuzeigen. Dazu wird der Korrelationskoeffizient der Zeitreihe

mit sich selbst berechnet, wobei diese um einen bestimmten Zeitraum versetzt ist. Durch die Verschiebung der Werte x_i um den Zeitraum t erhält man gemäß Formel (8.1):

$$\text{Kor}(x, x_t) = \frac{\sum_{i=1}^{n-t} (x_i - \bar{x})(x_{i+t} - \bar{x})}{\sqrt{\sum_{i=1}^{n-t} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n-t} (x_{i+t} - \bar{x})^2}} \quad (3.4)$$

Das Ergebnis einer Berechnung ist ein Autokorrelationskoeffizient mit einem Wertebereich von $[-1, \dots, 1]$. Abbildung 3.5 zeigt eine Grafik der errechneten Werte bei sich vergrößerndem Zeitversatz für die Zeitreihe einer Schule. Die Kurve besitzt lokale Maxima bei 24 Stunden Zeitversatz und dessen vielfache. Es ist auffällig, dass jedes siebte Maximum höher ausfällt. Dies ist ein Hinweis darauf, dass der Lastgang sich im Tages- und im Wochenrhythmus zyklisch wiederholt.

Zieht man diese Eigenschaft zur Prognose heran, so würde eine Zeitreihe durch vergangene ausgewählte Tage oder Wochen prognostiziert. Der Autokorrelationskoeffizient könnte somit eine Aussage über die Prognosequalität bei einem solchen Verfahren treffen.

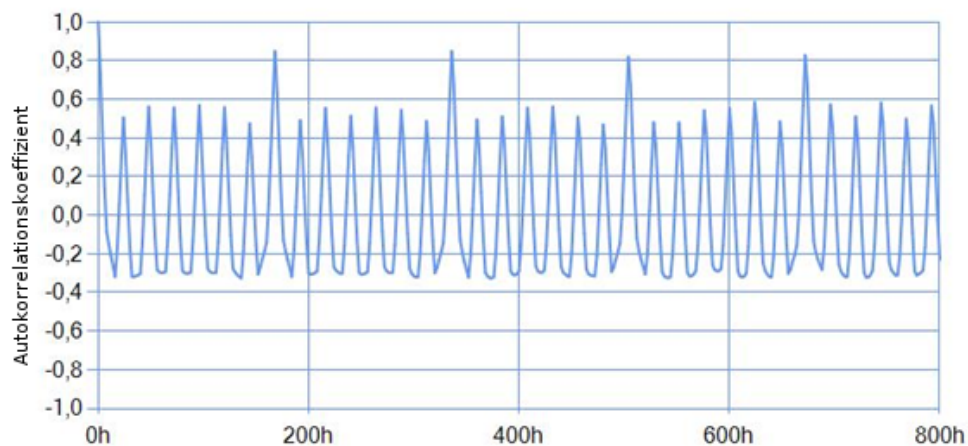


Abbildung 3.5.: Autokorrelationsdiagramm einer Schule

3.3. Zusammenfassung

In diesem Kapitel wurden verschiedene Zeitreihen beeinflussende Größen vorgestellt und Methoden aufgezeigt, um diese sichtbar und damit quantifizierbar zu machen. Dazu wurden verschiedene Analysatoren eingeführt:

Der Streuungs-Analysator gibt Auskunft über die Ähnlichkeit von Archivtagen und kann somit zur Bestimmung eines geeigneten Datenarchivs genutzt werden. Mittels des Korrelations-Analysators kann eine Aussage darüber getroffen werden, wie gut eine Zeit-

reihe durch eine Einflussgröße prognostizierbar ist. Der Periodizitäts-Analysator erkennt Zyklen innerhalb von Zeitreihen. Erwartet wird, dass sich diese Eigenschaft ebenfalls auf die Prognosequalität auswirkt.

4. Modellierung von Optimierungsproblem und Algorithmus

Das folgende Kapitel ist zweigeteilt. Zunächst wird die Zeitreihenprognose als Optimierungsproblem mathematisch definiert. Anschließend wird der Algorithmus beschrieben, mit dem das Optimierungsproblem gelöst und die Prognose durchgeführt wird.

4.1. Optimierungsproblem

Im Folgenden wird das Problem der optimalen Zeitreihenprognose als lineares Optimierungsproblem definiert und eine Prognosefunktion vorgestellt, die dieses Problem minimiert. Die Prognosefunktion bedient sich dabei der in Kapitel 3 vorgestellten spezifischen Eigenschaften von Energiezeitreihen.

4.1.1. Zielfunktion

Zur Optimierung der Zeitreihenprognose steht für jede Zeitreihe ein umfangreiches Datenarchiv zur Verfügung. Ziel ist es, ausgehend von einem Datenarchiv, eine Prognosefunktion zu ermitteln.

Die Güte dieses Ermittlungsalgorithmus kann gemessen werden, indem er auf eine Situation angewendet wird, in der die Ist-Daten am zu prognostizierenden Tag bereits bekannt sind und dann die Prognose auf Basis von Archivtagen durchgeführt wird, die in der Vergangenheit liegen.

Metrischer Raum der Datumswerte

Das Datenarchiv ist gemäß der starken Abhängigkeit des Verbrauchs vom Kalender jeweils in Tageslastgänge unterteilt.

Es sei also \mathbb{D} die Menge der Datumswerte, auf der mittels (4.2) eine Ordnungsrelation definiert wird.

$$\mathbb{D} := \left\{ \begin{pmatrix} d \\ m \\ y \end{pmatrix} \in \mathbb{N}^3 \mid \text{Im Jahr } y \text{ hat der Monat } m \text{ mindestens } d \text{ Tage} \right\} \quad (4.1)$$

$$\begin{aligned} \begin{pmatrix} d_1 \\ m_1 \\ y_1 \end{pmatrix} < \begin{pmatrix} d_2 \\ m_2 \\ y_2 \end{pmatrix} &: \Leftrightarrow (d_1 < d_2 \wedge m_1 = m_2 \wedge y_1 = y_2) \\ &\vee (m_1 < m_2 \wedge y_1 = y_2) \\ &\vee (y_1 < y_2) \end{aligned} \quad (4.2)$$

Auf \mathbb{D} wird dann wie folgt eine Metrik eingeführt:

$$\rho \left(\begin{pmatrix} d_1 \\ m_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} d_2 \\ m_2 \\ y_2 \end{pmatrix} \right) := \text{Anzahl Tage zwischen den beiden Tagen} \quad (4.3)$$

Nach (4.3) ist der Abstand zwischen $(2, 3, 2015)^T$ und $(27, 2, 2015)^T$ definiert als $\rho = 3$. Im Sinne dieser Metrik wird mit

$$\overline{B}_r(\tilde{d}) := \{d \in \mathbb{D} \mid \rho(d, \tilde{d}) \leq r\} \quad \tilde{d} \in \mathbb{D}, r \geq 0 \quad (4.4)$$

die kompakte Kugel mit dem Radius r bezeichnet. Für eine Teilmenge $J \subset \mathbb{D}$, $J \neq \emptyset$ wird zudem geschrieben:

$$\overline{B}_r(J) := \bigcup_{\tilde{d} \in J} \overline{B}_r(\tilde{d}) \quad (4.5)$$

Eine kompakte Kugel mit Radius 2 um $(5, 1, 2015)^T$ wäre demnach:

$$\overline{B}_2 \left(\begin{pmatrix} 5 \\ 1 \\ 2015 \end{pmatrix} \right) = \left\{ \begin{pmatrix} 3 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 5 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 6 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 7 \\ 1 \\ 2015 \end{pmatrix} \right\} \quad (4.6)$$

Datenarchiv

Das Datenarchiv \mathcal{A} wird definiert als eine Funktion, die eine Menge von Tagen D auf einen Tageslastgang $x \in \mathbb{R}^{96}$ mit den Komponenten x_1, \dots, x_{96} abbildet. D sei dabei die Menge der Tage, an denen Tageslastgänge bekannt sind:

$$\mathcal{A} : D \subset \mathbb{D} \rightarrow \mathbb{R}^{96} \quad D \neq \emptyset \quad (4.7)$$

Da es sich um eine Prognose in die Zukunft handelt, wird für jeden Tag $\mathring{d} > \min(D)$ ein Archiv $\mathcal{A}(\mathring{d})$ gebildet, welches diejenigen Tage, die zeitlich vor dem zu prognostizierenden Tag liegen, auf \mathbb{R}^{96} abbildet.

$$\mathcal{A}(\mathring{d}) = \mathcal{A}|_{\{d \in D \mid d < \mathring{d}\}} \quad \forall \mathring{d} \in \mathbb{D}, \mathring{d} > \min D \quad (4.8)$$

Man nehme beispielsweise an, zu allen Tagen in D mit

$$D = \left\{ \begin{pmatrix} 1 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 5 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 6 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 7 \\ 1 \\ 2015 \end{pmatrix} \right\} \quad (4.9)$$

seien Zeitreihen verfügbar. Dann ergibt sich ein Datenarchiv \mathcal{A} von $\mathring{d} = (3, 1, 2015)^T$ wie folgt, welches jeden Tag d auf einen Lastgang $z \in \mathbb{R}^{96}$ abbildet.

$$\mathcal{A} \left(\begin{pmatrix} 5 \\ 1 \\ 2015 \end{pmatrix} \right) : \left\{ \begin{pmatrix} 1 \\ 1 \\ 2015 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2015 \end{pmatrix} \right\} \rightarrow \mathbb{R}^{96} \quad (4.10)$$

$$\begin{pmatrix} 1 \\ 1 \\ 2015 \end{pmatrix} \mapsto \begin{pmatrix} 10 \\ 15 \\ \cdots \\ 10 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2015 \end{pmatrix} \mapsto \begin{pmatrix} 3 \\ 10 \\ \cdots \\ 10 \end{pmatrix} \quad (4.11)$$

Prognosefunktion

Die Prognosefunktion P berechnet aus dem Prognosetag \mathring{d} und dem Datenarchiv $\mathcal{A}(\mathring{d})$ einen Tageslastgang $P(\mathring{d}, \mathcal{A}(\mathring{d}))$, der die Prognose des Tageslastgangs für den Tag \mathring{d} darstellt.

$$P(\mathring{d}, \mathcal{A}(\mathring{d})) \in \mathbb{R}^{96} \quad (4.12)$$

Es gilt, eine möglichst einfache Funktion P unter allen Prognosefunktionen \mathfrak{P} zu finden, so dass die prognostizierten Werte möglichst nahe bei den Ist-Werten liegen. Wenn P die algorithmisch ermittelte Prognosefunktion für den Tag \mathring{d} ist und $\mathring{x} \in \mathbb{R}^{96} = (\mathring{x}_1, \dots, \mathring{x}_{96})^T$ die Ist-Werte des zu prognostizierenden Tages beschreibt, dann ist

$$\sum_{t=1}^{96} \frac{|\mathring{x}_t - [P(\mathring{d}, \mathcal{A}(\mathring{d}))]_t|}{\mathring{x}_t} \quad (4.13)$$

als ein Maß für die Güte der Prognosefunktion-Auswahl. Die zusätzliche Division der Differenz durch den Ist-Wert \mathring{x}_t berücksichtigt die unterschiedlichen Verbrauchsniveaus der abgenommenen Energie, da die prozentuale Abweichung von den Ist-Werten betrachtet wird.

Da Formel 4.13 als Maß der Prognosegüte die Prognosefehler aufsummiert, ist eine Prognose dann optimal, wenn ein $P \in \mathfrak{P}$ gefunden wird, so dass der Prognosefehler minimal ist. Es handelt sich hier also um ein Minimierungsproblem.

4.1.2. Definition der Prognosefunktion

Wie im vorangegangenen Kapitel erörtert, wird zur Optimierung eine möglichst einfache Prognosefunktion P aus allen Prognosefunktionen \mathfrak{P} gesucht, die den Prognosefehler nach Formel (4.13) minimiert. Zur Ermittlung der Prognosefunktion P wird das Datenarchiv $\mathcal{A}(\mathring{d})$ auf eine Menge repräsentativer Archivtage reduziert. Die Prognose selbst ergibt sich aus der Durchschnittszeitreihe jener Archivtage.

Tagesgruppierungen

Aus der Literatur konnte bereits entnommen werden, dass die meisten Zeitreihen Tages- und Wochenzyklen enthalten (siehe Kapitel 3.1). Ebenfalls wurde gezeigt, dass Archivtage anhand von Tageseigenschaften gruppierbar sind. Daraufhin wurde die Vermutung aufgestellt, dass die Ähnlichkeit von Archivtagen mit einem geringen Streuungswert innerhalb einer Tagesgruppe einhergeht (siehe Kapitel 3.2.1).

Eine Tagesgruppierung entsteht durch eine Äquivalenzrelation aus \mathbb{D} , mit Hilfe derer die Tage als gleichartig (z.B. gleicher Wochentag) angesehen werden. Die Tagesgruppen sind hierbei die durch diese Äquivalenzrelation induzierten Äquivalenzklassen.

Bsp.: Auf \mathbb{D} wird durch

$$\begin{aligned} d_1 \sim d_2 : \Leftrightarrow & \quad d_1, d_2 \text{ sind beide Samstage} \\ & \text{oder beide Sonntage} \\ & \text{oder beide sind weder Samstag noch Sonntag} \end{aligned} \quad (4.14)$$

eine Äquivalenzrelation definiert. Sie erzeugt drei Äquivalenzklassen: die Samstage, die Sonntage und alle anderen Tage.

Es gilt, die am besten den Prognosetag \mathring{d} geeignete Tagesgruppe G_\sim^* zu finden. Hierfür wird eine Vorauswahl von Äquivalenzrelationen $G_\sim \in \mathbb{G}$ getroffen. Für eine Tagesgruppierung $\sim \in \mathbb{G}$ und einen zu prognostizierenden Tag \mathring{d} sei $[\mathring{d}]_\sim \subset \mathbb{D}$ die Äquivalenzklasse, in der \mathring{d} liegt. Dann ist

$$G(\mathring{d})_\sim := [\mathring{d}]_\sim \cap \{d \in D \mid d < \mathring{d}\} \quad (4.15)$$

die Menge der zur Prognose von \mathring{d} heranzuziehenden Archivtage, die äquivalent zu \mathring{d} bezüglich \sim sind. Das Ziel ist es, die Tagesgruppierung $G_\sim^* \in \mathbb{G}$ zu ermitteln, die analog zu Formel 3.1

$$\frac{1}{|\mathcal{A}_{\mathring{d}}(G(\mathring{d})_\sim)|} \cdot \sum_{x \in \mathcal{A}_{\mathring{d}}(G(\mathring{d})_\sim)} \|x - \bar{x}\| \quad (4.16)$$

mit

$$\bar{x} := \frac{1}{|\mathcal{A}_{\mathring{d}}(G(\mathring{d})_\sim)|} \cdot \sum_{x \in \mathcal{A}_{\mathring{d}}(G(\mathring{d})_\sim)} x \quad (4.17)$$

minimiert. Dabei wird die 1-Norm auf dem \mathbb{R}^{96} betrachtet, also

$$\|x\| := \|x\|_1 := \sum_{t=1}^{96} |x_t| \quad (4.18)$$

Zeitfenster

In Kapitel 3.1 wurde bereits über die Saisonalität von Zeitreihen gesprochen. Der Energieverbrauch verändert sich langsam im Jahresverlauf. Diese Veränderungen wiederholen sich im Jahresrhythmus. Demnach kann sich der Lastgang an einem Montag im Januar auch bei starkem wochenzyklischen Verhalten von einem Montag im August signifikant unterscheiden.

Abbildung 4.1 zeigt Wochenlastgänge eines Schwimmbades zu unterschiedlichen Jahreszeiten im Jahr 2013. Es ist ersichtlich, dass die verschiedenen Verbrauchskurven eine ähnliche Form haben, sich aber im Leistungsniveau unterscheiden.

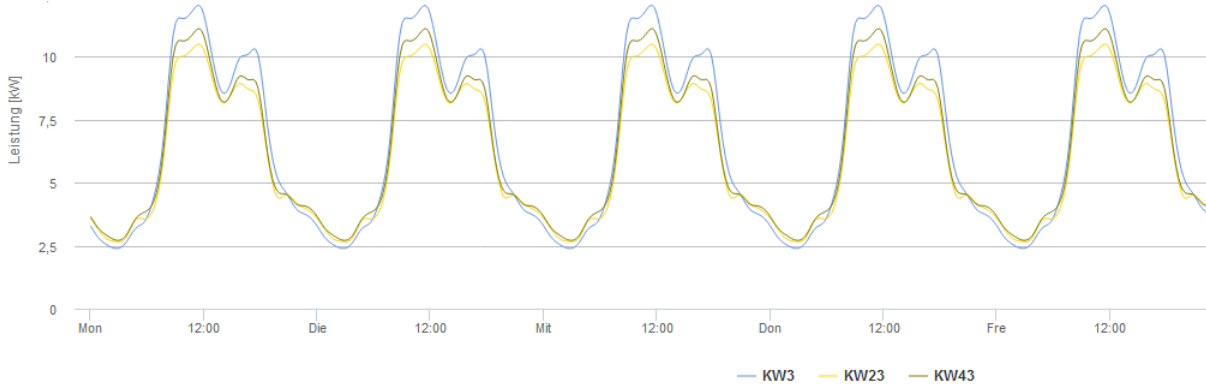


Abbildung 4.1.: Energieverbrauch eines Schwimmbades an Wochentagen zu unterschiedlichen Kalenderwochen im Jahr 2013

Um die Saisonalität einer Zeitreihe zu berücksichtigen, erscheint es sinnvoll, dass nur solche Tage innerhalb einer Prognosegruppe in die Prognose einbezogen werden, die eine zeitliche Nähe zum Prognosetag aufweisen. An dieser Stelle wird die Saisonalität durch ein gleitendes Zeitfenster abgebildet. Somit werden nur Archivtage verwendet, die (abgesehen von der Jahresangabe) sich in Tag und Monat um höchstens eine fest angegebene Zahl an Tagen unterscheidet.

Konkret wenn $\mathring{d} = (d_0, m_0, y_0)^T$ der zu prognostizierende Tag ist und

$$J(\mathring{d}) = \begin{pmatrix} d \\ m \\ y \end{pmatrix} \quad (4.19)$$

die Menge aller Tage, die in Tag und Monat mit \mathring{d} übereinstimmen, dann werden als Archivtage bei einem Zeitfenster von $z \geq 0$ Tage nur

$$[\mathring{d}]_{\sim} \cap \{d \in D \mid d < \mathring{d}\} \cap \overline{B}_z(J(\mathring{d})) \quad (4.20)$$

als Archivtage akzeptiert.

Es sei beispielsweise ein Prognosetag $\mathring{d} = (5, 1, 2015)^T$, ein Datenarchiv $D = \{(1, 1, 2013)^T, \dots, (31, 12, 2015)^T\}$ und eine Äquivalenzklasse $[\mathring{d}]_{\sim} = \{d \in D \mid d \text{ ist ein Montag}\}$, dann ist die Menge der akzeptierten Archivtage bei einem Zeitfenster $z = 10$ nach Formel (4.20):

$$\left\{ \begin{pmatrix} 31 \\ 12 \\ 2014 \end{pmatrix}, \begin{pmatrix} 13 \\ 1 \\ 2014 \end{pmatrix}, \begin{pmatrix} 6 \\ 1 \\ 2014 \end{pmatrix}, \begin{pmatrix} 30 \\ 12 \\ 2013 \end{pmatrix}, \begin{pmatrix} 14 \\ 1 \\ 2013 \end{pmatrix}, \begin{pmatrix} 7 \\ 1 \\ 2013 \end{pmatrix} \right\} \quad (4.21)$$

Ziel ist es, das optimale Zeitfenster z^* zu finden, aus dem die Archivtage gewählt werden sollen. Die Länge des Fensters sollte sich danach bemessen, wie stark sich die Lastgänge innerhalb des Fensters verändern, oder in anderen Worten, wie stark sie einander ähneln.

Bisher wurde die Streuung als Maß für die Ähnlichkeit von Archivtagen genutzt, diese wird hier ebenfalls herangezogen. Unter der Annahme, dass ein gutes Zeitfenster eine geringe Streuung aufweist, gilt es bei optimaler Tagesgruppe G_\sim^* ein z^* zu finden, das

$$\frac{1}{|\mathcal{A}_{\dot{d}}(G_\sim^*(\dot{d}) \cap \overline{B}_z(J(\dot{d})))|} \cdot \sum_{x \in \mathcal{A}_{\dot{d}}(G_\sim^*(\dot{d}) \cap \overline{B}_z(J(\dot{d})))} \|x - \bar{x}\| \quad (4.22)$$

mit

$$\bar{x} := \frac{1}{|\mathcal{A}_{\dot{d}}(G_\sim^*(\dot{d}) \cap \overline{B}_z(J(\dot{d})))|} \cdot \sum_{x \in \mathcal{A}_{\dot{d}}(G_\sim^*(\dot{d}) \cap \overline{B}_z(J(\dot{d})))} x \quad (4.23)$$

minimiert.

Prognosefunktion

Der Durchschnittslastgang aller repräsentativen Archivtage ergibt die Zeitreihe des Prognosetages \dot{d} . Zu diesem Zwecke wird das Datenarchiv $\mathcal{A}(\dot{d})$ anhand einer Tagesgruppierung G und der Zeitfensterbreite z eingeschränkt. Die Prognose eines Lastgangs $\hat{x} = (\hat{x}_1, \dots, \hat{x}_{96})^T$ wird folglich berechnet aus:

$$\hat{x}_t = P_t(\dot{d}, \mathcal{A}(\dot{d}), G, z) := \frac{1}{|\mathcal{A}_{\dot{d}}(G_\sim(\dot{d}) \cap \overline{B}_z(J(\dot{d})))|} \cdot \sum_{x \in \mathcal{A}_{\dot{d}}(G_\sim(\dot{d}) \cap \overline{B}_z(J(\dot{d})))} x_t \quad (4.24)$$

4.2. Algorithmus

Dieses Kapitel stellt einen vierstufigen Algorithmus vor, der das in Kapitel 4.1 vorgestellte Optimierungsproblem implementiert und löst. Ziel ist es, eine Menge repräsentativer Tage im Datenarchiv zu identifizieren, aus denen sich ein möglichst guter Lastgang für den jeweiligen Prognosetage errechnen lässt.

Zunächst wird ein Überblick über den Aufbau des Algorithmus gegeben. Die folgenden Kapitel stellen die Lösung der einzelnen Optimierungsprobleme vor. Dabei wird jeder Algorithmusteil in Form von Pseudocode aufgeschrieben und erläutert.

4.2.1. Struktur des Algorithmus

Zu Beginn des Algorithmus muss der Prognosezeitraum T festgelegt werden (siehe Algorithmus 1). Der Prognosezeitraum umfasst alle Tage \hat{d} , für die ein Lastgang prognostiziert werden soll. Anschließend werden auf Basis des Datenarchivs stufenweise die einzelnen Parameter der Prognosefunktion ermittelt.

In Stufe eins des Algorithmus wird zunächst für jede Tagesgruppierung die Streuung innerhalb der zugehörigen Archivtage berechnet. Die Liste wird aufsteigend mittels Quicksort nach dem Streuungswert sortiert und pro Prognosetag \hat{d} aus T diejenige Tagesgruppierung G^* ermittelt, die am besten geeignet erscheint (siehe Kapitel 4.2.2). Das Ergebnis der Berechnung wird als Tupel (\hat{d}, G^*) in einer Liste L gespeichert und in Stufe zwei weiterverarbeitet.

Die nächste Stufe des Algorithmus bildet die Saisonalität von Zeitreihen mittels eines Zeitfensters ab. Es wird das Datenarchiv auf diejenigen Tage eingeschränkt, die innerhalb eines Abstandes z vom Prognosetag \hat{d} entfernt sind (siehe Kapitel 4.2.3). Das ermittelte Zeitfenster z^* wird zusammen mit Prognosetag \hat{d} und Tagesgruppe G^* gesichert.

Nach der Ermittlung eines Zeitfenster stellt sich die Frage, ob eine andere Tagesgruppierung innerhalb des Fensters bessere Ergebnisse verspricht. Dies sollte insbesondere dann der Fall sein, wenn die saisonalen Schwankungen sehr groß sind, oder sich das Verbrauchsverhalten abrupt ändert. Aus diesem Grund wird in Stufe drei die Tagesgruppierung innerhalb des Zeitfensters neu ermittelt (siehe Kapitel 4.2.4).

Anschließend erfolgt eine erneute Berechnung des Zeitfensters (Kapitel 4.2.4) und die Berechnung der Prognosezeitreihe für den Prognosezeitraum (Kapitel 4.2.3).

4.2.2. Ermittlung der Tagesgruppierung

Zur Ermittlung der am besten geeigneten Tagesgruppierung nach Formel (4.16) wird wie in Algorithmus 2 dargestellt vorgegangen. Das Sortieren der Liste *StreuungProGruppierung* vereinfacht nun die Suche nach der besten Tagesgruppe. Die Liste wird sequentiell durchlaufen bis eine Tagesgruppierung gefunden ist, die äquivalent zum Prognosetag \hat{d} ist. Diese wird als beste Gruppierung zurückgegeben.

Algorithmus 1 : Vierstufiger Algorithmus zu Ermittlung optimaler Prognoseparameter

```

1  lege Prognosezeitraum  $T \subset \mathbb{D}$  fest
   // Stufe 1: Optimale Tagesgruppe
   // Berechne die Streuung für alle Tagesgruppen
2  foreach  $G \in \mathbb{G}$  do
3    berechne Streuung  $s \leftarrow s(\mathcal{A}_{\mathring{d}}(G))$  nach Formel (3.1)
4    StreuungProGruppierung.add( $G, s$ )
5  end
   // Ermittlung der Tagesgruppe
6  StreuungProGruppierung.sort( $s$ )
7  foreach  $\mathring{d} \in T$  do
8     $G^* \leftarrow \text{ermittleTagesgruppierung}(\mathring{d}, \text{StreuungProGruppierung})$ 
9     $L.\text{add}(\mathring{d}, G^*)$ 
10 end
    // Stufe 2: Optimales Zeitfenster
11 foreach  $(\mathring{d}, G^*) \in L$  do
12    $z^* \leftarrow \text{ermittleZeitfenster}(\mathring{d}, G^*)$  nach Formel (4.22)
13    $K.\text{add}(\mathring{d}, G^*, z^*)$ 
14 end
    // Stufe 3: Überarbeitung der Tagesgruppe
15 foreach  $(\mathring{d}, G^*, z^*) \in K$  do
16    $G^{**} \leftarrow \text{ueberpruefeTagesgruppe}(\mathring{d}, z^*)$  nach Formel (4.16)
17    $L'.\text{add}(\mathring{d}, G^{**})$ 
18 end
    // Stufe 4: Überarbeitung Zeitfenster
19 foreach  $(\mathring{d}, G^{**}) \in L'$  do
20    $z^{**} \leftarrow \text{ermittleZeitfenster}(\mathring{d}, G^{**})$  nach Formel (4.22)
21    $K'.\text{add}(\mathring{d}, G^{**}, z^{**})$ 
22 end
    // Prognose ausführen
23 foreach  $(\mathring{d}, G^{**}, z^{**}) \in K'$  do
24    $x \leftarrow \text{prognostiziere}(\mathring{d}, G^{**}, z^{**})$ 
25   Prognose.add( $x$ )
26 end

```

Algorithmus 2 : Erste Stufe des Algorithmus, Berechnung der optimalen Tagesgruppierung

Eingabe : \mathring{d} , der Prognosetag

Eingabe : *StreuungProGruppierung*, Auflistung der Tagesgruppierungen mit ihrem Streuungswert

Ausgabe : G^* : *besteGruppe*, enthält die beste Prognosekonfiguration für \mathring{d}

```

1 function ermittleTagesgruppierung( $\mathring{d}$ , StreuungProGruppierung) begin
2    $i = 0$ 
3   do
4      $(G_{\sim}, s) \leftarrow \text{StreuungProGruppierung}[i]$ 
5      $\text{besteGruppe} \leftarrow G_{\sim}$ 
6      $i++$ 
7   until ( $\mathring{d} \sim [G]_{\sim}$ )
8   return besteGruppe
9 end

```

4.2.3. Festlegung des Fensters

Die Implementierung zu Berechnung der optimalen Zeitfenstergröße wird in Algorithmus 3 beschrieben. Das Zeitfenster schränkt eine Auswahl von Archivtagen auf einen Zeitraum um den Prognosetag in jedem Archivjahr ein (siehe Kapitel 4.1.2). Ein Zeitfenster der Länge z umfasst dabei die Tage vor und nach dem Prognosetag, die höchstens z Tage entfernt sind.

Dazu werden in einer Schleife für jede Zeitfensterbreite von 1 bis zu einem Maximalwert z_{max} die zugehörigen Archivtage bestimmt und, sofern Archivtage vorhanden sind, der Streuungswert berechnet. Die Daten werden in einer Liste abgespeichert.

Um auftretende statistische Ausreißer zu eliminieren werden die Streuungswerte zunächst mit dem Verfahren des gleitenden Mittelwertes geglättet. Anschließend wird die Fensterbreite z ermittelt, deren zugehöriger Streuungswert s' am geringsten ist. Bei der Ermittlung werden alle Fenster verworfen, die eine höhere Streuung als aufweisen, als die Tagesgruppierung ohne Zeitfenster. Deren Streuungswert s_G^* wurde bereits im vorherigen Algorithmus Schritt ermittelt und ist verfügbar.

Wurde ein passendes Zeitfenster gefunden, wird als Zeitfensterbreite z^* der Wert in *bestesZeitfenster* zurückgegeben. Andernfalls wird die Zeitfensterbreite auf 183 gesetzt, was einem halben Jahr entspricht und daher keine Einschränkung der Archivtage darstellt.

Algorithmus 3 : Zweite Stufe des Algorithmus, Berechnung der optimalen Zeitfensters

Eingabe : \mathring{d} , der Prognosetag
Eingabe : G^* , enthält die ermittelte beste Tagesgruppierung für \mathring{d}
Eingabe : s_{G^*} , enthält den Streuungswert für G^*
Ausgabe : z^* : *bestesFenster*, enthält beste Zeitfensterbreite für \mathring{d}

```

1 function ermittleZeitfenster( $\mathring{d}, G^*, s_{G^*}$ ) begin
2   foreach  $z \in [1 \dots z_{max}]$  do
3      $Archivtage \leftarrow G^* \cap \overline{B}_z(J(\mathring{d}))$ 
4     if  $Archivtage \neq \emptyset$  then
5        $s \leftarrow$  berechne Streuung über  $\mathcal{A}(Archivtage)$  nach Formel (3.1)
6        $fensterBewertung.add(z, s)$ 
7     end
8   end
9   // Glätten der Werte
10  erzeuge Liste  $fensterBewertungGeglaettet$ 
11  foreach  $p$  in  $[1 \dots fensterBewertung.Length]$  do
12     $s' \leftarrow \sum_{p-2}^{p+2} fensterBewertung[p]$ 
13     $fensterBewertungGeglaettet.add(z, s')$ 
14  end
15  // Ermittlung des Besten Fensters
16   $besteStreuung \leftarrow s_{G^*}$ 
17  foreach  $(z, s')$  in  $fensterBewertungGeglaettet$  do
18    if  $s' < besteStreuung$  then
19       $besteStreuung \leftarrow s'$ 
20       $bestesFenster \leftarrow z$ 
21    end
22  end
23  if  $besteStreuung = s_{G^*}$  then
24    return 183
25  else
26    return  $bestesFenster$ 
27  end
28 end

```

4.2.4. Überprüfung der Tagesgruppe

In der nächsten Stufe wird die Tagesgruppe innerhalb des zuvor berechneten Zeitfensters neu ermittelt (siehe Algorithmus 4).

Dazu werden zunächst alle Tagesgruppierungen $G \in \mathbb{G}$ auf Äquivalenz zum Prognosetag \mathring{d} geprüft. Ist die Prüfung erfolgreich, werden die zugehörigen Archivtage innerhalb des Zeitfensters z^* ermittelt (siehe Kapitel 4.1.2). Andernfalls wird mit der nächsten fortgefahren.

Sobald Archivtage ermittelt werden konnten, wird die Streuung s innerhalb der zugehörigen Archivdaten berechnet. Dieser Wert wird daraufhin mit dem bisher besten Streuungswert verglichen und auf diese Weise die Tagesgruppierung ermittelt, die am besten zur Prognose geeignet ist.

Algorithmus 4 : Dritte Stufe des Algorithmus, Neuberechnung der optimalen Tagesgruppe

Eingabe : \mathring{d} , der Prognosetag

Eingabe : z^* , bislang bestes Zeitfenster

Ausgabe : G^{**} : *besteGruppe*, enthält beste Prognosekonfiguration für \mathring{d}

```

1 function überprüfeTagesgruppe( $\mathring{d}$ , StreuungProGruppierung) begin
2     // Berechne die Streuung für alle Tagesgruppen
3     besteStreuung  $\leftarrow$  maxDouble
4     foreach  $G_{\sim} \in \mathbb{G}$  do
5         if  $\mathring{d} \sim G_{\sim}$  then
6             Archivtage  $\leftarrow G_{\sim} \cap \overline{B}_{z^*}(J(\mathring{d}))$ 
7             if Archivtage  $\neq \emptyset$  then
8                  $s \leftarrow$  berechne Streuung über  $\mathcal{A}(\textit{Archivtage})$  nach Formel (3.1)
9                 if  $s < \textit{besteStreuung}$  then
10                     besteStreuung  $\leftarrow s$ 
11                     besteGruppe  $\leftarrow G$ 
12                 end
13             end
14         end
15     end
16 return besteGruppe
end
```

4.2.5. Berechnung der Prognosezeitreihe

Zum Zeitpunkt der Prognose steht die errechnete optimale Tagesgruppierung G^{**} und die Zeitfensterbreite z^{**} für den Prognosetag \mathring{d} fest. Aus diesen Informationen wird nun Prognoselastgang \hat{x} errechnet (siehe Algorithmus 5).

Zunächst wird das Datenarchiv auf Basis von Tagesgruppierung und Zeitfenster errechnet. Die Funktion \mathcal{A} bildet dabei jeden Archivtag auf einen Tageslastgang $x \in R^{96}$ ab (siehe Kapitel 4.1.1).

Die Berechnung der Prognosezeitreihe \hat{x} für \mathring{d} erfolgt für jeden der 96 Viertelstundenwerte \hat{x}_t separat. Hierzu wird pro Viertelstunde t der Durchschnittswert aus allen Lastgängen x im Archiv *Datenarchiv* zur selben Viertelstunde gebildet. Dieser ergibt dann den Prognosewert zu entsprechenden Viertelstunde im Prognoselastgang \hat{x} .

Algorithmus 5 : Berechnung der Prognosezeitreihe

Eingabe : \mathring{d} , der Prognosetag
Eingabe : G^{**} , enthält beste Tagesgruppierung für \mathring{d}
Eingabe : z^{**} , enthält die beste Zeitfensterbreite
Ausgabe : \hat{x} , der prognostizierte Tageslastgang für \mathring{d}

```

1 function prognostiziere( $\mathring{d}, G^{**}, z^{**}$ ) begin
    // Ermittlung Datenarchiv
2    $Archivdaten \leftarrow \mathcal{A}(G_{\sim}^{**} \cap \overline{B}_{z^{**}}(J(\mathring{d})))$ 
    // Berechnung Prognosezeitreihe
3   foreach  $t \in [1, \dots, 96]$  do
4      $\hat{x}_t \leftarrow \frac{1}{|Archivdaten|} \sum (x_t \in Archivdaten)$ 
5   end
6   return  $\hat{x}$ 
7 end
```

4.3. Zusammenfassung

Im vorangegangenen Kapitel wurde das Optimierungsproblem der Zeitreihenprognose definiert und ein vierstufiger Algorithmus zur Lösung des Problems vorgestellt. Die Zielfunktion des Optimierungsproblems minimiert die durchschnittliche relative Abweichung der Ist-Werte vom jeweiligen Prognoselastgang innerhalb eines Tages.

Es wurde eine Prognosefunktion entwickelt, die einen Prognosetag aus dem Durchschnittslastgang repräsentative Archivtage bildet. Dazu wurden Methoden vorgestellt, um das Datenarchiv auf geeignete Tage einzuschränken:

Im Wesentlichen wurde versucht, durch die Einführung von Tagesgruppierungen und Zeitfenster die Tages-, Wochen- und Jahreszyklen von Energiezeitreihen abzubilden. Als Bewertungsmethode für gute Tagesgruppierungen und Zeitfenster dient dabei die in Kapitel 3.2.1 eingeführte Streuungs-Analyse.

Anschließend wurde der Algorithmus beschrieben, der eine Lösung des Optimierungsproblems implementiert. Er führt die Lösung in mehreren Stufen herbei, und soll damit sukzessiv die Prognosequalität verbessern:

Zunächst wurde die Struktur und Abfolge der einzelnen Stufen aufgezeigt und deren Funktion innerhalb des Algorithmus erklärt. In den nachfolgenden Abschnitten wurde die Ermittlung der Prognoseparameter mit Hilfe von Pseudocode beschrieben, mittels derer das Datenarchiv reduziert wird. Anschließend wird die Berechnung des Lastgangs aus den ausgewählten Archivdaten erläutert.

5. Analyse von Datenbestand und Algorithmus

Die Auswertung des in Kapitel 4.2 vorgestellten Algorithmus erfolgt nach einer bestimmten Methode, welche im folgenden Abschnitt beschrieben wird. Die Struktur der ausgewählten Daten wird vorgestellt. Weiterhin wird ein Auswerteverfahren festgelegt, anhand dessen die Prognosequalität bestimmt wird.

5.1. Deskriptive Datenanalyse

Die deskriptive Datenanalyse beschreibt die Struktur des Zeitreihenportfolios, das der Evaluation des vorgestellten Algorithmus zugrunde liegt. Dazu werden verschiedene charakteristische Kennzahlen wie Jahresverbräuche, Benutzungsstunden und die Anzahl der Archivtage über alle Zeitreihen ermittelt und grafisch aufbereitet. Ebenfalls kommen die in Kapitel 3 vorgestellten Analysatoren zum Einsatz.

5.1.1. Kundenprofil

Das auszuwertende Zeitreihenportfolio umfasst die Zeitreihen von 355 Verbrauchern. Abbildung 5.1 zeigt die verschiedenen Branchen, die im Datenarchiv vertreten sind.

Die größte Gruppe stellen die Gewerbekunden dar. Darunter fallen Groß- und Einzelhandel sowie Banken, Hotels, Gastronomie und Handwerksbetriebe. Unter dem Begriff Industrie sind in erster Linie produzierende Gewerbe zusammengefasst. Hier finden sich chemische Industrie, Maschinen-, Metall- und Werkzeugbau sowie große Druckereien, Sägewerke und Keltereien sowie Zeitreihen landwirtschaftlicher Betriebe.

Zu den öffentlichen Einrichtungen gehören zumeist Behörden und Schulen, während Krankenhäuser, Altenpflegeheime und andere soziale Einrichtungen gesondert ausgewiesen sind.

Windkraft- und Biogasanlagen sind in dieser Auswertung als separate Kategorie aufgeführt. Dabei handelt es sich nicht um Daten, die bei der Einspeisung ins Stromnetz

anfallen, sondern um Verbräuche, die zum Betrieb der Anlage notwendig sind, wenn die Anlage beispielsweise selbst keinen Strom produziert. Ein nicht unerheblicher Teil von 60 Zeitreihen, also 17 %, konnte aufgrund fehlender Informationen nicht zugeordnet werden.

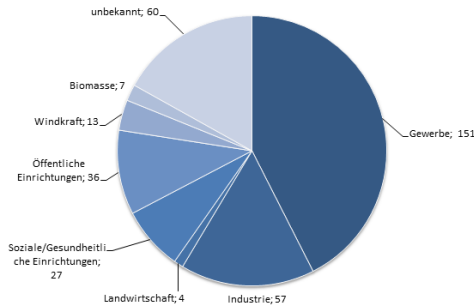


Abbildung 5.1.: Kategorisierung der Zeitreihen nach Branchen

5.1.2. Datenarchiv

Abbildung 5.2a zeigt den durchschnittlichen Jahresverbrauch der untersuchten Zeitreihen. Die Bundesnetzagentur verfügt, dass Zähler mit einem Verbrauch von über 100 MWh im Jahr Lastgang-gemessen sein müssen. Ist der Verbrauch geringer, kann der Betreiber des Stromnetzes entscheiden, welche Art der Messung er bevorzugt (Deutscher Bundestag 2013). Während etwas mehr als ein Viertel der Zeitreihen einen durchschnittlichen Jahresverbrauch von über 250 MWh im Jahr aufweisen, liegen 40 % unter der Marke von 100 MWh.

Zum Test des Algorithmus wurden ausschließlich solche Zeitreihen ausgewählt, die über einen Zeitraum von mindestens zwei Jahren an mehr als 90 % der Tage einen Lastgang besitzen. Das Archiv reicht bis ins Jahr 2007 zurück, wobei fast drei Viertel der Daten den Jahren 2012 und 2013 entstammen (siehe Abbildung 5.2b).

Abbildung 5.2c zeigt eine Verteilung der Archivtage pro Zeitreihe, die innerhalb des Datenarchivs zur Verfügung stehen. Ersichtlich wird, dass über die Hälfte der Zeitreihen ein Datenarchiv von maximal 900 Tagen besitzt. Nur 13 % der Zeitreihen verfügen über 1100 Archivtage, was etwa 3 Jahren entspricht.

Eine weitere wichtige Kennzahl für die Prognostizierbarkeit von Zeitreihen ist die Anzahl der Benutzungsstunden. Sie gibt Auskunft über die Gleichmäßigkeit der Entnahme und wird berechnet als

$$\text{Benutzungsstunden} = \frac{\text{Jahresverbrauch [kWh]}}{\text{maximale Leistung [kW]}}$$

Für eine Zeitreihe mit einem Jahresverbrauch von 100.000 kWh und 25 kW Maximalleistung, könnte dieser Verbraucher 4000 von 8760 Stunden im Jahr unter Vollast gelaufen sein. Ein anderes Szenario wäre ein konstanter Verbrauch von 11,4 kW. Im Allgemeinen ist bei 4000 Benutzungsstunden ein eher kontinuierlicher Verbrauch mit geringen Schwankungen naheliegend und damit ein Hinweis auf gute Prognostizierbarkeit.

Ein Überblick über die Benutzungsstunden der Zeitreihen im untersuchten Datenarchiv ist in Abbildung 5.2d gegeben. Lediglich 14 % der Zeitreihen haben über 4200 Benutzungsstunden, während 22 % einen Wert von unter 1200 Stunden aufweisen. Die Verteilung der Benutzungsstunden legt nahe, dass es sich bei dem Datenarchiv um ein recht anspruchsvoll zu prognostizierendes Portfolio handelt. Eine Übersicht über die Maximalleistung der Zeitreihen ist in Abbildung 5.2e dargestellt.

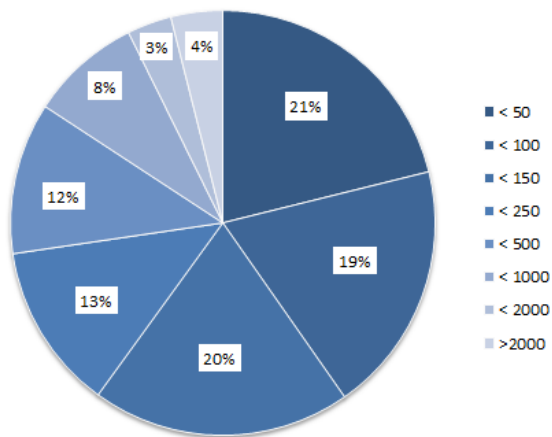
5.1.3. Korrelationsanalysen

Wie in Kapitel 3 dargelegt, wird der Stromverbrauch von verschiedenen Faktoren beeinflusst. Dazu zählen die Außentemperatur und kalendarische Ereignisse sowie der Einfluss der Zeitreihe auf sich selbst. Letzteres drückt sich vor allem durch Verbrauchszyklen innerhalb der Zeitreihe aus.

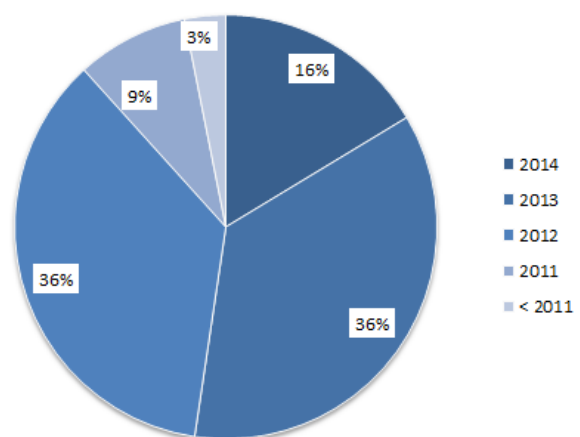
Abbildung 5.3 zeigt die Ergebnisse verschiedener Korrelationsanalysen aller Zeitreihen im Datenarchiv. Die angegebenen Korrelationskoeffizienten sind die Betragswerte der errechneten Koeffizienten, da zur Übersicht nur die Intensität der Abhängigkeit, nicht die Richtung (gleichläufig oder gegenläufig) als relevant angesehen wird.

Die Zyklizität von Zeitreihen innerhalb des Datenarchives kann mittels Autokorrelationskoeffizient sichtbar gemacht werden (siehe Kapitel 3.2.3). Die Histogramme in Abbildung 5.3a und 5.3b zeigen die Verteilung der Autokorrelationskoeffizienten bei einer Verschiebung der Zeitreihe um einen bzw. sieben Tage. Am unteren Ende ($\leq 0,2$) der Skala befinden sich vergleichsweise viele Windparks und Biomasse-Anlagen, während am oberen Ende hauptsächlich ($\geq 0,8$) öffentliche Einrichtungen zu finden sind.

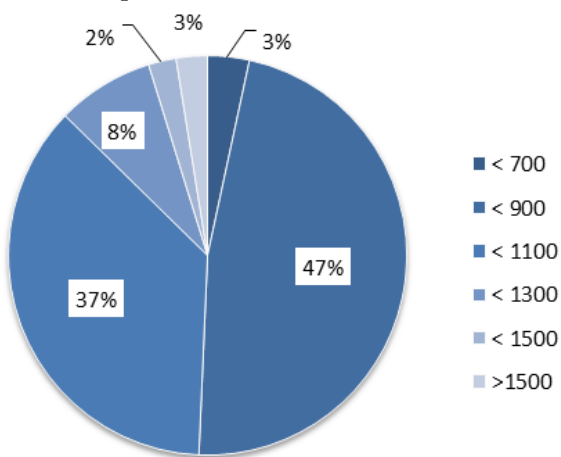
Des weiteren wurde der Einfluss der Schulferien im Zeitreihenarchiv untersucht (siehe Kapitel 3.2.2). Abbildung 5.3c zeigt das Ergebnis der Korrelationsanalyse. Im Bereich größer als 0,4 sind, wie erwartet, hauptsächlich öffentliche Einrichtungen zu finden, die Schule im Eingangsbeispiel hat einen Koeffizienten von 0,57. Im unteren Bereich ($< 0,1$) finden sich vor allem Einzelhandel, Banken und Windparks.



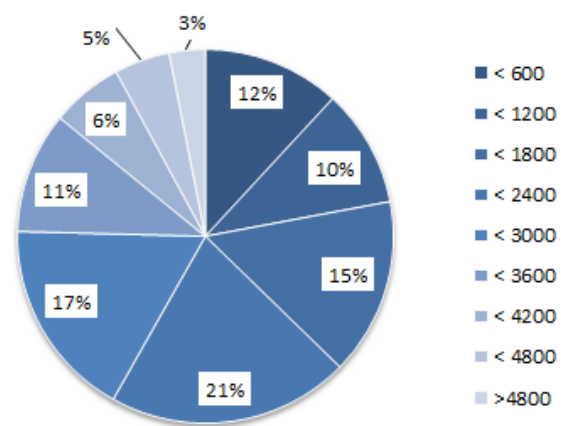
(a) Durchschnittliche Jahresverbräuche in MWh pro Zeitreihe



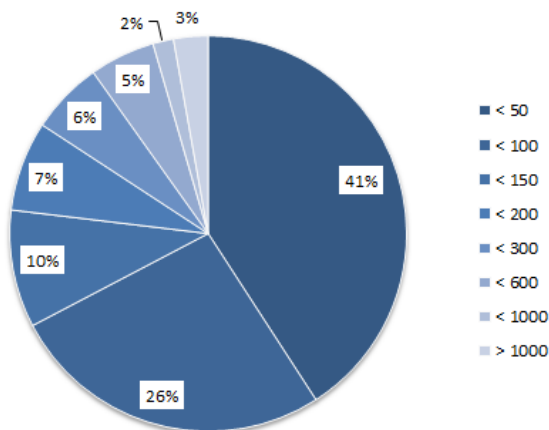
(b) Aufteilung der Archivtage auf Kalenderjahre



(c) Archivgröße in Tagen pro Zeitreihe

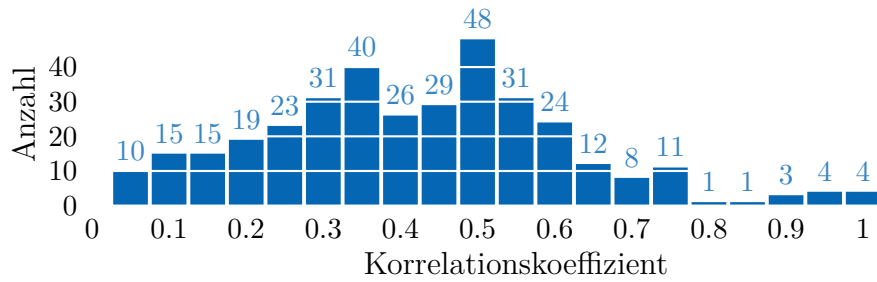


(d) Benutzungsstunden pro Zeitreihe

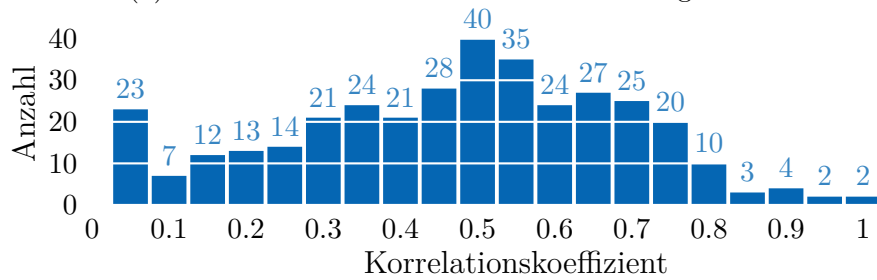


(e) Maximalleistung pro Zeitreihe in kW

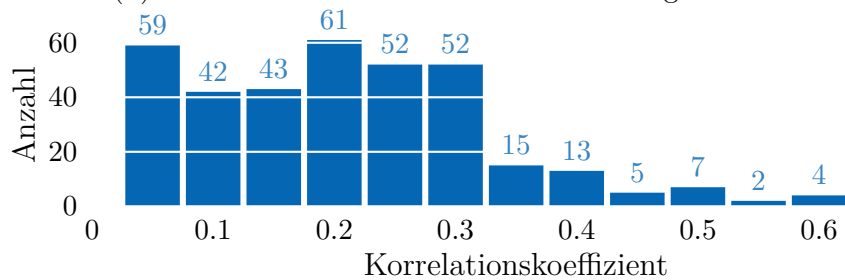
Abbildung 5.2.: Übersicht über das Datenarchiv



(a) Autokorrelation der Zeitreihen bei 1 Tag Versatz



(b) Autokorrelation der Zeitreihen bei 7 Tagen Versatz



(c) Korrelation der Zeitreihen mit Schulferien

Abbildung 5.3.: Ergebnisse der Korrelationsanalysen über das Datenarchiv

5.2. Methodik Auswertung

Im Folgenden wird die Umgebung, in der die Zeitreihenprognose durchgeführt und bewertet wird, näher erläutert. Zur Umgebung gehört die Konfiguration wichtiger Stammdaten, wie Tagesgruppen und die verwendeten Kalender, die der Prognosealgorithmus benötigt. Auch die Konfiguration der Originalprognose, deren Ergebnis als Vergleich dient, ist beschrieben. Anschließend wird das Verfahren vorgestellt, nach dem im späteren Verlauf der Arbeit die Prognosegüte berechnet und ausgewertet wird.

5.2.1. Konfiguration der Tagesgruppierungen

In Kapitel 4.1.2 wurden Tagesgruppen als Äquivalenzrelationen definiert, die die Gleichheit verschiedener Archivtage anhand von Tageseigenschaften festlegen. Die ausgewerteten

Eigenschaften und deren Werte sind in Tabelle 5.1 aufgelistet. Es wird der Wochentag, ein Tagestyp und die Information, ob ein Brückentag vorliegt, verwendet. Alle Eigenschaften können in beliebigen Kombinationen auftreten.

Eigenschaft	Werte
Wochentag	Montag, Dienstag, ..., Samstag, Sonntag
Tagestyp	Werktag, Samstag, Sonntag, Brückentag, Ferien-Werktag, Ferien-Samstag, Ferien-Sonntag, Ferien-Brückentag
Brückentag	Ja, Nein

Tabelle 5.1.: Übersicht über Tageseigenschaften

Kalender

Offensichtlich kann die Bestimmung der Tageseigenschaften eines Datums nur mit Hilfe eines Kalenders erfolgen. Hier können regionale Besonderheiten wie Feiertage, Bückentage oder Schulferien abgebildet werden. Feiertage werden in diesen Kalendern üblicherweise mit dem Tagestyp Sonntag gekennzeichnet.

In Tabelle 5.2 ist die Zuordnung der Tageseigenschaften anhand von verschiedenen Kalendern für den 06.01.2015 beispielhaft aufgeführt. Dieser Tag ist in Nordrhein-Westfalen ein normaler Werktag, während er in Bayern ein Feiertag und deshalb dem Tagestyp Sonntag zugeordnet ist. Gleichzeitig liegt er in Rheinland Pfalz innerhalb der Schulferien, was in Bayern nicht der Fall ist.

Kalender	Wochentag	Tagestyp	Brückentag
Nordrhein-Westfalen	Dienstag	Werktag	Nein
Nordrhein-Westfalen Ferien	Dienstag	Werktag	Nein
Bayern	Dienstag	Sonntag	Nein
Bayern Ferien	Dienstag	Sonntag	Nein
Rheinland Pfalz	Dienstag	Werktag	Nein
Rheinland Pfalz Ferien	Dienstag	Ferien-Werktag	Nein

Tabelle 5.2.: Beispiel der Tageseigenschaften des 06.01.2015 anhand von unterschiedlichen Kalendern

Schema Tagesgruppierungen

Tagesgruppierungen werden der Übersichtlichkeit halber zu einem Schema zusammengefasst, nach dem ein Verbraucher prognostiziert wird. Ein mögliches, sehr einfaches Schema wäre zum Beispiel, alle Werktage und die Wochenenden jeweils zusammenzufassen. Dies

ergibt dann ein Schema *Werktage - Wochenende*, das zwei Tagesgruppierungen enthält. Eine denkbare Anwendung für dieses Schema wäre ein Bürogebäude.

In *FirstSale* wurden zur Prognose von Zeitreihen verschiedene Schemata mit ihren Tagesgruppierungen angelegt, von denen der Algorithmus eine passende Tagesgruppierung pro Prognosetag auswählen soll. Alle konfigurierten Schemata, die zur Prognose von Zeitreihen im Rahmen dieser Arbeit herangezogen wurden, sind in Kapitel A in Tabelle A.1 aufgelistet.

5.2.2. Konfiguration der Originalprognose

Um die Performance des Algorithmus bewerten zu können, muss zunächst die Prognosegüte des alten Verfahrens berechnet werden, im Folgenden auch Originalprognose genannt. Sie legt für alle Zeitreihen unabhängig von ihren Charakteristika dasselbe Prognoseverfahren zugrunde.

Tabelle 5.3 zeigt die für den Prognosezeitraum konfigurierten Tagesgruppenschemata, die eingesetzt werden. Ersichtlich wird, dass zu verschiedenen Phasen im Jahr unterschiedliche Schemata angewendet werden. Eine Variation findet sich hauptsächlich um den Jahreswechsel. Hier wurde versucht, den speziellen Lebensumständen zu dieser Zeit Rechnung zu tragen. Eine umfangreichere Konfiguration der Schemata kann in Tabelle A.1 nachgeschlagen werden.

Der Kalender, der zur Ermittlung der Tageseigenschaften genutzt wird, ist der Kalender des Bundeslandes, in dem der Verbraucher liegt. Ein Zeitfenster ist nicht eingetragen. Stattdessen ist die maximale Anzahl an Archivtagen, die zur Prognose genutzt werden, auf die zwei Tage begrenzt, die im Jahresverlauf am nächsten liegen (analog zu Kapitel 4.1.2). Auf diese Weise wird versucht die Saisonalität von Zeitreihen abzubilden.

Zeitraum		Schema Tagesgruppe
01.08.2012	22.12.2013	Mo, Di - Do, Fr, Sa, So
23.12.2013	27.12.2013	je Wochentag
28.12.2013	31.12.2013	Mo, Di - Do, Fr, Sa, So
01.01.2014	01.01.2014	je Wochentag
02.01.2014	12.01.2014	Mo, Di - Do, Fr, Sa, So
13.01.2014	13.01.2014	Mo - Fr, Sa, So
14.01.2014	31.12.4999	Mo, Di - Do, Fr, Sa, So

Tabelle 5.3.: Konfiguration der Originalprognose

5.2.3. Auswertungsverfahren der Prognosegüte

Im Folgenden soll das Verfahren näher erläutert werden, nach dem die Prognosegüte des neuen Prognosealgorithmus ausgewertet wird.

Archiv- und Prognosezeitraum

Mit Blick auf die unterschiedliche Anzahl von Archivtagen pro Zeitreihe gilt es, einen sinnvollen Prognosezeitraum festzulegen und den dazugehörigen Archivzeitraum zu definieren. Er muss dabei mehrere Kriterien erfüllen:

- Der Prognosezeitraum einer Zeitreihe muss innerhalb ihres Datenarchivs liegen. Andernfalls ist ein Vergleich mit Ist-Werten nicht möglich.
- Prognose- und Archivzeitraum dürfen sich nicht überschneiden. Die Zeitreihe des Prognosetages darf nicht zur Prognose herangezogen werden.
- Jede Zeitreihe muss die gleiche Anzahl von Prognosetagen haben, damit die Ergebnisse nicht verzerrt werden.

Zum Test des Algorithmus wird eine Langfristprognose simuliert. Sie ist dadurch gekennzeichnet, dass auf Basis der vorhandenen Zeitreihen ein zusammenhängender Zeitraum in der Zukunft prognostiziert wird. Jede Zeitreihe im Datenarchiv wird demnach in einen festen Prognose- und einen Analysezeitraum aufgeteilt. Das Datenarchiv, das vom Prognosealgorithmus verarbeitet wird (siehe Kapitel 4.1.1), enthält folglich ausschließlich Archivtage aus dem Analysezeitraum.

Zwischen Prognose und Analysezeitraum wird eine einmonatige Lücke generiert. Ohne diese Lücke lägen die Archivtage zu nahe am Prognosetag. Die Ergebnisse wären dann kein realistischer Test für eine langfristige Prognose. Als Prognosezeitraum sind dadurch elf Monate festgelegt, was einen Archivzeitraum von mindestens einem Jahr pro Zeitreihe ergibt.

Der Umstand, dass kein volles Jahr prognostiziert wird, könnte dazu führen, dass die Ergebnisse durch saisonale Einflüsse verzerrt werden. Dieser Effekt wird dadurch vermieden, dass die verfügbaren Zeitreihen im Datenarchiv unterschiedliche Zeiträume abdecken (siehe auch Kapitel 4.1.1).

Prognosegüte

Die Prognosegüte wird im Rahmen dieser Arbeit aus der Summe der Fehlmenge, die zu viel oder zu wenig prognostiziert wurde, und dem Ist-Wert der Zeitreihe gebildet. Analog

zur Zielfunktion in Kapitel 4.1.1 wird die Fehlmenge als betragsmäßige Differenz aus den prognostizierten viertelstündlichen Werten \hat{x}_i für eine Zeitreihe und den gemessenen Werten \check{x}_i berechnet.

Da eine einfache Differenz Großverbrauchern ein übermäßiges Gewicht in den Ergebnisdatsätzen verschaffen würde, wird an dieser Stelle die prozentuale Abweichung, nicht die absolute Abweichung, betrachtet. Die durchschnittliche prozentuale Abweichung für jeden Prognosetag ergibt sich gemäß Formel (5.1) aus der Fehlermenge eines Tages und den 96 Viertelstunden.

$$\frac{1}{96} \sum_{i=1}^{96} \frac{|\check{x}_i - \hat{x}_i|}{\check{x}_i} \quad (5.1)$$

Nachteilig bei diesem Verfahren ist der Umstand, dass die berechnete prozentuale Prognosefehler bei geringen Verbräuchen nahe 0 kWh sehr groß werden kann. Dies sollte bei einem genügend umfangreichen Datenarchiv vernachlässigbar sein.

In den folgenden Kapiteln wird die Prognosequalität pro Prognosetag berechnet und ausgewertet. Erwartet wird, dass bei einer Verbesserung der Prognose mehr Tage eine geringere prozentuale Abweichung aufweisen als bei Durchführung der Prognose mit der ursprünglichen Konfiguration, wie in Kapitel 5.2.2 beschrieben.

Auswertung nach Algorithmus-Stufen

Der neue Prognosealgorithmus, wie in Kapitel 4.2 vorgestellt, verfeinert die Prognose in mehreren Schritten: Zunächst wird eine Tagesgruppe festgelegt, danach das Zeitfenster. Anschließend werden die Einstellungen noch einmal überprüft und verbessert. Es wird erwartet, dass sich die Prognosegüte bei jeder der Algorithmusstufen sukzessiv verbessert. Dazu wird nach jeder Stufe eine Prognose durchgeführt und deren Prognosequalität berechnet.

Als Ergebnis dieser Auswertung entstehen je untersuchter Zeitreihe vier Prognoselastgänge: Einer als Resultat der Originalkonfiguration und drei Lastgänge, die durch den Algorithmus berechnet werden. Deren Prognosequalität kann anschließend ermittelt und gegenübergestellt werden.

Finanzielle Auswertung der Prognosegüte

Obwohl eine Optimierung nach finanziellen Kriterien nicht zulässig ist, ist die monetäre Bewertung der Fehlermengen durchaus von Interesse. Sie zeigt die Kosten auf, die einem Energielieferanten durch ungenaue Prognosen entstehen. Erwartet wird, dass der vorge-

stellte Prognosealgorithmus nicht nur die Prognosegüte verbessert, sondern auch Kosten seitens der Energielieferanten einspart.

Die finanzielle Bewertung einer Prognosezeitreihe wird hier anhand des Beispiels erläutert. Abbildung 5.4 zeigt die Prognosezeitreihe eines Verbrauchers und dessen gemessene Werte sowie eine Preiszeitreihe für einen Tag. Somit wird ersichtlich, dass tagsüber zu wenig Leistung und in den Morgen- und Abendstunden zu viel prognostiziert wurde. Zur Bewertung der Differenz wird nun für jede Fehlmenge ein Kauf beziehungsweise Verkauf unterstellt. Der Preis für die gehandelte Energie für jede Viertelstunde ergibt sich aus der Preiszeitreihe. Dabei wird angenommen, dass der simulierte Handel den Preis nicht beeinflusst.

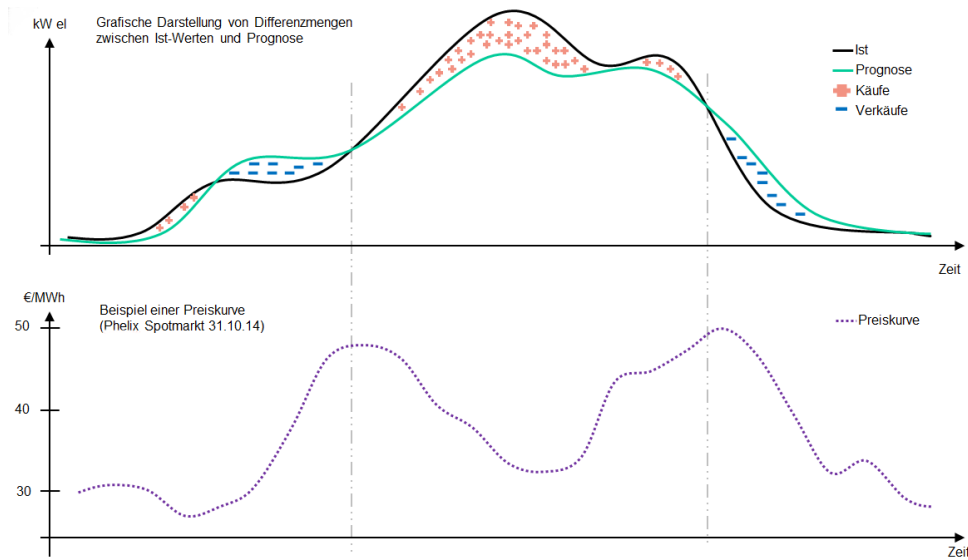


Abbildung 5.4.: Schematische Ansicht der finanziellen Bewertung einer Differenzzeitreihe

Im Allgemeinen berechnen sich die entstehenden Kosten aus der Summe der Differenz zwischen Prognosewert \hat{x}_i und Ist-Wert $\hat{\hat{x}}_i$, multipliziert mit dem Preis p_i je Viertelstunde i . Die Konstante A stellt den Umrechnungsfaktor der Leistung in Arbeit dar. Bei 96 Werten pro Tage ist $A = 15[\text{min}]$.

$$\sum_{i=1}^{96} (\hat{\hat{x}}_i - \hat{x}_i) p_i * A \quad (5.2)$$

Zur Bewertung wird die Preiszeitreihe der Regelenenergie genutzt. Sie beinhaltet die größten Preisschwankungen (siehe Kapitel 1.2.1) und ergibt ein realistisches Bild der Kosten, die entstünden, wenn die Langfristprognose nicht mehr verbessert würde.

5.3. Zusammenfassung

Im vorangegangenen Kapitel wird die Vorgehensweise bei der Auswertung des in Kapitel 4.2 vorgestellten Prognosealgorithmus beschrieben und das zugrundeliegende Datenarchiv analysiert.

Das untersuchte Zeitreihenportfolio wird in Wirtschaftsbereiche unterteilt und geographisch eingeordnet. Der Großteil der Zeitreihen lässt sich Gewerbe und Industrie zuordnen.

Zur Datenbasis werden verschiedene aussagekräftige Kennzahlen ermittelt: Jahresverbräuche, Maximalleistung, Benutzungsstunden, Zeitbereich und Archivgröße. Das Datenarchiv stammt zum überwiegenden Teil aus den Jahren 2012 und 2013 und umfasst zumeist zwischen zwei und drei Jahre an Archivtagen je Zeitreihe. Verbräuche und Maximalleistung lassen auf ein gut durchmisches Datenarchiv schließen, während die Benutzungsstunden auf ein recht anspruchsvolles Portfolio hinweisen.

Zuletzt kommen die in Kapitel 3 entwickelten Analysatoren zum Einsatz. Die Aufbereitung des Datenarchivs erfolgt anhand von verschiedenen Korrelationskoeffizienten mit einer anschließenden Diskussion der Ergebnisse.

Im zweiten Teil des Kapitels wird die Systemumgebung des Prognosealgorithmus und dessen Auswertung beschrieben. Die Umsetzung der in Kapitel 4.1.2 definierten Äquivalenzklassen wird in *FirstSale* anhand von Wochentag, Tagestyp und der Information, ob ein Brückentag vorliegt, vorgenommen. Die Notwendigkeit mehrerer Kalendertage wurde diskutiert, die einem Prognosetag die Eigenschaften zuordnen.

Anschließend wird die Konfiguration der Originalprognose und das Verfahren beschrieben, nach dem der neuen Algorithmus anhand des Datenarchiv ausgewertet wird. Für die Ermittlung der Prognosegüte wird eine Langfristprognose simuliert. Der gewählte Prognosezeitraum beträgt elf Monate und fällt zeitlich mit dem Datenarchiv zusammen, so dass ein Vergleich mit den Ist-Daten möglich ist.

Die Auswertung erfolgt nach jeder Algorithmusstufe mit den zu diesem Zeitpunkt ermittelten Prognoseparametern. Diese ergibt drei Prognoselastgänge, die den Ergebnissen der Originalprognose gegenüber gestellt werden können. Zuletzt wird eine Methode zur finanziellen Bewertung der Lastgänge anhand von Regelenergiepreisen vorgestellt.

Dabei wurde näher auf die Konfiguration der bisherigen Prognose und die Evaluierung der beiden Prognoseverfahren eingegangen. Zusätzlich wurde die Methoden, mit denen die Prognosegüte in dieser Arbeit finanziell bewertet wird, vorgestellt.

6. Ergebnisse der Algorithmusanalyse

In diesem Kapitel wird die Prognosegüte des neuen Algorithmus ausgewertet und der der Originalprognose gegenübergestellt. Dabei werden die drei Prognosestufen des neuen Algorithmus separat aufgeführt und finanziell bewertet. Anschließend werden die Ergebnisse diskutiert und deren Performance ergründet.

6.1. Auswertung der Prognosegüte

Zur Messung der Prognosegüte wurde der entwickelte Prognosealgorithmus auf ein Archiv aus 355 Zeitreihen angewendet. Je Zeitreihe wurde die Prognose für ein Kalenderjahr durchgeführt und die durchschnittliche prozentuale Abweichung pro Tag gemessen. Die Messung wurde für die Prognose mit ursprünglicher Konfiguration und nach jeder Prognosestufe durchgeführt (siehe Kapitel 5.2.3). Zu den verschiedenen Algorithmusstufen wurde zunächst die Tagesgruppierung festgelegt, ein Zeitfenster zur Abbildung von saisonalen Abhängigkeiten ermittelt und anschließend auf Basis der gesetzten Parameter neu berechnet (siehe Kapitel 4.2).

Die Ergebnisse zur Prognosequalität liegen als Kombination aus Zeitreihe, Prognosetag und prozentualer Abweichung vor. Je ausgewerteter Prognose wurden ca. 115.000 Datensätze berechnet. Die Anzahl der ausgewerteten Prognosetage liegt unterhalb der erwarteten 118.570 Tage, da es immer wieder Lücken innerhalb der Prognosezeiträume gibt, zu denen keine Daten verfügbar sind.

Eine Übersicht über die errechneten Prognosefehler ist in Tabelle 6.1 gegeben. Die dargestellten Ergebnisse sind teils widersprüchlich, der maximale Fehler und der Median sind in der Originalprognose am höchsten, während der durchschnittliche Fehler hier am niedrigsten ist. Ebenfalls ist die Bestimmung, welche der Algorithmusstufen die beste Prognose erbringt, nicht möglich: Der Durchschnittswert ist nicht verwertbar, da er offensichtlich zu von Ausreißern stark beeinflusst wurde; die Mediane liegen zu nahe beieinander, um aussagekräftig zu sein.

Stufe	Prognosefehler in %			
	Min	Ø	Median	Max
Original Prognose	0	132	30	69.236
Stufe 1: Tagesgruppierung	0	162	25	33.547
Stufe 2: Zeitfenster	0	150	24	42.778
Stufe 3: Überarbeitung	0	151	25	38.721

Tabelle 6.1.: Übersicht über Prognosefehler der Originalprognose und der verschiedenen Algorithmusstufen

Aus diesem Grund wird zur Evaluierung des Ergebnisses eine andere Darstellung der Ergebnisse gewählt. Abbildung 6.1 stellt die Performance der unterschiedlichen Messungen in einem kumulierten Histogramm dar. Auf der horizontalen Achse ist der prozentuale Prognosefehler aufgetragen; die vertikale Achse gibt den prozentualen Anteil an Tagen an, die höchstens über den angegebenen Fehler verfügen. Die Ergebnisse der Gruppen sind jeweils als Kurve dargestellt. So ist es beispielsweise möglich, die Quantile direkt abzulesen: Der Median einer jeden Kurve liegt auf Höhe von 50% der Prognosetage auf der vertikalen Achse.

Eine Prognose wird folglich mehr Tage mit geringer Abweichung und weniger Tage mit hoher Abweichung besitzen als eine Prognose, die schlechter ist. Daraus folgt, dass die Kurve in einem kumulierten Histogramm zunächst stärker ansteigt. Verläuft sie grundsätzlich oberhalb, und nähert sich damit früher 100% aller Prognosetage, kann die zugrundeliegende Prognose als besser angesehen werden.

Demnach zeigt Abbildung 6.1, dass die Prognosegüte durch Anwendung des neuen Algorithmus deutlich verbessert werden konnte. Auch die Optimierung in den verschiedenen Prognosestufen zeigt, wie erwartet, ein besseres Ergebnis.

6.2. Finanzielle Bewertung

Bei der finanziellen Bewertung sollen die Kosten betrachtet werden, die durch die Prognose innerhalb des Prognosezeitraums entstehen. Zur Ermittlung wird angenommen, dass die fehlprognostizierte Energie an der Börse verkauft bzw. beschafft wird. Hierzu wird eine Differenzzeitreihe von Prognosestufen und Ist-Werten erstellt und mit einer Preiszeitreihe multipliziert (siehe Kapitel 5.2.3).

In Abbildung 6.2 sind die errechneten Kosten für die verschiedenen Prognosestufen der Originalprognose gegenübergestellt. Ersichtlich wird, dass die Originalprognose zu zusätzlichen Kosten von ca 140.000 € geführt hätte, während bei einem Energieversorger

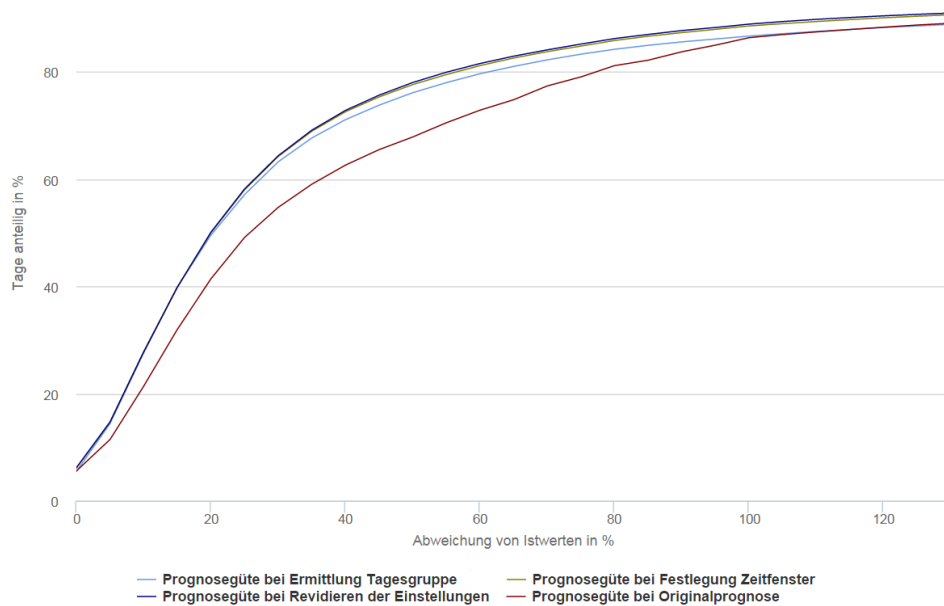


Abbildung 6.1.: Relative Prognosegüte im Vergleich neuer Algorithmus und alte Prognose

bei vorliegendem Zeitreihenportfolio durch den neuen Algorithmus lediglich ca. 11.000 € Kosten angefallen wären.

6.3. Diskussion

Die Auswertung der Prognosequalität für die Langfristprognose zeigt, dass der eingesetzte Algorithmus die Prognose im Vergleich zu dem zuvor eingesetzten Verfahren wesentlich verbessert hat. In den verschiedenen Stufen des Algorithmus wurde das Datenarchiv, das der Prognose eines einzelnen Prognosetages zugrunde liegt, weiter eingeschränkt. Es konnte gezeigt werden, dass die Prognose mit jeder Stufe genauer wird.

Auch die finanzielle Auswertung ergibt eine Verbesserung durch den Algorithmus. Rein nominell würde man die erste Stufe vorziehen, da hier am meisten Erlös durch den Verkauf von Regelenergie zustande kam. Da man aber in erster Linie daran interessiert ist, den Prognosefehler zu minimieren und damit das Preisrisiko zu reduzieren, ist auch hier die dritte Algorithmusstufe zu bevorzugen.

Trotz der guten Ergebnisse sollte nicht außer Acht gelassen werden, dass die Spanne der prozentualen Prognosefehler (siehe Tabelle 6.1) sehr groß ist. Eine Ursache liegt sicherlich bei sehr geringer Abnahmeleistung nahe 0 kW. Dadurch ergibt eine Abweichung der Prognosezeitreihe um wenige kW einen sehr großen prozentualen Prognosefehler.

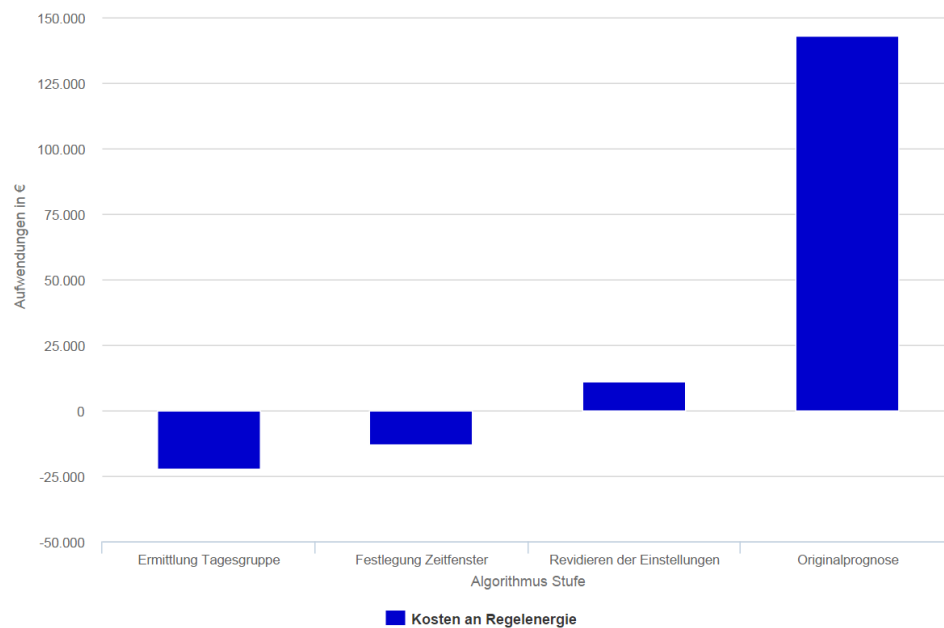


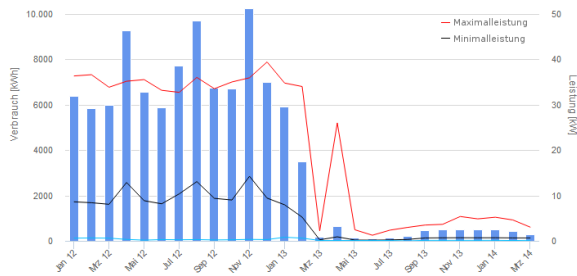
Abbildung 6.2.: Finanzielle Bewertung der Fehlmengen mit Spotmarktpreisen

Da dies sicherlich nicht als alleinige Ursache für die große Abweichung herhalten kann, werden im Folgenden drei Zeitreihen exemplarisch aufgegriffen, deren Prognose aus unterschiedlichen Gründen schlechte Ergebnisse geliefert haben.

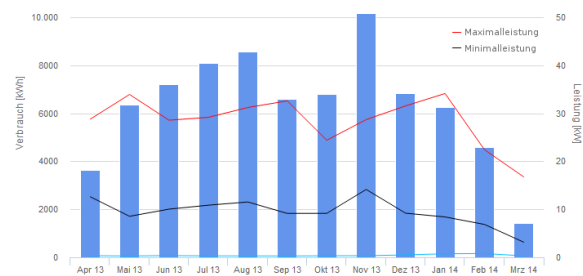
6.3.1. Zeitreihen mit Trendanteilen

Sehr hohe Abweichungen bei der Prognose wurden bei Zeitreihen gefunden, die einen Trend aufweisen oder dauerhaft ihr Verbrauchsverhalten ändern. In Abbildung 6.3a sind die monatlichen Verbräuche einer solchen Zeitreihe dargestellt. Man kann erkennen dass der Verbrauch zwischen November 2012 und März 2013 stark abfällt und danach auf niedrigem Niveau stagniert. Wählt man einen Prognosezeitraum ab April 2013, wird der Abwärtstrend im vorgestellten Algorithmus nicht fortgesetzt. Die Prognose orientiert sich am Vorjahresverbrauch, der Abwärtstrend wird im Folgejahr wiederholt (siehe Abbildung 6.3b).

Um solche Trends innerhalb von Zeitreihen zu erkennen, stehen aus der Statistik verschiedene Verfahren zur Verfügung, um die der Algorithmus ergänzt werden könnte. Dennoch ist es fraglich, ob ein Trend fortgesetzt werden sollte. Ob ein solcher vorliegt, ist erst eindeutig erkennbar, wenn der weitere Verlauf der Verbrauchskurve bekannt ist. Ebenfalls kann das Fortsetzen eines Aufwärtstrends dazu führen, dass dauerhaft zu viel Energie beschafft wird.



(a) Istzeitreihe



(b) Prognosezeitreihe

Abbildung 6.3.: Zeitreihe mit Trendanteil und ihre Prognose

Ist die Änderung im Verbrauchsverhalten von Dauer, wird sich die Prognose mit der Zeit anpassen, da bei der Prognose immer mehr Archivtage mit dem neuen Verbrauchsverhalten einfließen.

6.3.2. Zeitreihen mit abweichendem Verbrauchsprofil zu unterschiedlichen Tageszeiten

Auffällig und durch hohe Prognosefehler gekennzeichnet waren Zeitreihen, die zu verschiedenen Tageszeiten unterschiedliche Verbrauchsmuster zeigen. Als Beispiel sei hier eine Schule aufgeführt. Sie hat in der Regel mehrere verschiedene Verbrauchsmuster: den regulären Unterrichtsbetrieb, Ferienbetrieb und eine abendliche Nutzung der Räumlichkeiten durch Vereine.

Aus diesem Grunde ist die Bestimmung eines einheitlichen Prognoseparameters für einen Tag von Nachteil. Abbildung 6.4 veranschaulicht die Problematik anhand eines Beispiels. Hier ist die errechnete Prognosegüte eines Prognosetages abhängig von Uhrzeit und Größe des gewählten Zeitfensters für die Auswahl von Archivtagen abgebildet. Die Prognosegüte selbst ist als Farbe von grün (geringe Abweichung) bis rot (hohe Abweichung) eingetragen.

Die gewählten Prognoseparameter ergeben zu verschiedenen Tageszeiten eine unterschiedliche Prognosequalität. In den Abendstunden ist das Verbrauchsverhalten wie erwartet abweichend zum regulären Schulbetrieb am Tage. Offensichtlich sind kurze Zeitfenster tagsüber in diesem Beispiel von Vorteil, während sie abends zu einem deutlich höheren Prognosefehler führen.

Demnach sollte die Prognose sich verbessern lassen, wenn die Analyse für jede Viertelstunde des Tages separat ausgeführt wird und der Algorithmus 96 Prognoseparameter pro Prognosetag erzeugt.

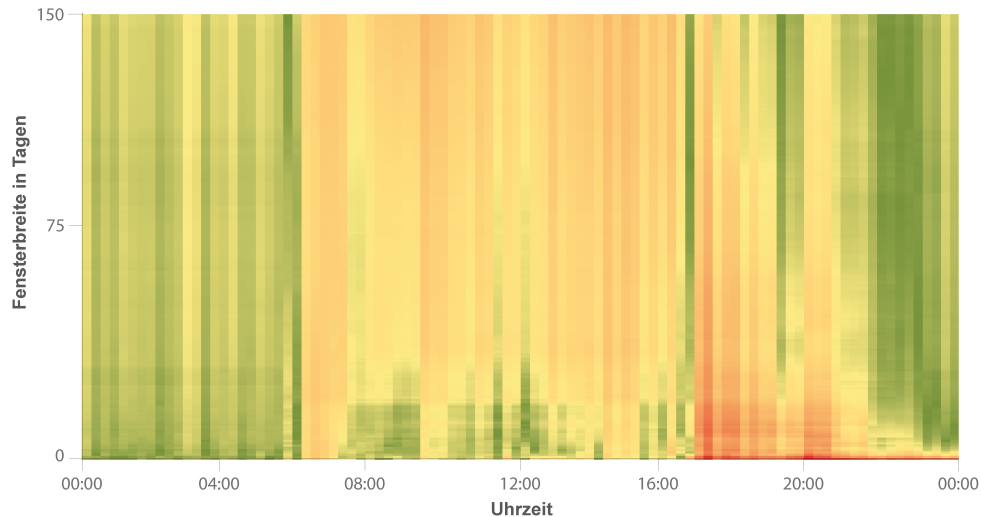


Abbildung 6.4.: Heatmap der Prognosequalität nach Uhrzeit und Zeitfenstergröße einer Schule für einen Prognosetag

6.3.3. Zeitreihen mit zufälligem Verbrauch

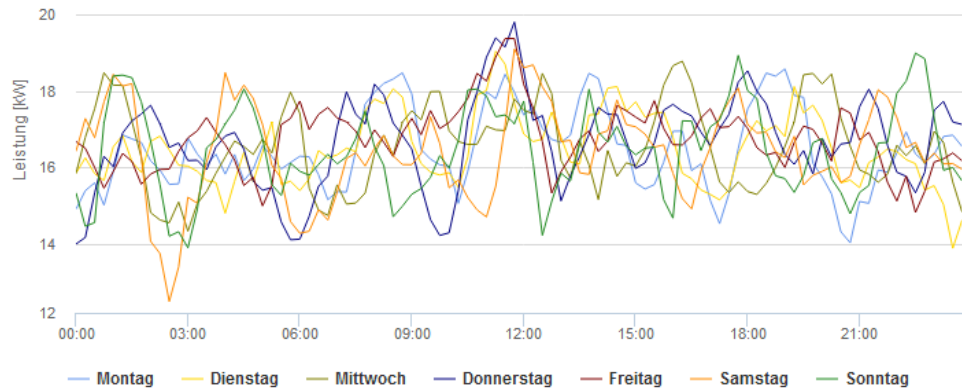
Am schlechtesten in der Auswertung schnitten jene Zeitreihen ab, deren Verbrauch keinem mit den vorgestellten Methoden erkennbaren Muster folgt. Ein Beispiel für eine solche Zeitreihe ist in Abbildung 6.5a gegeben. Der durchschnittliche Verbrauch pro Wochentag besitzt keine erkennbaren Regelmäßigkeiten. Die Zeitreihen erscheinen völlig zufällig.

Derlei Verbraucher stellen ein großes Risiko im Portfolio eines Energielieferanten dar und sollten identifiziert werden können. Ein KorrelationsDiagramm gibt hier Aufschluss (siehe Abbildung 6.5b). Der Korrelationskoeffizient bleibt dauerhaft zwischen $-0,2$ und $0,2$ und ist damit, wie bereits vermutet, ein Hinweis auf schlechte Prognostizierbarkeit (siehe auch Kapitel 3.2.3). Dies sollte im weiteren Verlauf der Arbeit noch quantifiziert werden, und kann somit zur Riskobewertung einer Zeitreihe herangezogen werden.

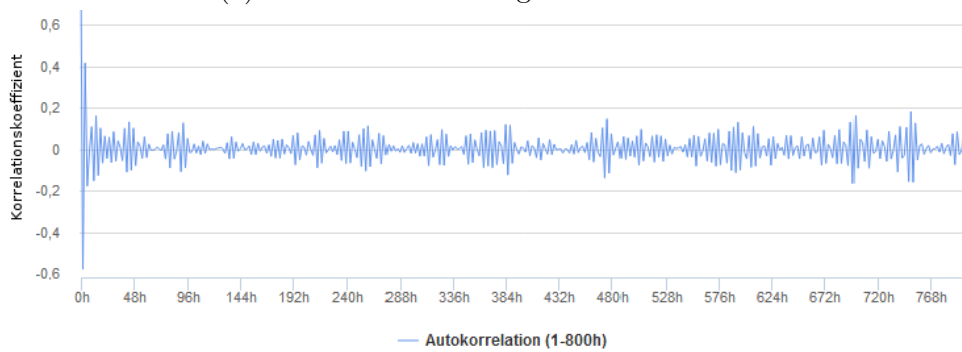
6.4. Zusammenfassung

Im vorangegangenen Kapitel wurden die bei der Auswertung des Algorithmus entstandenen Ergebnisse vorgestellt und diskutiert. Durch den entwickelten Algorithmus konnte die Genauigkeit der Prognose gesteigert werden. Auch die zu erwartenden Kosten für Regenergie blieben im hier vorgestellten Fall unterhalb der aktuellen Kosten.

Auffallend waren jedoch die sehr hohe Abweichungen an mehr als 10% der Prognosetage. Bei der Suche nach Ursachen konnten mehrere Störfaktoren beobachtet werden: Trends in Zeitreihen, unterschiedliche Verbrauchsmuster innerhalb eines Tages und Zeitreihen mit



(a) durchschnittliche tägliche Lastkurve



(b) Autokorrelationsdiagramm

Abbildung 6.5.: Zeitreihe ohne erkennbare Zyklen und tägliche Verbrauchsmuster

zufälligem Verbrauch. Diese wurden anhand von drei Beispielen vorgestellt und mögliche Verbesserungen des Algorithmus aufgezeigt.

7. Überarbeitung Algorithmus

Im vorhergehende Kapitel wurde über Zeitreihen gesprochen, die abhängig von der Tageszeit unterschiedlichen Verbrauchsmustern folgen (siehe Kapitel 6.3.2). Eine Möglichkeit den Algorithmus zu verbessern besteht darin, einen Satz Prognoseparameter pro Prognosestag und Viertelstunde zu erzeugen. Der Prognosefehler soll so weiter reduziert werden. Das In Kapitel 4.1 vorgestellte Modell wird nun dahingehend angepasst und die daraus resultierende Prognosegüte evaluiert.

7.1. Anpassung Modell

Zur Berechnung von viertelstündlichen Parametern muss im vorgestellten Modell die Optimierung der Tagesgruppierung und des Zeitfensters angepasst werden. Demnach muss je eine Tagesgruppierung $G_{t\sim}^* \in \mathbb{G}$ pro Viertelstunde t gefunden werden, die analog zu Formel (4.16) - (4.17)

$$\frac{1}{|\mathcal{A}_{\dot{d}}(G(\dot{d})_{\sim})|} \cdot \sum_{x \in \mathcal{A}_{\dot{d}}(G(\dot{d})_{\sim})} \|x_t - \bar{x}_t\| \quad (7.1)$$

mit

$$\bar{x}_t := \frac{1}{|\mathcal{A}_{\dot{d}}(G(\dot{d})_{\sim})|} \cdot \sum_{x \in \mathcal{A}_{\dot{d}}(G(\dot{d})_{\sim})} x_t \quad (7.2)$$

minimiert.

Dementsprechend gilt es zur Berechnung der Zeitfensterbreite bei optimaler Tagesgruppe $G_{t\sim}^*$, ein z_t^* zu finden, welches

$$\frac{1}{|\mathcal{A}_{\dot{d}}(G_{t\sim}^*(\dot{d}) \cap \overline{B}_z(J(\dot{d})))|} \cdot \sum_{x \in \mathcal{A}_{\dot{d}}(G_{t\sim}^*(\dot{d}) \cap \overline{B}_z(J(\dot{d})))} \|x_t - \bar{x}_t\| \quad (7.3)$$

mit

$$\bar{x}_t := \frac{1}{|\mathcal{A}_{\dot{d}}(G_{t\sim}^*(\dot{d}) \cap \overline{B}_z(J(\dot{d})))|} \cdot \sum_{x \in \mathcal{A}_{\dot{d}}(G_{t\sim}^*(\dot{d}) \cap \overline{B}_z(J(\dot{d})))} x_t \quad (7.4)$$

minimiert. Damit wird die Zeitreihe eines Prognosetages \mathring{d} berechnet als $x = \{\bar{x}_1, \dots, \bar{x}_{96}\}^T$ aus (7.4) mit $G_{t\sim}^*$ und z_t^* .

Die Prognosefunktion ergibt damit anstelle eines Tageslastgangs nur noch einzelne Viertelstundenwerte \hat{x}_t (siehe auch (4.24)).

$$\hat{x}_t = P_t(\mathring{d}, \mathcal{A}(\mathring{d}), G_t, u) := \frac{1}{|\mathcal{A}_{\mathring{d}}(G_{t\sim}(\mathring{d}) \cap \overline{B}_u(J(\mathring{d})))|} \cdot \sum_{x \in \mathcal{A}_{\mathring{d}}(G_{t\sim}(\mathring{d}) \cap \overline{B}_u(J(\mathring{d})))} x_t \quad \text{mit } u = z_t \quad (7.5)$$

7.2. Auswertung und Diskussion

Die Auswertung der Ergebnisse erfolgt analog zu Kapitel 5.2.3 in mehreren Stufen. Die erste Stufe beinhaltet das Ergebnis nach der Festlegung der Tagesgruppierungen für jedes Intervall. Die zweite Auswertung erfolgt nach der Berechnung der Zeitfenster. Um eine Vergleichbarkeit der Ergebnisse zu erhalten, erfolgt die Auswertung derselben Zeitreihen über den bereits zuvor ausgewerteten Zeitraum.

Die Auswertung des abgeänderten Algorithmus ist ebenfalls als kumuliertes Diagramm dargestellt (siehe Abbildung 7.1). Als Vergleichskurve ist die letzte, beste Algorithmusstufe ohne viertelstündliche Parameter eingetragen. Nach der Bestimmung der Tagesgruppierung bleibt die Prognosegüte des abgeänderten Algorithmus noch hinter der Vergleichskurve zurück. Nach Festlegung eines Zeitfensters liegt die Kurve deutlich oberhalb.

Obwohl die Kurven sich im unteren Fehlerbereich noch schneiden, ergibt die Umstellung auf 96 Prognoseparameter pro Tag eine deutliche Steigerung in der Prognosequalität. Jedoch zeigt die finanzielle Bewertung Kosten von ca. 35.000 € und 60.000 € für Regelenergie (siehe Abbildung 7.2).

Nach den in Kapitel 4.1.1 Kriterien ergeben die hier vorgestellten Änderungen die höchste Prognosequalität. Der abgeänderte Algorithmus ist zu bevorzugen. Dennoch spricht die finanzielle Bewertung dagegen. Eine mögliche Erklärung ist sicherlich in der Art der Auswertung zu finden: während die Prognosegüte die relative Abweichung vom Ist-Wert betrachtet, rechnet die finanzielle Bewertung mit den absoluten Werten.

Ungeachtet der Tatsache, dass eine finanzielle Optimierung nach Regelenergiepreisen nicht zulässig ist, sollte hier noch einmal nachforschen, ob nicht einzelne Zeitreihen das Ergebnis stark negativ beeinflussen. Diese könnten dann bei der Risikobewertung in einen entsprechend schlechteren Tarif eingestuft werden.

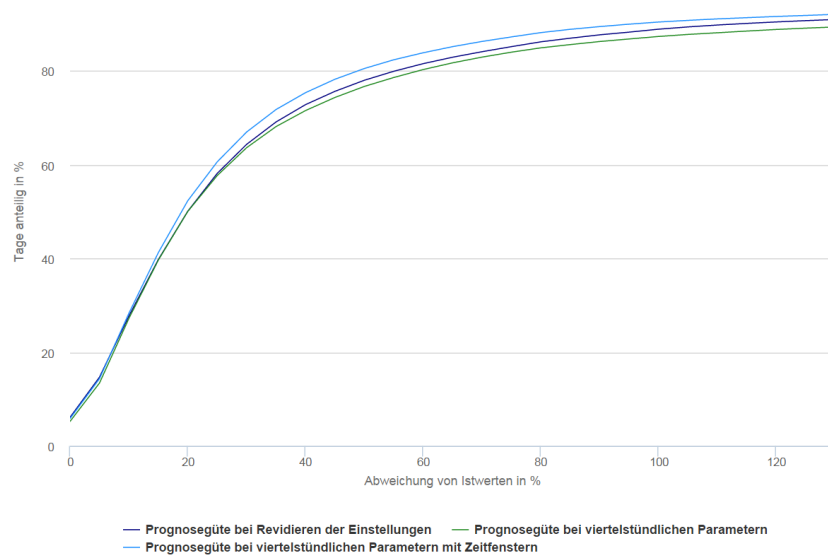


Abbildung 7.1.: Vergleich relativer Prognosegüte zwischen bester Algorithmusstufe und dem Algorithmus mit viertelstündlichen Prognoseparametern

7.3. Zusammenfassung

In diesem Kapitel wurde der in Kapitel 4.2 vorgestellte Algorithmus um die Erstellung von viertelstündlichen Prognoseparametern erweitert. Hierzu wurde das Optimierungsproblem aus Kapitel 4.1 erweitert. Die Auswertung ergab eine Steigerung der Prognosequalität und der Kosten für Fehlmengen.

Nach den Vorgaben der Bundesnetzagentur müsste dieses Verfahren trotz der schlechteren finanziellen Ergebnisse bevorzugt werden. Warum die Ergebnisse voneinander abweichen sollte Gegenstand weiterer Forschung sein.

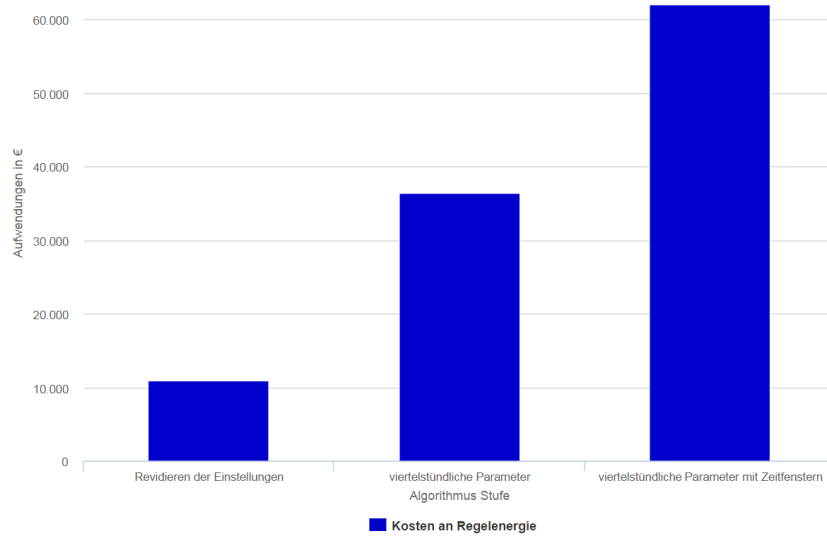


Abbildung 7.2.: Finanzielle Bewertung der Fehlmengen mit Spotmarktpreisen nach Berechnung von viertelstündlichen Prognoseparametern

8. Prognostizierbarkeit von Zeitreihen

Teilziel dieser Arbeit ist eine quantitative Aussage über die Vorhersagbarkeit einer Zeitreihe. Zu diesem Zweck wurden mehrere Kennzahlen ermittelt, die im folgenden Kapitel der Prognostizierbarkeit einer Zeitreihe gegenübergestellt werden.

8.1. Idee

In Kapitel 3 wurde bereits über Größen gesprochen, die eine Aussage über die Prognostizierbarkeit von Zeitreihen treffen. Als mögliche Kennzahlen wurden die Streuung und verschiedene Korrelationskoeffizienten diskutiert. Es stellt sich die Frage, wie genau die Größen die Prognosegüte beeinflussen.

Nach Testen des Algorithmus ist es nun möglich, diese Kennzahlen der ermittelten Prognosegüte gegenüberzustellen. Abbildung 8.1 veranschaulicht den Zusammenhang zwischen den Autokorrelationskoeffizienten der untersuchten Zeitreihen und der Prognosegüte in einer Heatmap.

Auf der horizontalen Achse sind die verschiedenen Autokorrelationskoeffizienten aufgetragen, während die vertikale Achse die Prognosegüte anzeigt. Die Farbskala gibt den prozentualen Anteil der Prognosetage mit gleichem Koeffizienten an, die eine bestimmte Prognosegüte besitzen. Werden beispielsweise alle Zeitreihen betrachtet, für die ein Koeffizient von 0,7 berechnet wurde, so haben etwa 25% aller Prognosetage einen Prognosefehler zwischen 10 und 15%.

Wie erwartet kann man deutlich einen positiven Zusammenhang zwischen Koeffizienten und Prognosegüte erkennen. Je höher der Koeffizient, desto stärker konzentrieren sich die Prognosefehler im unteren Prozentbereich. Jedoch ist auffällig, dass der Prognosefehler im oberen und unteren Bereich der Koeffizienten nicht dieser Regel folgt. Die schlechten Ergebnisse am rechten Rand lassen sich durch die geringe ausgewertete Datenmenge erklären (siehe Abbildung 5.3b). In den niedrigen Wertebereichen der Koeffizienten befinden sich hauptsächlich Zeitreihen, die nur gelegentlich und dann zufällig Leistung abnehmen,

während der Verbrauch an den meisten Tagen null ist. Dies ergibt einen geringen Prognosefehler bei schlechtem Autokorrelationswert.

Kann der Einfluss dieser und weiterer Größen auf die Prognosequalität quantifiziert werden, ist eine Vorhersage derselben aus einer unbekannten Zeitreihe möglich. So ist ein Maß für das Risiko gefunden, das einer Zeitreihe innewohnt.

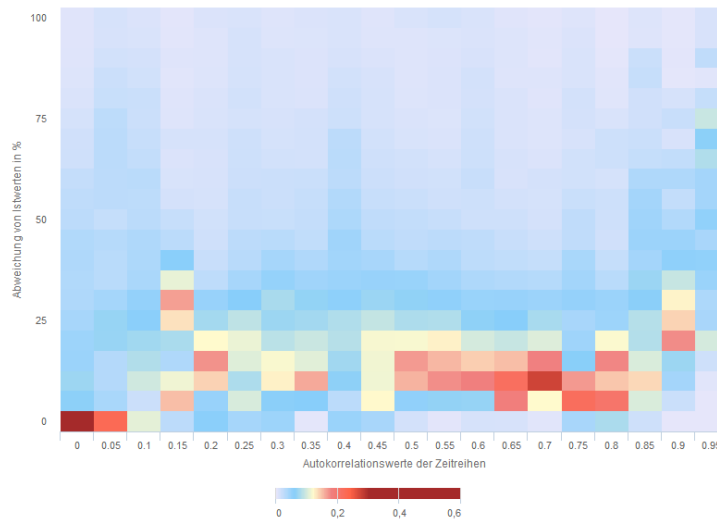


Abbildung 8.1.: Grafik von Prognosegüte und Betrag der Autokorrelationswerte bei einem Versatz von einer Woche

8.2. Herangehensweise

Zusammenhänge zwischen einer Zeitreihe und mehreren Variablen können mittels mehrdimensionaler Regressionsanalyse geschätzt werden. Dadurch kann ein Modell aufgestellt werden, welches die konkrete Beziehung zwischen Einflussgrößen und Zeitreihe beschreibt. Im Modell wird der Prognosefehler auch als abhängige und eine Einflussgröße als unabhängige Variable bezeichnet.

Der vorgestellte Algorithmus ging bisher davon aus, dass die Streuung innerhalb des Datenarchivs und die Breite des Zeitfensters den Prognosefehler beeinflussen. Weiterhin wurde angenommen, dass ein großes Datenarchiv grundsätzlich bessere Ergebnisse liefert. Diese Annahmen sollen hier überprüft werden. Außerhalb des Algorithmus wurden weitere unabhängige Variablen identifiziert, die einen Einfluss auf die Prognostizierbarkeit einer Zeitreihe haben: den Korrelationskoeffizienten mit Schulferien und Autokorrelationskoeffizienten (siehe Kapitel 3.2.3 und 3.2.2).

	Min	25% Q.	Median	Mittelwert	75% Q.	Max
Fehler	0	12,92	23,99	169,10	48,43	48060
Streuung	0	28,10	43,68	50,54	62,82	382
Fensterbreite	1	20,00	37,00	36,58	54,00	70
Archivtiefe	1	1	2	2	3	3
Fensterb.* Streuung	0	590	1436	1920	2641	26220
Autokorr. 1 Tag	0	0,25	0,38	0,39	0,52	0,98
Autokorr. 1 Woche	0	0,28	0,46	0,44	0,60	0,97
Korr. Ferien	0	0,23	0,38	0,39	0,52	0,98

Tabelle 8.1.: Datenbeschreibung: Quantile der unabhängigen Variablen

Zusätzlich bietet sich hier die Möglichkeit, die Annahme zu verifizieren, dass ein gutes Zeitfenster durch eine geringe Streuung innerhalb der Archivtage gekennzeichnet ist. Dazu wird an dieser Stelle eine Regression gegen das Produkt von Streuung und Fensterbreite berechnet.

Zur Schätzung der Koeffizienten der unabhängigen Variablen muss zunächst eine geeignete Datengrundlage erstellt werden. Ausgehend von der besten ermittelten Tagesgruppierung und einem Zeitfenster werden Prognosegüte und Streuung für jeden Prognosetag innerhalb des verarbeiteten Datenarchivs ermittelt. Hierzu wird schrittweise das Zeitfenster verbreitert und die Berechnung wiederholt.

Die Korrelations- und Autokorrelationskoeffizienten je Zeitreihe können unabhängig davon bestimmt werden. Da hier nur die Intensität der Abhängigkeit, jedoch nicht die Gleich- und Gegenläufigkeit der Größen von Interesse sind, wird ihr Betrag zur Regressionsanalyse herangezogen.

Tabelle 8.1 gibt einen Überblick über abhängige und unabhängige Variablen und deren Werte. Im folgenden wird auf lineare und quadratische Abhängigkeit getestet. Damit ergibt sich folgende Regressionsfunktion:

$$\begin{aligned}
 \text{Fehler}(x_{\text{Streu}}, x_{\text{Fens}}, x_{\text{Arch}}, &= a_0 + a_1 x_{\text{Streu}} + a_2 x_{\text{Fens}} + a_3 x_{\text{Fens}}^2 \\
 K_{\text{Tag}}, K_{\text{Woche}}, K_{\text{Ferien}}) &+ a_4 x_{\text{Arch}} + a_5 x_{\text{Arch}}^2 + a_6 x_{\text{Streu}} x_{\text{Fens}} + a_7 x_{\text{Streu}} x_{\text{Fens}}^2 \\
 &+ a_8 K_{\text{Tag}} + a_9 K_{\text{Tag}}^2 + a_{10} K_{\text{Woche}} + a_{11} K_{\text{Woche}}^2 \\
 &+ a_{12} K_{\text{Ferien}} + a_{13} K_{\text{Ferien}}^2
 \end{aligned} \tag{8.1}$$

8.3. Umsetzung

Zur Berechnung der Koeffizienten wurden die ermittelten Daten in das Statistik-Programm R geladen. R ist eine open source Software zur Lösung von statistischen Problem und erstellen von Grafiken. Es wurde R in der Version 3.1.1 für Windows eingesetzt.

Anhand der Prognosefehler in Tabelle 8.1 ist erkennbar, dass es signifikante Ausreißer nach oben in den Werten gibt: Mittelwert und Median klaffen weit auseinander, der Mittelwert liegt oberhalb des 75%-Quantil. Die Regressionsanalyse wurde deshalb mit einem robusten Schätzer durchgeführt (siehe Koller und Stahel 2011).

Die Analyse und anschließende Regression wird über ein Teil des Datenarchivs ausgeführt, da die Ergebnisse zu einer weiteren Prognose genutzt werden sollen (siehe Kapitel 8.6). Daher werden die Archivdaten aus dem Jahr 2014 nicht verarbeitet.

8.4. Auswertung

Die Ergebnisse der Regressionsanalyse sind in Tabelle 8.2 festgehalten. In den ersten beiden Spalten befinden sich die geschätzten Koeffizienten und deren Standardfehler, die für Formel (8.1) ermittelt wurden.

Die dritte Spalte enthält den Quotienten aus Koeffizient und Standardfehler, t-Wert genannt. Er zeigt an, wie groß der Einfluss des Koeffizienten auf das Gesamtergebnis ist. Ein t-Wert nahe 0 bedeutet, dass der Einflussfaktor zu vernachlässigen ist.

Der Wert der letzten Spalte gibt die Wahrscheinlichkeit an, dass der Wert des tatsächlichen Koeffizienten außerhalb des angegebenen Intervalls liegt. Umso niedriger dieser Wert ist, desto vertrauenswürdiger ist der Koeffizient. Man spricht hier von statistischer Signifikanz (siehe auch Auer 2011).

Zunächst werden die im Algorithmus verwendeten Größen ausgewertet. Anschließend finden die Korrelationskoeffizienten Beachtung, die Auskunft über die Prognostizierbarkeit einer Zeitreihe gehen sollen.

Coefficients:	Estimate	Standard Error	t ratio	Pr(> t)
(y-Achsenabschnitt)	1,586 E+01***	8,846E-02	179,322	<2E-16
Fensterbreite	-1,207 E-01***	4,889E-03	-24,698	<2E-16
Fensterbreite ²	1,267 E-03***	6,676E-05	18,977	<2E-16
Archivtiefe	7,814 E-03***	1,430E-04	54,647	<2E-16
Archivtiefe ²	-1,128 E-05***	1,786E-07	-63,196	<2E-16
Streuung	2,052 E-01***	1,721E-03	119,278	<2E-16
Streuung ²	-3,377 E-04***	5,800E-06	-58,214	<2E-16
Fensterbreite * Streuung	-2,721 E-04**	9,427E-05	-2,886	0.0039
Fensterbreite ² * Streuung	-8,744 E-07	1,250E-06	-0,700	0.4841
AutokorrWoche	3,001 E+01***	2,627E-01	114,208	<2E-16
AutokorrWoche ²	-3,845 E+01***	3,018E-01	-127,390	<2E-16
AutokorrTag	-1,502 E+01***	2,866E-01	-52,420	<2E-16
AutokorrTag ²	1,637 E+01***	3,200E-01	51,177	<2E-16
KorrFerien	-9,801 E+00***	9,845E-02	-99,551	<2E-16
KorrFerien ²	7,983 E+00***	3,082E-01	25,907	<2E-16

Signif. codes: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Robust residual standard error: 18,52

Multiple R-squared: 0,07139, Adjusted R-squared: 0,07138

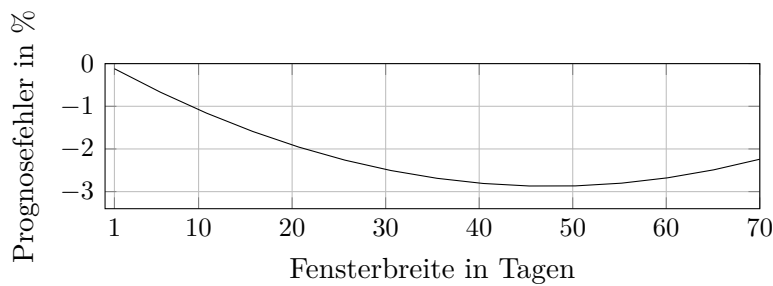
Convergence in 11 IRWLS iterations

Tabelle 8.2.: Ergebnisse der Regressionsanalyse

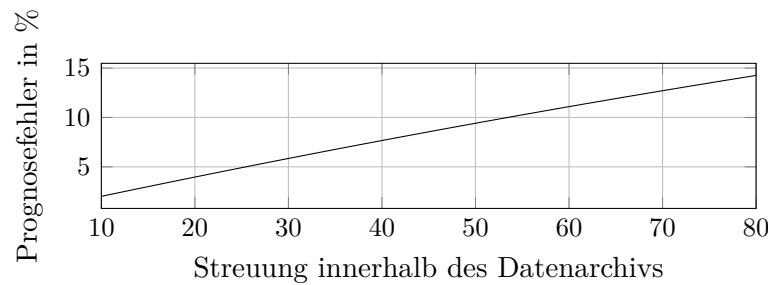
8.4.1. Parameter des Prognosealgorithmus

In den Daten finden sich Hinweise auf einen Zusammenhang zwischen Fensterbreite, Archivtiefe, Streuung und dem Prognosefehler. Allein der Einfluss der Fensterbreite² und der Streuung in Kombination besitzt einen Signifikanzwert von weit über 5%. Der Koeffizient ist deshalb als zufällig einzustufen. Folglich liegt kein Zusammenhang mit der Prognosegüte vor.

In Abbildung 8.2 wird der Zusammenhang der unabhängigen Variablen mit dem Prognosefehler je Größe dargestellt. Dazu wurde aus den Koeffizienten jeder unabhängigen Variablen eine Funktion generiert. Ihr Graph ist innerhalb eines relevanten Intervalls abgebildet. Beispielsweise ergeben die Koeffizienten der Fensterbreite das Polynom $f(x) = 0.001267 * x^2 - 0.1207 * x$, während die Fensterbreite laut Tabelle 8.1 ausschließlich auf dem Intervall $[1, \dots, 70]$ definiert ist.



(a) Prognosefehler nach Fensterbreite



(b) Prognosefehler nach Streuung

Abbildung 8.2.: Darstellung der Regressionskurven, abgeleitet aus Tabelle 8.2

umfassen kleine Zeitfenster wahrscheinlich zu wenige Archivtage, so dass statistische Ausreißer überbewertet werden. Sehr große Zeitfenster führen dazu, dass Wintertage mit Verbräuchen aus dem Sommer prognostiziert werden.

Der Zusammenhang zwischen Fensterbreite und Prognosefehler ist quadratisch mit einem Minimum bei etwa 48 Tagen (siehe Abbildung 8.2a). Wie erwartet umfassen kleine Zeitfenster wahrscheinlich zu wenige Archivtage, so dass statistische Ausreißer überbewertet werden. Sehr große Zeitfenster führen dazu, dass Wintertage mit Verbräuchen aus dem Sommer prognostiziert werden (siehe Kapitel 4.1.2).

Bei Streuung und Archivtiefe wurde ebenfalls auf quadratischen Zusammenhang getestet. Sein Anteil ist jedoch marginal, so dass die Kurve stark einer steigenden Geraden ähnelt. Geringe Streuungswerte und Archivtiefen ergeben folglich eine gute Prognosequalität (siehe Abbildung 8.2b und 8.2c).

Der Einfluss der Streuung auf das Zeitfenster wurde durch das Produkt aus beiden Variablen gemessen. Es ist ersichtlich, dass hohe Werte des Produktes einen geringen Prognosefehler zur Folge haben (siehe Abbildung 8.2d).

In Abbildung 8.3 sind die oben genannten unabhängigen Variablen noch einmal als 3D-Diagramm dargestellt. Deutlich erkennbar ist, dass die Streuung innerhalb des Daten-

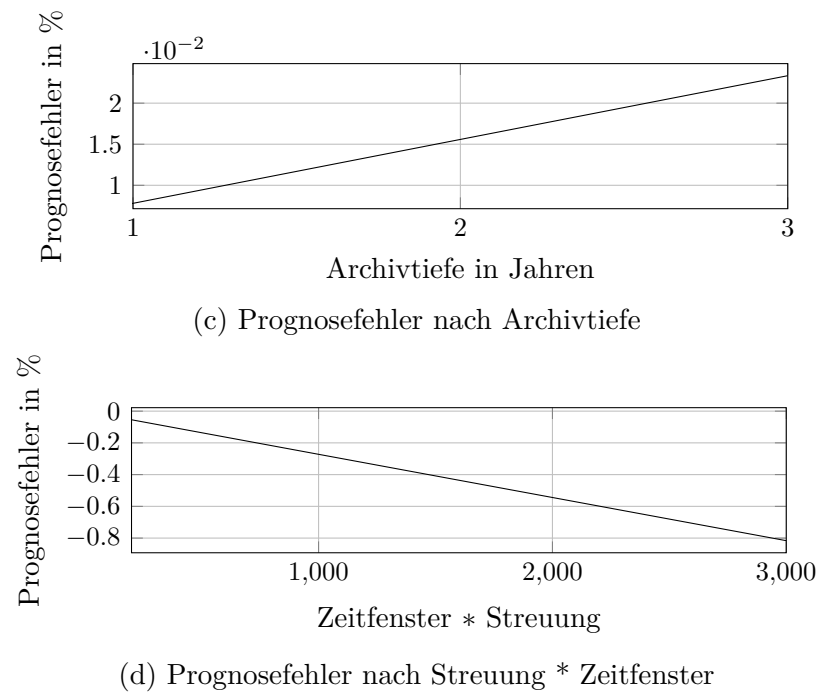


Abbildung 8.2.: Darstellung der Regressionskurven, abgeleitet aus Tabelle 8.2

archivs die für den Prognosefehler ausschlaggebende Größe ist. Die Zeitfensterbreite und insbesondere die Archivtiefe sind nachrangig.¹

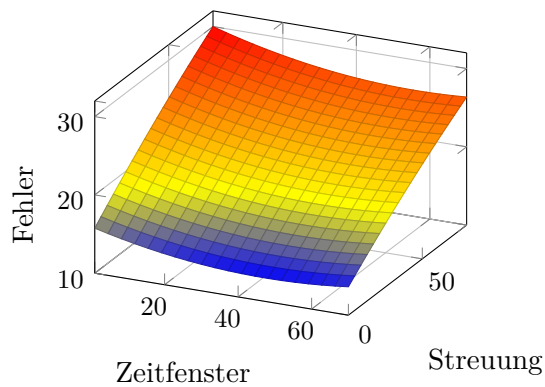


Abbildung 8.3.: Darstellung des Zusammenhangs von Zeitfensterbreite x , Streuung y und Archivtiefe $z = 1$ mit dem Prognosefehler

¹3DDiagramme mit den Archivtiefen $z \in \{1, 2\}$ sind in Anhang A.1 abgebildet

8.4.2. Einfluss der Korrelationskoeffizienten

Um den Einfluss von externen Faktoren zu verifizieren, wurde die Abhängigkeit der errechneten Prognosegüte von verschiedenen Korrelationskoeffizienten getestet (siehe auch Kapitel 3). Die Signifikanzniveaus der Ergebnisse in Tabelle 8.2 sind für die jeweiligen getesteten unabhängigen Variablen hinreichend, so dass ein Zusammenhang angenommen werden kann.

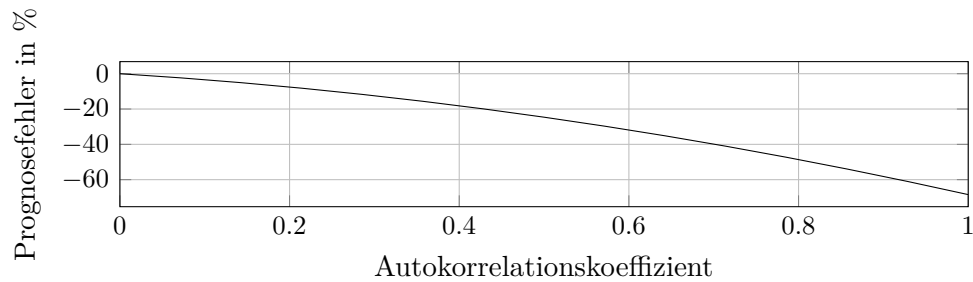
Zur besseren Anschaulichkeit wurden die errechneten Koeffizienten aus Tabelle 8.2 ebenfalls für jede Variable grafisch dargestellt (siehe Abbildung 8.4). Die Vermutung, dass die Abhängigkeit der Autokorrelationskoeffizienten nicht linear ist, wird in Abbildung 8.4a und Abbildung 8.4b bestätigt.

Die Grafik zeigt, dass hohe Autokorrelationkoeffizienten bei wochenzyklischen Zeitreihen von Vorteil sind, während bei Tageszyklen Korrelationskoeffizienten zwischen 0,4 und 0,5 von Vorteil sind. Eine starke Korrelation mit einer Ferienzeitreihe ergibt ebenfalls bessere Prognoseergebnisse. Auch Unkorreliertheit ist von Vorteil, denn die Kurve besitzt die Form eines umkehrten U (siehe Abbildung 8.4c). Im Vergleich der verschiedenen Korrelationskoeffizienten wird deutlich, dass der Wochenzyklus den größten Einfluss auf die Prognosegüte hat.

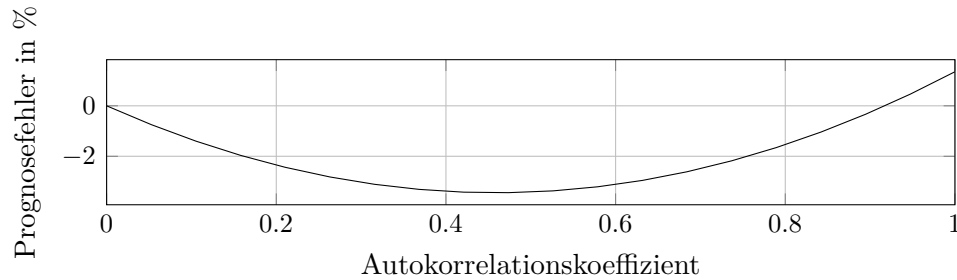
8.5. Diskussion

Rückblickend auf die Annahmen, die im Verlauf der Arbeit gemacht wurden, haben die Ergebnisse der Regressionsanalyse viele bestätigt und einige widerlegt. Die These, dass die Prognosegüte stark von der Streuung innerhalb des gewählten Datenarchiv abhängt (siehe Kapitel 3.2.1), konnte durch die Analyse noch einmal verifiziert werden. Auch die Vermutung, dass ein zu kleines Zeitfenster das Archiv übermäßig einschränkt, während ein zu großes die Saisonalität von Zeitreihen nicht berücksichtigt (siehe Kapitel 4.1.2), konnte durch die Form der Regressionskurve untermauert werden.

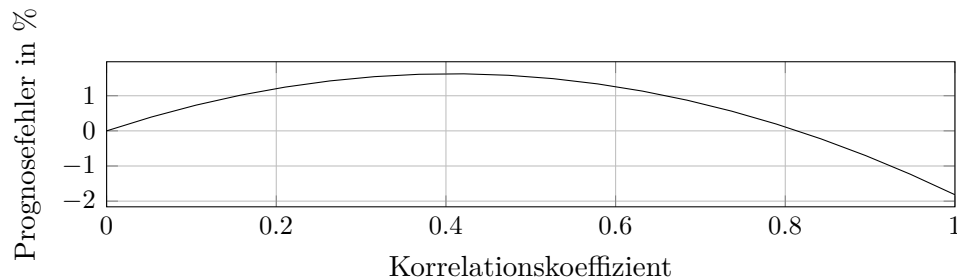
Überraschend war das Resultat im Bezug auf den Einfluss der Streuung auf das Zeitfenster. Erwartet wurde, dass ein gutes Zeitfenster durch eine niedrige Streuung gekennzeichnet ist (siehe Kapitel 4.1.2). Das Ergebnis der Regressionsanalyse deutet auf das Gegenteil hin. Trotzdem wurden bei dem Einsatz der entsprechenden Algorithmusstufe die Ergebnisse besser (siehe Kapitel 6.1). Dies ist wahrscheinlich darauf zurückzuführen, dass zuvor kein Zeitfenster eingesetzt wurde, was in einer noch schlechteren Prognosegüte resultiert. Auch ist der Einfluss der Variable auf das Ergebnis sehr gering (siehe Tabelle 8.2).



(a) Autokorrelationskoeffizient bei einer Woche Zeitversatz



(b) Autokorrelationskoeffizient bei einem Tag Zeitversatz



(c) Korrelationskoeffizient mit der Ferienzeitreihe

Abbildung 8.4.: Darstellung der Regressionskurven der Korrelationskoeffizienten, abgeleitet aus Tabelle 8.2

Wider Erwarten ist auch eine größere Archivtiefe, welche die Anzahl der Archivtage angibt, für die Prognosequalität von Nachteil. Die Ergebnisse sprechen dafür, dass ein besonders großes Datenarchiv den Prognosefehler erhöht. Die Vermutung liegt nahe, dass die Verbräuche sich über mehrere Jahre verändern und somit nicht-repräsentative Tage in die Prognose mit einfließen. Die Einführung einer weiteren Größe, die eine Aussage über die Ähnlichkeit der Jahresverläufe trifft, wäre möglich. Allerdings ist dieser Einfluss verhältnismäßig gering, sodass geprüft werden muss, ob eine weitere Forschung an dieser Stelle sinnvoll ist.

Anhand der Regressionsanalyse gegen die ausgewählten Korrelationskoeffizienten sieht man sehr deutlich, dass eine starke Wochenzyklizität den größten Einfluss auf die Prognose

hat. Jedoch ist dieses Ergebnis nicht verwunderlich, da der Prognosealgorithmus durch den Einsatz der Tagesgruppen-Schemata stark auf dieser Eigenschaft aufbaut (siehe Kapitel 4.1.2). Auch der Effekt der Korrelationskoeffizienten mit einer Ferienzeitreihe lässt sich auf denselben Umstand zurückführen.

Dies ist jedoch für den Prognosealgorithmus nicht von Nachteil. Es bedeutet lediglich, dass Zeitreihen mit hoher Zyklizität und starker Abhängigkeit von kalendarischen Ereignissen besonders gut mit dem hier vorgestellten Algorithmus prognostizierbar sind. Für Zeitreihen, die über schlechte Werte in diesen Kategorien verfügen, sollte ein anderes Verfahren gewählt werden.

Durch die Regressionanalyse lässt sich auch eine weitere zentrale Anforderung dieser Arbeit klären: die Frage nach einer Kennzahl, die eine Bewertung des Risikos, welches eine Zeitreihe mit sich bringt. Mit Hilfe der Koeffizienten ist es möglich, anhand einer zur Verfügung gestellten Zeitreihe eine Vorhersage des Prognosefehlers für einen Prognosetag zu treffen. Bei gewählter Tagesgruppierung und Zeitfenster ergibt folgende Formel einen Schätzwert für den relativen Prognosefehler:

$$\begin{aligned}
 \text{Fehler}(x_{\text{Streu}}, x_{\text{Fens}}, x_{\text{Arch}}, &= 15,86 + 0,2052x_{\text{Streu}} - 0,1207x_{\text{Fens}} + 0,001267x_{\text{Fens}}^2 \\
 K_{\text{Tag}}, K_{\text{Woche}}, K_{\text{Ferien}}) &+ 0,007814x_{\text{Arch}} - 0,00001128x_{\text{Arch}}^2 \\
 &- 0,0002721x_{\text{Streu}}x_{\text{Fens}} \\
 &- 15,02K_{\text{Tag}} + 16,37K_{\text{Tag}}^2 \\
 &+ 30,01K_{\text{Woche}} - 38,45K_{\text{Woche}}^2 \\
 &- 9,801K_{\text{Ferien}} + 7,983K_{\text{Ferien}}^2
 \end{aligned} \tag{8.2}$$

8.6. Prognose mit neuer Zielfunktion

Wie bereits dargestellt, gibt die durchgeführte Regressionsanalyse Aufschluss, wie die unabhängigen Variablen die Prognostizierbarkeit einer Zeitreihe beeinflussen. Auf Basis dieser Ergebnisse kann nicht nur der Prognosefehler vorausgesehen werden. Deren Koeffizienten können ebenfalls dazu genutzt werden, die Prognoseparameter einzustellen, so dass der Prognosefehler minimiert wird.

Aus der Regressionsanalyse resultiert eine neue Zielfunktion, die zur Ermittlung der Prognoseparameter herangezogen werden kann. Die Variablen K_{Tag} , K_{Woche} , K_{Ferien} entfallen, da sie je Zeitreihe konstant bleiben und das Ergebnis nicht beeinflussen.

$$\begin{aligned}
\text{Fehler}(x_{\text{Streu}}, x_{\text{Fens}}, x_{\text{Arch}}) &= 15,86 + 0,2052x_{\text{Streu}} - 0,1207x_{\text{Fens}} + 0,001267x_{\text{Fens}}^2 \\
&\quad + 0,007814x_{\text{Arch}} - 0,00001128x_{\text{Arch}}^2 \\
&\quad - 0,0002721x_{\text{Streu}}x_{\text{Fens}}
\end{aligned} \tag{8.3}$$

Wie bereits erwähnt, wurde der Analysezeitraum im Vergleich zu der Auswertung in Kapitel 5.2.3 verkürzt, damit sich Analyse und Prognosezeitraum nicht überschneiden. Die Auswertung beruht folglich nur auf einem Teil der Daten.

Zum Test der Zielfunktion wird der Algorithmus aus Kapitel 4.2.1 angepasst. Algorithmus 6 ersetzt dabei die Stufen zwei bis vier des neuen Prognosealgorithmus 1. Zu diesem Zeitpunkt steht bereits eine Tagesgruppierung G^* fest. Darauf basierend wird für alle Kombinationen aus Zeitfenster z und Archivtiefe a die Streuung im Datenarchiv ermittelt und nach der neuen Zielfunktion in Formel (8.3) optimiert.

Algorithmus 6 : Berechnung von Zeitfenstergröße und Archivtiefe mit Zielfunktion nach Regressionsanalyse

Eingabe : \mathring{d} , der Prognosetag

Eingabe : G^* , enthält die ermittelte beste Tagesgruppierung für \mathring{d}

Ausgabe : *bestKonfig*, enthält die optimale Archivtiefe und Fensterbreite für \mathring{d}

```

1 function ermittleParameter( $G^*, \mathring{d}$ ) begin
2   foreach  $z \in [1 \dots z_{\max}]$  do
3     foreach  $a \in [1 \dots a_{\max}]$  do
4        $\text{Archivtage} \leftarrow G^* \cap \overline{B}_z(J(\mathring{d})) \cap \{d \in D \mid \rho(d, \mathring{d}) \leq 365 * a\}$ 
5       if  $\text{Archivtage} \neq \emptyset$  then
6          $s \leftarrow$  berechne Streuung über  $\mathcal{A}(\text{Archivtage})$  nach (5.1)
7          $\text{fehler} \leftarrow \text{Prognosefehler}(s, z, a)$  nach (8.3)
8         if  $\text{fehler} < \text{maxFehler}$  then
9            $\text{bestKonfig} \leftarrow (z, a)$ 
10        end
11      end
12    end
13  end
14  return bestKonfig
15 end

```

Die Ergebnisse der Auswertung in Abbildung 8.5 zeigen keine Steigerung der Prognosequalität im Vergleich zu den vorangegangenen besten Prognosestufen. Die errechneten Erlöse aus dem Verkauf von Strom nach Regelenergiepreisen ist mit ca 21.000 € geringer

als bei der Prognose mit viertelstündlichen Parametern, kann aber die erste Version des neuen Algorithmus nicht unterbieten (siehe Abbildung 8.6).

Möglicherweise sind führen die über die Gesamtheit der Daten errechneten Koeffizienten über Effekte auf die einzelne Zeitreihe, die hier noch nicht berücksichtigt wurden. Die Regressionsanalyse sollte auf Zeitreihenebene wiederholt werden, um je einen Satz aus Koeffizienten zu erzeugen. Auf diese Weise könnte jede Zeitreihe nach einer individuellen Zielfunktion optimiert werden. Weitere Forschung wird zeigen, ob dieser Ansatz ein besseres Ergebnis liefert.

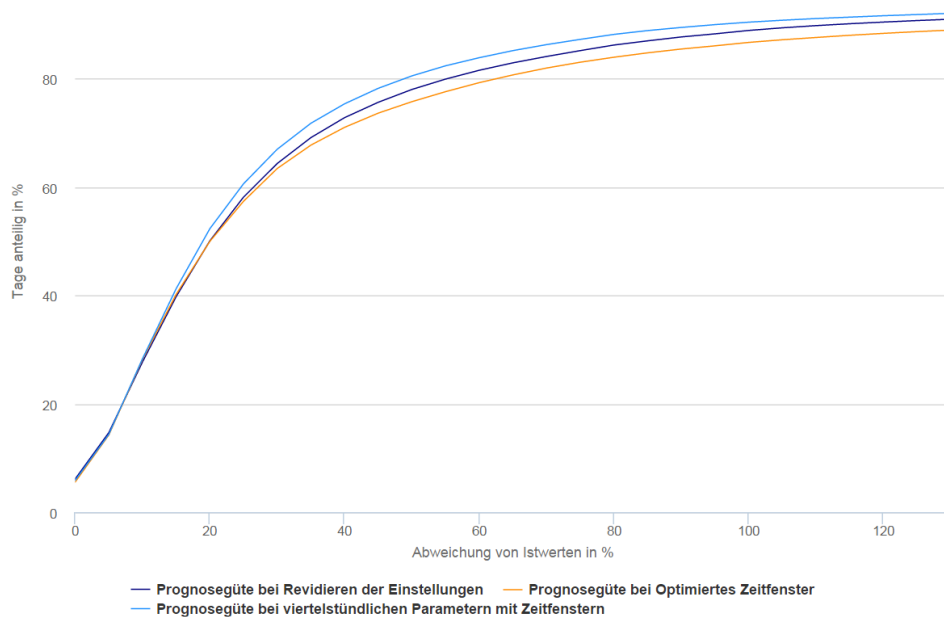


Abbildung 8.5.: Vergleich relativer Prognosegüte zwischen Algorithmus mit viertelstündlichen Prognoseparametern und der Optimierung nach neuer Zielfunktion

8.7. Zusammenfassung

Im vorangegangenen Kapitel wurde der Einfluss verschiedener Kennzahlen auf die Prognosequalität des Algorithmus untersucht. Zu diesem Zweck wurde eine Regressionsanalyse durchgeführt, die Auskunft über lineare und quadratische Abhängigkeiten der Einflussvariablen aufgezeigt hat. Die Ergebnisse wurden ausführlich diskutiert. Getestet wurden folgende Einflussvariablen: Streuung innerhalb des Datenarchivs, Archivtiefe, Fensterbreite, der Korrelationskoeffizient mit einer Ferienzeitreihe und verschiedene Autokorrelationskoeffizienten.

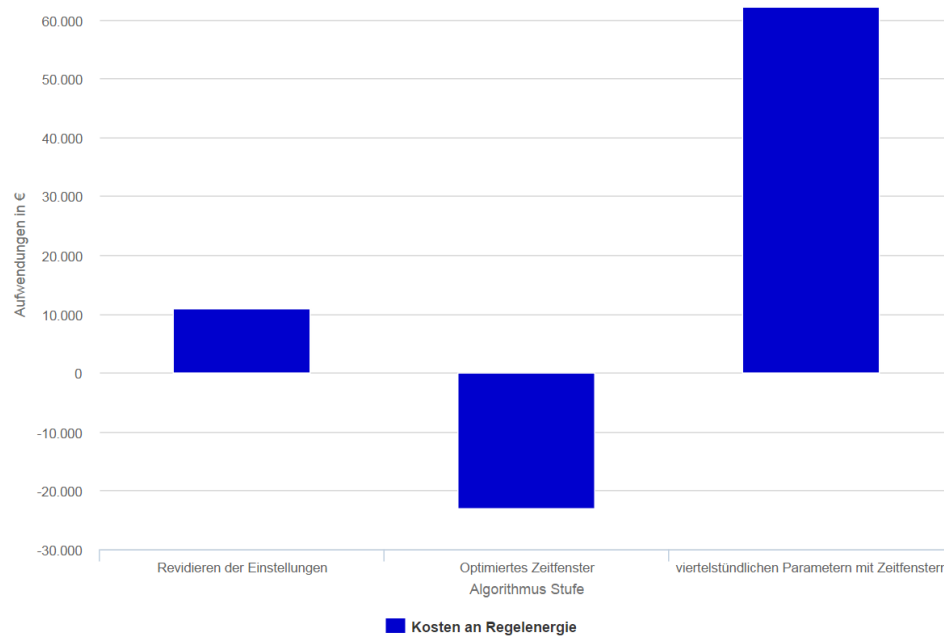


Abbildung 8.6.: Finanzielle Bewertung der Ergebnisse im Vergleich

Es wurden die Abhängigkeiten grafisch aufbereitet und gezeigt, dass die Prognosegüte des neu entwickelten Algorithmus stark von der Streuung im gewählten Datenarchiv und dem Autokorrelationskoeffizienten bei wöchentlicher Verschiebung beeinflusst wird.

Weiterhin wurden die Ergebnisse der Analyse genutzt, um eine Funktion zur Risikobewertung für eine Zeitreihe sowie eine Zielfunktion für einen abgeänderten Prognosealgorithmus aufzustellen. Leider blieben die Ergebnisse hinter den Erwartungen zurück. Eine Mögliche Verbesserung des Verfahrens in Form einer separaten Zielfunktion je Zeitreihe wurde vorgeschlagen.

9. Zusammenfassung und Ausblick

Dieses Kapitel dient als Zusammenfassung der vorliegenden Arbeit. Die Ergebnisse werden noch einmal herausgestellt und ihre Relevanz für Praxis und Forschung eingeordnet. Anschließend werden die Limitationen der Arbeit erörtert und Anreize für zukünftige Forschung geliefert.

9.1. Zusammenfassung

Die vorliegende Arbeit behandelt die Thematik der Prognose von Energiezeitreihen auf Basis historischer Verbrauchsdaten. Sie wurde im Rahmen des Prognosemoduls der Firma Klafka & Hinz Energie-Informationen-Systeme geschrieben. Zu deren Kunden gehören Energievertriebe, deren Aufgabenfeld auch die Prognose von Energiezeitreihen für Großkunden umfasst. Ziel der Arbeit ist die Entwicklung eines neuen Prognosealgorithmus und die Analyse von Zeitreihen im Hinblick auf eine Risikobewertung derselben.

Energie wird an verschiedenen Handelsplätzen mit unterschiedlichen Zeitperspektiven gehandelt. Eine möglichst genaue Zeitreihenprognose ist dabei die Grundlage der Energiebeschaffung, da die Strompreise auf dem Markt starken Schwankungen unterliegen. Differenzen zwischen prognostizierten und tatsächlichen Verbräuchen müssen zu hohen Preisen ausgeglichen werden und bedeuten gleichzeitig ein finanzielles Risiko in der Energiebeschaffung.

Das Prognosemodul von FirstSale besitzt bereits umfangreiche Prognosefunktionalitäten. Jedoch überfordert dies einige Kunden, sodass diese die Konfiguration vernachlässigen und schlecht Prognosen in Kauf nehmen.

Eine Risikobewertung von Neukunden ist mit dem aktuellen Funktionsumfang von *FirstSale* nur eingeschränkt möglich. Ziel dieser Arbeit ist es, einen selbstkonfigurierenden Prognosealgorithmus zu finden und damit die Prognose beim Kunden zu verbessern, sowie eine automatisierte Risikobewertung auf Basis einer Zeitreihe zu ermöglichen.

Das neue Prognoseverfahren soll in der Lage sein, historische Verbrauchsdaten zu analysieren und sich selbst optimal zu konfigurieren. Aus Gründen der Akzeptanz beim Kun-

den sollte die Arbeitsweise leicht verständlich und die Prognoseerstellung gut nachvollziehbar sein.

Das gewählte Verfahren errechnet den Lastgang für einen Prognosetag als Durchschnittslastgang mehrerer repräsentativer Archivtage. Dazu werden die Tageslastgänge im Datenarchiv anhand ihrer kalendarischen Eigenschaften gruppiert. Der Prognosealgorithmus ermittelt anhand von statistischen Kennzahlen die passende Gruppierung und wendet ein Zeitfenster zur Abbildung der Saisonalität in Zeitreihen an. Auf diese Weise wird das Datenarchiv auf einige wenige Lastgänge reduziert, die dann zur Prognose herangezogen werden.

Dieses Verfahren wird in einem mehrstufigen Algorithmus implementiert und durch Simulation einer Langfristprognose auf 355 Zeitreihen ausgewertet. Zur finanziellen Evaluierung werden Regelernergiepreise zugrunde gelegt.

Die Auswertung ergibt eine sukzessive Verbesserung der Prognose durch den Algorithmus. Auch in finanzieller Hinsicht sind Einsparungen möglich. Allerdings traten große statistische Ausreißer auf: Insbesondere stark zufallsbehaftete Zeitreihen oder solche mit Trendanteilen schneiden sehr schlecht ab. Anschließend wird das ebenfalls auftretende Problem von verschiedenen Verbrauchsmustern innerhalb eines Prognosetages näher betrachtet. Der Einsatz viertelstündlicher Prognoseparameter verbessert den Algorithmus entsprechend.

Für die Risikobewertung einer Zeitreihe wird der Einfluss, der im Algorithmus genutzten Parameter und weiterer statistische Kennzahlen ausgewertet, welche die Zyklizität einer Zeitreihe und die Abhängigkeit von kalendarischen Ereignissen beschreiben. Es wird ein lineares Modell der Abhängigkeiten aufgestellt und dessen Koeffizienten mittels Regressionsanalyse geschätzt. Ergebnis ist eine Funktion, die den erwarteten Fehler des vorgestellten Algorithmus je Prognosetag berechnet. Diese kann zur Risikobewertung einer Zeitreihe eingesetzt werden.

Als Nebenprodukt der Regressionsanalyse kann eine neue Zielfunktion für die Optimierung der Zeitreihenprognose aufgestellt werden. Deren Auswertung ergibt jedoch keine weitere Verbesserung der Prognosequalität.

9.2. Limitationen der Arbeit und zukünftige Forschung

Der vorgestellte Algorithmus implementiert eine Methode zur langfristigen Prognose von Zeitreihen. Über die Notwendigkeit einer kurzfristigen Prognose wird im Rahmen der Arbeit zwar gesprochen, mit Blick auf den Umfang jedoch bewusst ausgespart (siehe

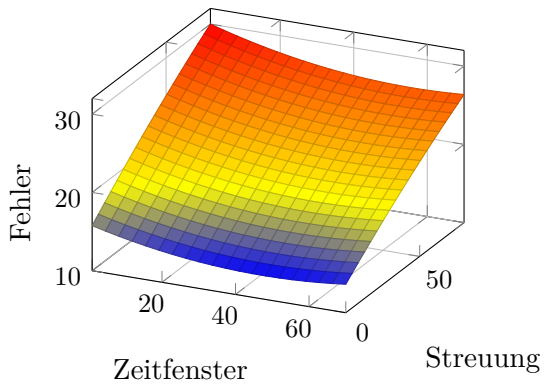
Kapitel 5.2.3). Der Algorithmus kann durch eine Veränderung des Archivzeitraums leicht auf eine Kurzfristprognose umgestellt werden. Von Interesse wäre es zu erfahren, in wie weit sich die Ergebnisse einer Regressionsanalyse nach Kapitel 8 verändern.

Ebenfalls ist es denkbar, das Instrument der Regressionsanalyse auf Zeitreihenebene durchzuführen und somit individuelle Parameter für eine Prognose nach dem Verfahren aus Kapitel 8.6 zu finden. Als weiteren Anreiz für die zukünftige Forschung drängt sich eine Kombination des Algorithmus, welcher viertelstündliche Parameter erzeugt (siehe 7), auf mit der Zielfunktion nach den Ergebnissen der Regressionsanalyse.

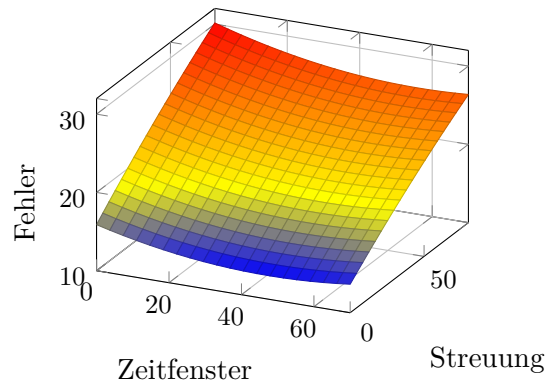
In der vorliegenden Arbeit wird der Einfluss von Temperaturzeitreihen ignoriert. Dies könnte ein Schwachpunkt des Algorithmus sein, stellt die Literatur das Wettergeschehen doch teilweise als wichtigsten Einflussfaktor vor (siehe Kapitel 3.1). Für die langfristige Prognose ist z.B. die Tagesdurchschnittstemperatur zwar schlecht vorhersagbar und teilweise durch die Saisonkomponenten abgedeckt, jedoch wird sie für eine Kurzfristprognose von wenigen Tagen oder Stunden in die Zukunft unverzichtbar sein. Weitere Analysen sollten auch den Einfluss der Temperatur auf die einzelne Zeitreihe analog zu den Schulferien berücksichtigen.

Als Maß für die Inhomogenität des Datenarchivs wurde bisher ausschließlich die Streuung herangezogen. Es ist jedoch denkbar, dass andere Kennzahlen, die weniger sensibel für statistische Ausreißer sind, bessere Ergebnisse liefern. Mögliche Kandidaten hierfür sind der Median und der Quantilsabstand. Dies sollte mit dem vorliegenden Algorithmus getestet und ausgewertet werden.

A. Anhänge



(a) Archivtiefe $z = 2$ Jahre



(b) Archivtiefe $z = 3$ Jahre

Abbildung A.1.: Darstellung des Zusammenhangs von Zeitfensterbreite x , Streuung y und Archivtiefe $z \in \{1, 2\}$ mit dem Prognosefehler

ID	Montag		Dienstag		Mittwoch		Donnerstag		Freitag		Samstag		Sonntag		
ID	WT	TT	BR	WT	TT	BR	WT	TT	BR	WT	TT	BR	WT	TT	BR
alle Tagekeine Gruppierung															
1	x			x			x			x			x		
BrueckenTag - kein BrueckenTag															
1	x		N	x		N	x		N	x		N	x		N
2	x		J	x		J	x		J	x		J	x		J
BrueckenTag - kein BrueckenTag pro Wochentag															
1	x		WT	N											
2			x			WT	N								
3					x										
4									WT	N					
5									x						
6												Sa	N		
7													x		So
8	x		WT	J											
9			x			WT	J								
10					x										
11									x						
12															
13												Sa	J		
14													x		So
Di - Fr															
2			x			WT	N	x		WT	N	x			
Feiertag kein Feiertag															

1	x	So	x	So	x	So	x	So
2	x	WT	x	WT	x	WT	x	Sa
Feiertag kein Feiertag pro Wochentag								
1	x	So	N	x	So	N	x	So
2				x	So	N		
3								
4								
5								
6								
7	x	WT	N	x	WT	N	x	Sa
8								
9								
10								
11								
12								
Feiertag, Brueckentag kein FT, BT								
1	x	So	N	x	So	N	x	So
2	x	WT	N	x	WT	N	x	Sa
Feiertag, Brueckentag kein FT, BT pro WT								
1	x	So	N					
1	x	So	J					
2	x	WT	N	x	WT	N	x	Sa
2	x	WT	N	x	So	N	x	Sa
3								
4								
5								

4	x	J	x	J	x	J	x	J	x	J	x	J
Mo, Di Do, Fr, Sa, So												
1	x	WT	N	x	WT	N	x	WT	N	x	WT	N
2												
3												
4	x	Sa	N	x	Sa	N	x	Sa	N	x	Sa	N
5	x	So	N	x	So	N	x	So	N	x	So	N
6	x	J	J	x	J	J	x	J	J	x	J	J
Schulferien keine Schulferien Pro Wochentag												
1	x	FWT	N	x	FWT	N	x	FWT	N	x	FWT	N
2												
3												
4												
5												
6												
7												
8	x	WT	N	x	WT	N	x	WT	N	x	WT	N
9												
10												
11												
12												
13												
14												
Schulferien Mo Fr, Sa, So												
1	x	FWT	N	x	FWT	N	x	FWT	N	x	FWT	N
2	x	WT	N	x	WT	N	x	WT	N	x	WT	N

[illegible]

[illegible]

B. Literatur

- 50Hertz Transmission GmbH (2014). *Ausgleichsenergiepreise*. URL: http://www.50hertz.com/cps/rde/xchg/trm_de/hs.xsl/2655.htm?rdeLocaleAttr=de%5C&rdeC0Q=SID-341222AA-96C37839.
- Auer, L von (2011). „Ökonometrie“. In: *Eine Einführung* 4, S. 131–170.
- Bastian, Jörgen (1985). *Optimale zeitreihenprognose*. Lang.
- Bruhns, Alexander, Gilles Deurveilher und Jean-Sébastien Roy (2005). „A non-linear regression model for mid-term load forecasting and improvements in seasonnality“. In: *Proceedings of the 15th Power Systems Computation Conference*, S. 22–26.
- Chen, Bo-Juen, Ming-Wei Chang und Chih-Jen Lin (2004). „Load forecasting using support vector machines: A study on EUNITE competition 2001“. In: *Power Systems, IEEE Transactions on* 19.4, S. 1821–1830.
- Deutscher Bundestag (2010). *Verordnung über den Zugang zu Gasversorgungsnetzen (Gasnetzzugangsverordnung GasNZV)*.
- (2013). *Verordnung über den Zugang zu Elektrizitätsversorgungsnetzen (Stromnetzzugangsverordnung StromNZV)*.
- EPEX Spot SE (2014). *Market Data Intraday*. URL: <http://www.epexspot.com/en/market-data/intraday/chart/intraday-chart/2014-02-28/DE>.
- European Energy Exchange AG (2014a). *Im Zentrum des europäischen Energiehandels*. Unternehmensbroschüre.
- (2014b). *Market Data Phelix Futures*. URL: <https://www.eex.com/de/marktdaten/strom/terminmarkt/phelix-futures%5C#!/2014/01/02>.
- Feinberg, Eugene A und Dora Genethliou (2005). „Load forecasting“. In: *Applied mathematics for restructured electric power systems*. Springer, S. 269–285.
- Fidalgo, J Nuno und Manuel A Matos (2007). „Forecasting Portugal global load with artificial neural networks“. In: *Artificial Neural Networks-ICANN 2007*. Springer, S. 728–737.

- Hahn, Heiko, Silja Meyer-Nieberg und Stefan Pickl (2009). „Electric load forecasting methods: Tools for decision making“. In: *European Journal of Operational Research* 199.3, S. 902–907.
- Hippert, Henrique Steinherz, Carlos Eduardo Pedreira und Reinaldo Castro Souza (2001). „Neural networks for short-term load forecasting: A review and evaluation“. In: *Power Systems, IEEE Transactions on* 16.1, S. 44–55.
- Hor, Ching-Lai, Simon J Watson und Shanti Majithia (2005). „Analyzing the impact of weather variables on monthly electricity demand“. In: *Power Systems, IEEE Transactions on* 20.4, S. 2078–2085.
- Koller, Manuel und Werner A Stahel (2011). „Sharpening wald-type inference in robust regression for small samples“. In: *Computational Statistics & Data Analysis* 55.8, S. 2504–2515.
- Russell, Stuart Jonathan u. a. (1995). *Artificial intelligence: a modern approach*. Bd. 2. Prentice hall Englewood Cliffs.
- Shively, Bob und John Ferrare (2008). *Understanding Today's Electricity Business*. Enerdynamics LLC.
- Taylor, James W, Lilian M De Menezes und Patrick E McSharry (2006). „A comparison of univariate methods for forecasting electricity demand up to a day ahead“. In: *International Journal of Forecasting* 22.1, S. 1–16.

C. Versicherung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften anderer entnommen sind, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form noch nicht als Prüfungsarbeit eingereicht worden.

Aachen, den 01.06.2015