



**BERGISCHE
UNIVERSITÄT
WUPPERTAL**

Fakultät für Elektrotechnik
Informationstechnik und Medientechnik

Zeitreihenanalyse von Prozessdaten der Aluminiumelektrolyse für die Vorhersage der Elektrolyttemperatur

Dissertation zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften (Dr.-Ing.)

eingereicht durch

Niclas Bartels

bei der



**BERGISCHE
UNIVERSITÄT
WUPPERTAL**

Fakultät für Elektrotechnik, Informationstechnik und Medientechnik

Gutachter

Erstgutachter Prof. Dr.-Ing. Dietmar Tutsch
Zweitgutachter Prof. Dr.-Ing. Tobias Meisen

Wuppertal, den 29. November 2024

Zeitreihenanalyse von Prozessdaten der Aluminiumelektrolyse für die Vorhersage der Elektrolyttemperatur

Niclas Bartels

29. November 2024

Abstract

The industrial production of aluminium is a very energy-intensive process. It is essential to conduct regular bath temperature measurements to ensure stable aluminium production. However, the use of sensors installed directly in the cells for bath temperature measurement is not feasible due to the corrosive nature of the production environment, which would rapidly corrode and dissolve the sensors. Nevertheless, continuous measurement of the bath temperature is desirable in order to gain further insights into the state of a reduction cell.

This thesis presents an analysis of data-based methods with the aim of predicting the bath temperature. First, the interactions between process variables in aluminium electrolysis and their influence on the bath temperature are presented. The temporal behaviour of the bath temperature is analysed with regard to autocorrelation, periodicity and stationarity. Two approaches for predicting the bath temperature are presented which utilise data-based models derived from the fields of time series analysis and machine learning.

The first approach is a daily forecast of the bath temperature, which is generated by a global autoregressive model (AR model). A comparison is made between the global AR model and local time series models. The second approach is the nowcast of the bath temperature, which is generated by a global random forest regression model. Furthermore, an interpretation of the generated nowcasts by the random forest regression model is presented. This approach represents a novel method for analysing the causes of changes in bath temperature, which, to the best of my knowledge, has not yet been demonstrated in the field of the aluminium electrolysis.

The results demonstrate that the global AR(10) model can replace local time series models in predicting the bath temperature. The global AR(10) model exhibits a mean absolute error of 4.55 °C on the test data. The random forest regression model achieves a mean absolute error of 4.53 °C for the present prediction on the same test data. In contrast to the global model, the random forest regression model predicts changes in the bath temperature instead of absolute bath temperature values, as this approach yields superior model performance within the cross-validation. The combination of the two global models improves the mean absolute error to 4.31 °C.

Zusammenfassung

Die industrielle Herstellung von Aluminium ist ein sehr energieintensiver Prozess. Eine regelmäßige Kontrolle der Elektrolyttemperatur ist erforderlich, um eine stabile Produktion von Aluminium zu gewährleisten. Eine zerstörungsfreie Messung der Elektrolyttemperatur über eine in den Öfen angebrachte Sensorik ist jedoch nicht möglich, da die korrosive Produktionsumgebung die Sensoren in kürzester Zeit auflösen würde. Eine kontinuierliche Messung der Elektrolyttemperatur ist dennoch erstrebenswert, um weitere Rückschlüsse auf den Zustand eines Elektrolyseofens zu gewinnen. Im Rahmen dieser Arbeit werden datenbasierte Methoden auf ihre Eignung zur Vorhersage der Elektrolyttemperatur untersucht. Zunächst werden die Wechselwirkungen von Prozessvariablen der Aluminiumelektrolyse sowie ihre Einflüsse auf die Elektrolyttemperatur aufgezeigt. Die zeitlichen Verläufe der Elektrolyttemperaturen werden hinsichtlich der Autokorrelation, Periodizität und Stationarität analysiert. Zwei Lösungsansätze für die Vorhersage der Badtemperatur werden präsentiert, die Vorhersagemodelle aus dem Bereich der Zeitreihenanalyse und des maschinellen Lernens aufgreifen.

Der erste Lösungsansatz ist eine Tagesprognose (Forecasting) der Elektrolyttemperatur, die mittels eines globalen autoregressiven Modells (AR-Modell) erstellt wird. Das globale AR-Modell wird im Rahmen der Untersuchung mit lokalen Zeitreihenmodellen verglichen. Der zweite Lösungsansatz ist eine Gegenwartsvorhersage (Nowcasting) der Elektrolyttemperatur, die mit einem globalen Random Forest (RF) Regressionsmodell erzeugt wird. Zudem wird ein Ansatz aufgezeigt, um die Gegenwartsvorhersagen des RF-Regressionsmodells interpretieren zu können. Dieser Ansatz stellt eine neuartige Methode zur Ursachenanalyse von Änderungen der Elektrolyttemperatur dar, die nach bestem Wissen im Bereich der Aluminiumelektrolyse bislang nicht aufgezeigt wurde.

Die Ergebnisse zeigen, dass ein globales AR(10)-Modell die lokalen Zeitreihenmodelle hinsichtlich der Vorhersage der Elektrolyttemperatur substituieren kann. Auf den Testdaten erreicht das globale AR(10)-Modell einen Mean Absolute Error (MAE) von 4,55 °C. Das RF-Regressionsmodell erreicht für die Gegenwartsvorhersage auf den gleichen Testdaten einen MAE von 4,53 °C. Anstatt der Badtemperatur sagt das RF-Regressionsmodell die Badtemperaturänderung vorher, da dies innerhalb der Cross-Validation zu einer besseren Modellperformanz führte. Eine Modellkombination, bestehend aus den beiden globalen Modellen, verbessert den MAE zusätzlich auf 4,31 °C.



Danksagung

Den Weg meiner Promotion ebnete mir mein Doktorvater, Professor Tutsch. Er ermöglichte mir, dass ich meinen wissenschaftlichen Interessen in voller Gänze und ohne Einschränkungen nachgehen konnte. Hierdurch konnte ich unterschiedliche Forschungsbereiche miteinander verknüpfen, die letztendlich zu dieser Dissertation führten. Ohne sein Vertrauen in mich wäre es wahrscheinlich nicht zu der vorliegenden Dissertation gekommen. Danke dir, Dietmar!

Auch meinem Zweitgutachter, Professor Meisen, möchte ich danken, der mich auf der Zielgeraden mit wertvollem Feedback versorgt hat. Ein großer Dank geht an die Mitarbeitenden der TRIMET Aluminium SE in Essen. Hier sind vor allem Roman Düssel und Hanno Vogel zu nennen, die mich von Anfang an begleitet haben. Sie ermöglichten nicht nur eine großartige Zusammenarbeit, sondern gaben mir auch die Möglichkeit, in den praktischen Teil der Aluminiumelektrolyse einzutauchen.

Große Unterstützung erhielt ich auch von meinen (ehemaligen) Kolleginnen und Kollegen, auf die ich immer zukommen konnte, die mir den Rücken freigehalten und mich nicht nur auf fachlicher Ebene unterstützt haben. Insbesondere sind hier Dimitrios Savvidis, Bernard Beitz, Tobias Wiegard, Robert Roth, Christoph Brandau, Inge Günther, Ron Kremser und Dena Zaiß zu nennen.

Meine Frau Raphaele stand mir in dieser Zeit immer zur Seite und hatte immer ein offenes Ohr für mich. Die Gespräche und Diskussionen mit ihr, die auch auf wissenschaftlicher Ebene stattgefunden haben, haben mich immer wieder dazu motiviert, am Ball zu bleiben. Unsere gemeinsame Reise nach Neuseeland im Jahr 2018, um meine Publikation auf der 12th Australasian Aluminium Smelting Technology Conference vorzustellen, war für mich ein ganz besonderes Highlight. Danke dir für deine Hingabe, Zuversicht und Unterstützung.

Ich danke meinen Eltern, die mich während meines Studiums und auf dem Weg der Dissertation in vielerlei Hinsichten unterstützt haben. Ihr Stolz und ihre Zuversicht haben dazu beigetragen, dass ich die Dissertation bis zum Ende durchgezogen habe. Ich danke meinen Schwiegereltern, die ständiges Interesse an meiner Arbeit gezeigt und mir immer viel Wertschätzung entgegengebracht haben. Meiner Schwester und meinen Freunden, Tobias und Nils, verdanke ich die gewisse Portion an Ablenkung, die in dieser Zeit hin und wieder auch nötig war.

Danke euch allen! Meine Dissertation wäre ohne eure Unterstützung und Zusammenarbeit nicht denkbar gewesen.

Verwendung von KI-unterstützten Technologien während des Schreibprozesses

Während des Schreibprozesses dieser Arbeit nutzte der Autor DeepL Write und DeepL Übersetzer des Unternehmens DeepL SE in der kostenlosen und kostenpflichtigen Variante, um Textverbesserungen, Umformulierungen und Übersetzungen zu erhalten. Nach der Nutzung dieser Tools/Dienste hat der Autor den Inhalt überprüft und nach Bedarf editiert. Der Autor übernimmt die volle Verantwortung für den Inhalt dieser Arbeit.

Inhaltsverzeichnis

Verzeichnisse	ix
Abbildungsverzeichnis.....	xi
Tabellenverzeichnis.....	xvii
Abkürzungsverzeichnis.....	xix
KAPITEL 1	
Einleitung	1
1.1 Aufbau der Arbeit	3
1.2 Projektpartner	5
KAPITEL 2	
Ziel und Lösungsansätze	9
2.1 Ist-Zustand.....	10
2.2 Ziel-Zustand	11
2.3 Literaturüberblick	12
KAPITEL 3	
Grundlagen der Aluminiumelektrolyse	17
3.1 Hall-Héroult-Prozess	17
3.2 Elektrolyttemperatur (Badtemperatur)	20
3.3 Zusammenfassung	29
KAPITEL 4	
Grundlagen der Zeitreihenanalyse	31
4.1 Zeitreihe	31
4.1.1 Komponenten einer Zeitreihe	32
4.1.2 Zeitreihenmerkmale/Anwendungsbereiche	34
4.1.3 Stationäre Zeitreihen	37
4.1.4 Autokorrelation	38
4.1.5 Periodogramm	41
4.2 Vorhersagemethoden	41
4.2.1 Lokale und globale Vorhersagemodelle.....	42
4.2.2 Autoregressives Modell.....	43
4.2.3 Random Forest.....	44
4.3 Validieren und Testen.....	47
4.3.1 Three-Way Holdout	48
4.3.2 Blocked Cross-Validation (CV-BI)	50
4.3.3 Repeated Holdout (Rep-Holdout).....	51
4.3.4 Prequential-Methode	51
4.4 Zusammenfassung	52

KAPITEL 5

Proof of Concept	55
5.1 Hypothesen/Lösungsansätze	56
5.2 Datenbereitstellung	57
5.2.1 Anpassung der Zeitstempel	57
5.2.2 Verknüpfung mit weiteren Betriebsdaten	60
5.2.3 Zusammenfassung	61
5.3 Zeitreihenanalyse der Badtemperaturdaten	62
5.3.1 Trainings- und Testdaten	62
5.3.2 Verteilung und Ausreißer	62
5.3.3 Badtemperaturänderungen	67
5.3.4 Datenlücken	72
5.3.5 Autokorrelation	74
5.3.6 Periodogramm	76
5.3.7 Stationarität	77
5.4 Lösungsentwicklung	78
5.4.1 Forecasting (AR-Modell)	79
5.4.2 Nowcasting (Random Forest)	81
5.4.3 Analyse (Nowcasting)	90
5.4.4 Evaluation	93
5.5 Zusammenfassung	95

KAPITEL 6

Fazit und Ausblick	99
---------------------------	-----------

Anhang	105
A Abbildungen	105
Literaturverzeichnis	111

Abbildungsverzeichnis

1.1	Die sechs Phasen des Vorgehensmodells “Machine Learning for Production (ML4P)”. Abbildung abgeändert nach [BP20, S. 3].	4
1.2	Ende-zu-Ende-Anordnung der PBPF-Öfen, die von der TRIMET in einer der drei Hallen in Essen betrieben werden. Das Bild wurde von der TRIMET zur Verfügung gestellt.	6
1.3	Arbeiten an den Wärmetauschern, die an einem Ofen installiert sind. Die Wärmetauscher sind direkt an der Ofenwand angebracht und mit davor befestigten Sammelboxen über flexible Lüftungsschläuche verbunden. Das Bild wurde von der TRIMET zur Verfügung gestellt.	7
2.1	Initiales ML-Pipeline-Diagramm.	9
2.2	Zwei Messgeräte der Firma Heraeus Electro-Nite, mit denen die Badtemperatur der Öfen bei der TAE gemessen wird. Die Verwendung dieser Bilder wurde von der Firma Heraeus Electro-Nite für diese Arbeit genehmigt.	11
3.1	Querschnittsansicht eines Aluminiumelektrolyseofens. Abgeändert nach [GB19, S. 2137].	18
3.2	Zusammenhang zwischen der Aluminiumoxidkonzentration und der Ofenspannung bei unterschiedlichen Anode-Kathode-Abständen (ACD). Abgeändert nach [Hau16b, S. 158].	19
3.3	Wechselwirkungen zwischen den einzelnen Prozessvariablen eines Aluminiumelektrolyseofens. Abgeändert nach [STC+09, S. 312].	21
3.4	Mit der Erhöhung des Energieeintrags stellt sich eine neue Wärme- und Materialbilanz eines Ofens ein. Abgeändert nach [RIW+16, S. 818].	22
3.5	Zeitlicher Badtemperaturverlauf der Öfen 1099, 1103, 1107 und 1109 vor, während und nach einer Hallenschaltung, die am 15.03.2021 gegen 12 Uhr stattgefunden hat. Für eine bessere Lesbarkeit der Verläufe sind die gemittelten Badtemperaturen durch Linien miteinander verbunden.	23
3.6	Mit der Erhöhung der Zugabe an Aluminiumfluorid (AlF_3) stellt sich eine neue Wärme- und Materialbilanz des Ofens ein. Abgeändert nach [RIW+16, S. 818].	26
3.7	Entstehung eines Ansatzes unter einer Anode. Abgeändert nach [KGD+21, S. 766; Mar17, S. 817; Mar18, S. 487].	27

3.8	Der Schlamm auf den Boden eines Ofens wirkt wie ein zusätzlicher elektrischer Widerstand und führt zu einer ungleichmäßigen Stromverteilung. Abgeändert nach [Tan10, S. 27].	28
4.1	Zwölf Zeitreihen aus dem M1- und M3-Wettbewerb [Hyn18] besitzen unterschiedliche Stärken bezüglich der saisonalen und Trend-Zyklus Komponente. Abbildung abgeändert nach [GKD+23, S. 2].	35
4.2	Beispiel für ein Korrelogramm der AKF. Entnommen aus [CM09, S. 81–83]. . .	39
4.3	Beispiel für ein Korrelogramm der PAKF. Entnommen aus [CM09, S. 81–83]. .	40
4.4	Aufbau eines RF-Modells, das aus drei Entscheidungsbäumen besteht. Abbildung abgeändert nach [KND15, S. 166].	46
4.5	Bootstrap Aggregating (Bagging) und Subspace Sampling angewendet auf einen Datensatz, der die Merkmale M1, M2 und M3 und die Zielvariable Y beinhaltet. Abbildung abgeändert nach [KND15, S. 166].	47
4.6	Einteilung der Daten mit der Three-Way-Holdout-Methode in einen Trainings-, Validierungs- und Testdatensatz. Die zeitliche Komponente der Daten (Out-of-Time Sampling) [KND15, S. 412] wird dabei berücksichtigt.	49
4.7	Ablauf der CV-BI mit $k = 10$. Die zeitliche Anordnung der Beobachtungen bleibt bei der CV-BI erhalten. In einem finalen Schritt werden alle Trainingsdaten für das Training des Modells mit der höchsten Performanz verwendet, das für einen abschließenden Test auf die Testdaten angewendet wird. Abbildung abgeändert nach [CTM20, S. 2001; RM17, S. 192].	50
4.8	Vorgehensweise der Rep-Holdout-Methode. Aus einem vorher festgelegten Zeitfenster wird zufällig ein Zeitpunkt a gewählt, der die Zeitreihe in einen Trainings- und Validierungsdatensatz einteilt. Die Einteilung wird mehrmals durchgeführt. Die ermittelten Performanzwerte werden anschließend zu einem Gesamtwert aggregiert. Abbildung abgeändert nach [CTM20, S. 2000].	51
4.9	Vorgehensweise der Prequential-Methode mit elf Blöcken. Die zeitliche Anordnung der Beobachtungen bleibt bei dieser Methode erhalten. Die schraffierten Blöcke werden in dem jeweiligen Schritt nicht berücksichtigt. In einem finalen Schritt werden alle Trainingsdaten für das Training des Modells mit der höchsten Performanz verwendet, das für einen abschließenden Test auf die Testdaten angewendet wird. Abbildung abgeändert nach [CTM20, S. 2000; GKD+18, S. 5].	52
5.1	Iterative Vorgehensweise innerhalb der Proof-of-Concept-Phase. Abbildung abgeändert nach [BP20, S. 9].	55
5.2	Angepasstes ML-Pipeline-Diagramm für die Vorhersage der Badtemperatur. .	56

5.3	Die Boxplots zeigen die Verteilung der zeitlichen Differenzen für die Öfen der ersten Schicht. Die Differenzen wurden jeweils zwischen dem Messzeitpunkt (UTC) der Badtemperatur und dem Beginn der ersten Schicht berechnet. . . .	58
5.4	Der Badtemperaturverlauf des Ofens 1001 ab dem 01.09.2022 bis zum 30.09.2022 vor und nach der Anpassung der Badtemperaturdaten.	61
5.5	Zeitliche Aufteilung der Badtemperaturdaten in einen Trainings- und Testdatensatz.	63
5.6	Ein Boxplot, der anhand aller Badtemperaturdaten berechnet wurde.	63
5.7	Ein Boxplot für jeden der 60 untersuchten Öfen, berechnet aus den jeweiligen Badtemperaturen. Eine größere Version dieser Abbildung befindet sich in Anhang A.2.	64
5.8	Badtemperaturverläufe von vier Öfen in einem Zeitraum vom 01.01.2022 bis zum 31.12.2022. Die vertikalen Markierungen zeigen Datenlücken in den Badtemperaturverläufen an.	65
5.9	Badtemperaturverläufe von vier Öfen in einem Zeitraum vom 01.10.2022 bis zum 31.10.2022. Die vertikalen Markierungen zeigen Datenlücken in den Badtemperaturverläufen an.	65
5.10	Ein Boxplot für jeden der 60 untersuchten Öfen, berechnet aus den jeweiligen Badtemperaturänderungen. Eine größere Version dieser Abbildung befindet sich in Anhang A.3.	66
5.11	Badtemperaturverläufe von vier Öfen, die eine starke Badtemperaturänderung aufweisen.	67
5.12	Zeitlicher Verlauf der Prozessvariablen des Ofens 1019.	68
5.13	Zeitlicher Verlauf der Prozessvariablen des Ofens 1034.	69
5.14	Zeitlicher Verlauf der Prozessvariablen des Ofens 1054.	70
5.15	Zeitlicher Verlauf der Prozessvariablen des Ofens 1058.	71
5.16	Anzahl an Datenlücken in den Badtemperaturdaten für jeden der 60 untersuchten Öfen aus Halle 1 im Jahr 2022. Eine größere Version dieser Abbildung befindet sich in Anhang A.4.	73
5.17	Zeitlich aufgetragene Datenlücken in den Badtemperaturdaten, die für 60 Öfen aus Halle 1 im Jahr 2022 ermittelt wurden. Ein grüner Punkt stellt eine Datenlücke dar.	73
5.18	Badtemperaturverläufe von den Öfen 1001, 1058, 1110 und 1115. Die grünen Punkte stellen die Werte dar, die für das Auffüllen der Datenlücken mithilfe der LOCF-Methode verwendet wurden.	74

5.19	Für jeden Ofen wurden die AKF und PAKF bis zur Verzögerungsstelle $k = 35$ berechnet. Die Koeffizienten wurden anschließend gemittelt und gegen die Verzögerungsstellen aufgetragen. In (a) sind die gemittelten Koeffizienten der AKF und in (b) die gemittelten Koeffizienten der PAKF jeweils mit einem Konfidenzintervall von 95 % zu sehen.	75
5.20	Für jeden einzelnen Badtemperaturverlauf wurde ein Periodogramm ohne anschließende Glättung berechnet. Anschließend wurden die einzelnen Periodogramme gemittelt, um einen globalen Eindruck über sich wiederholende Muster in den Badtemperaturverläufen zu erhalten.	76
5.21	Stärke der zweitägigen Saisonalität, die für jeden Badtemperaturverlauf in den Trainingsdaten berechnet wurde. Die Verteilung der berechneten Werte ist mit dem Boxplot dargestellt.	77
5.22	Badtemperaturverläufe (nicht saisonal differenziert) von den Öfen 1009, 1027, 1028 und 1036, für die die Nullhypothese des KPSS-Tests abgelehnt wird. . . .	78
5.23	Gemittelter MAE, der jeweils gegen die berücksichtigten Verzögerungsstellen eines globalen AR-Modells aufgetragen ist.	80
5.24	Ablauf der Gegenwartsvorhersage der Badtemperatur. Abgeändert nach [SBMW22, S. 376].	82
5.25	Erstellung von Zeitreihenmerkmalen über die 5 min-Daten in einem 24 Stunden Intervall. Abgeändert nach [KGD+21, S. 771].	84
5.26	Berechnete PFI für jeden Versuch. (a) Die Zielvariable ist die Badtemperatur. (b) Die Zielvariable ist die Badtemperatur. Der Wert der letzten Badtemperaturmessung wird als Merkmal berücksichtigt. (c) Die Zielvariable ist die Badtemperaturänderung. (d) Die Zielvariable ist die Badtemperaturänderung. Der Wert der letzten Badtemperaturmessung wird als Merkmal berücksichtigt. . .	87
5.27	PFI des RF-Regressionsmodells mit reduzierten Merkmalen.	88
5.28	MAE für jeden Ofen, den das RF-Regressionsmodell auf den Testdaten erzielt.	89
5.29	Zeitlich aufgetragene Gegenwartsvorhersagen des RF-Regressionsmodells für die Öfen 1001 und 1058. Zum Vergleich sind weiterhin die tatsächlich gemessenen Badtemperaturen aufgetragen. Die Standardabweichung (σ RF) gibt die Streuung der Gegenwartsvorhersagen der einzelnen Entscheidungsbäume des RF-Modells an.	90
5.30	Ein Waterfall-Chart, der die Zusammensetzung der Gegenwartsvorhersage vom 14.04.2023 des Ofens 1001 darstellt. Angaben sind auf zwei Nachkommastellen gerundet.	91
5.31	Ein Waterfall-Chart, der die Zusammensetzung der Gegenwartsvorhersage vom 19.04.2023 des Ofens 1058 aufzeigt. Angaben sind auf zwei Nachkommastellen gerundet.	92
5.32	Verteilungen der Vorhersagefehler der globalen Modelle auf den Testdaten. . .	94

5.33	Boxplot für die Vorhersagefehler der Modellkombination auf den Testdaten. .	95
6.1	Dashboard für die Analyse der Gegenwartsvorhersagen des RF-Regressionsmodells. Abgeändert nach [Abd23, S. 51].	102
A.1	Die Boxplots zeigen die Verteilung der zeitlichen Differenzen für die Öfen der zweiten und dritten Schicht. Die Differenzen wurden jeweils zwischen dem Messzeitpunkt (UTC) der Badtemperatur und Beginn (UTC) der ersten Schicht berechnet.	105
A.2	Ein Boxplot für jeden der 60 untersuchten Öfen, berechnet aus den jeweiligen Badtemperaturen.	106
A.3	Ein Boxplot für jeden der 60 untersuchten Öfen, berechnet aus den jeweiligen Badtemperaturänderungen.	107
A.4	Anzahl an Datenlücken in den Badtemperaturdaten für jeden der 60 untersuchten Öfen aus Halle 1 im Jahr 2022.	108
A.5	Korrelationen nach Pearson zwischen den berechneten Zeitreihenmerkmalen in den Trainingsdaten. Weiterhin sind die Korrelationen der Badtemperatur und der Badtemperaturänderung zu sehen.	109
A.6	Korrelationen nach Spearman zwischen den berechneten Zeitreihenmerkmalen in den Trainingsdaten. Weiterhin sind die Korrelationen der Badtemperatur und der Badtemperaturänderung zu sehen.	110

Tabellenverzeichnis

3.1	Badtemperaturdifferenz für jeden Ofen zwischen der letzten Messung vor der Hallenschaltung und der letzten Messung während der einstündigen Hallenschaltung am 15.03.2021. Die mittlere Differenz, die aus den einzelnen Differenzen berechnet wurde, liegt bei $-12,5^{\circ}\text{C}$ mit einer Standardabweichung von $2,7^{\circ}\text{C}$	24
5.1	Zusätzliche Badtemperaturmessungen, die am 08.03.2022 und 09.03.2022 am Ofen 1041 durchgeführt wurden.	58
5.2	In der Tabelle liegen einige Badtemperaturmessungen aus dem Jahr 2022 vor, die am Ofen 1090 durchgeführt wurden. Für den 12.01.2022 und den 16.01.2022 ist jeweils kein Messeintrag vorhanden.	60
5.3	Die zuvor fehlenden Einträge für den Ofen 1090 sind nun für den 12.01.2022 und den 16.01.2022 in der Tabelle vorhanden, wobei der jeweilige Badtemperaturwert als fehlend mit n.a. (nicht angegeben) gekennzeichnet wurde.	60
5.4	MAE, den die Modelle auf den Testdaten erzielten. Angaben sind auf zwei Nachkommastellen gerundet.	81
5.5	Für die Gegenwartsvorhersage verwendete Prozessvariablen.	83
5.6	MAE mit Standardabweichung, die in vier Versuchen innerhalb der CV-BI für das RF-Regressionsmodell ermittelt wurden. Angaben sind auf zwei Nachkommastellen gerundet.	85
5.7	MAE, der von dem jeweiligen Modell auf den Testdaten (ohne Öfen mit neuer Prozessteuerung) erzielt wurde. Die Angaben wurden zunächst aufsteigend sortiert und anschließend auf zwei Nachkommastellen gerundet. Als Forecasting wird die Tagesprognose bezeichnet.	93



Abkürzungsverzeichnis

In diesem Verzeichnis sind die wichtigsten Abkürzungen aufgeführt.

ACD Anode-Cathode Distance

AKF Autokorrelationsfunktion

AR-Modell Autoregressives Modell

ARIMA Autoregressive Integrated Moving Average

Bagging Bootstrap Aggregating

CV Cross-Validation

CV-BI Blocked Cross-Validation

ETS ExponenTial Smoothing

FFT Fast Fourier Transform

FlexTherm Thermische Flexibilisierung der Aluminiumelektrolyse

ID3 Iterative Dichotomizer 3

IOSB Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung

IQA Interquartilabstand

KPSS Kwiatkowski-Phillips-Schmidt-Shin

LfA Lehrstuhl für Automatisierungstechnik/Informatik

LOCF Last Observation Carried Forward

MA Moving-Average

MA-Modell Moving-Average-Modell

MAE Mean Absolute Error

ML-Pipeline-Diagramm Machine Learning Pipeline Diagramm

ML4P Machine Learning for Production

MLP Multilayer Perceptron

PAKF Partielle Autokorrelationsfunktion

PBPF PreBaked Point Feeder

PCA Principal Component Analysis

PFI Permutation Feature Importance

Rep-Holdout Repeated Holdout

RF Random Forest

RMSE Root Mean Squared Error

STL Seasonal and Trend decomposition using Loess

TAE TRIMET Aluminium SE Essen

TRIMET TRIMET Aluminium SE

UTC Coordinated Universal Time

XGBoost eXtreme Gradient Boosting

KAPITEL 1

Einleitung

Die Nutzung erneuerbarer Energien spielt im Zuge der Energiewende eine wesentliche Rolle. So wurde im Jahr 2023 ein Anteil von 51,8 % des gesamten Bruttostromverbrauchs in Deutschland von regenerativen Energiequellen gedeckt. Der Großteil davon wird von Windkraftanlagen an Land und von Photovoltaikanlagen erzeugt. Die Erzeugung von regenerativer Energie ist daher maßgeblich vom Wetter abhängig, was zu Schwankungen in der Energieerzeugung führt. Das macht neue intelligente Lösungen notwendig, um Angebot und Nachfrage auf dem Energiemarkt gleichermaßen bedienen zu können [Umw24, S. 52; Bun].

Ein Lösungsansatz zur Bewältigung dieser Herausforderung am Energiemarkt kann in der Flexibilisierung der industriellen Nachfrage nach Strom gesehen werden [Dü16, S. 2; DMB19, S. 533; DDP16, S. 571]. Der Kerngedanke besteht darin, den Stromverbrauch industrieller Produktionsanlagen in Abhängigkeit vom aktuellen Angebot anzupassen. Infolge eines Angebotsengpasses im Energienetz kann die Produktion entsprechend gedrosselt werden, während bei einem Angebotsüberschuss eine Erhöhung der Produktion erfolgt. Ein vielversprechender Bereich für eine solche Leistungsanpassung ist die industrielle Herstellung von Aluminium, die mit einem hohen Energieverbrauch verbunden ist und bislang mit einer nahezu konstanten Energiezufuhr durchgeführt wird [DDP16, S. 571; Dü16, S. 1–3]. Die Flexibilisierung des Energieeintrags der Aluminiumherstellung eröffnet die Möglichkeit, auf entsprechende Schwankungen im Energienetz zu reagieren [Dü16, S. 2]. Der Wechsel zu einer flexiblen Produktion an Aluminium bringt jedoch Schwierigkeiten im Herstellungsprozess mit sich, die es mit entsprechenden Anpassungen der Produktionsanlagen zu bewältigen gilt. Die Anpassungen zielen darauf ab, eine hohe Effizienz der Aluminiumproduktion aufrechtzuerhalten und einen Ausfall der Anlagen zu verhindern [Dü16, S. 4; DMB19, S. 533].

Die industrielle Aluminiumherstellung greift auf das etablierte Verfahren nach Hall-Héroult zurück. Das Hall-Héroult-Verfahren wurde 1886 entwickelt und ist bis heute das einzige Verfahren, mit dem Aluminium industriell hergestellt wird [GK93, S. 1; RP20, S. 1895–1896]. In diesem Verfahren wird Aluminiumoxid (Al_2O_3) in flüssigem Kryolith (Na_3AlF_6) mit einem Überschuss an Aluminiumfluorid (AlF_3) gelöst. Das gewonnene flüssige Aluminium setzt sich nach der

Reduktion in der Schmelze mittels der Schmelzflusselektrolyse auf dem Boden des Elektrolyseofens ab und wird in regelmäßigen Abständen abgesaugt [GK93, S. 1, 40; Dü16, S. 15, 44]. Eine Flexibilisierung des Energieeintrags bewirkt nicht nur eine Änderung der Produktionsmenge an Aluminium, sondern gleichfalls eine Veränderung der Energiebilanz des Elektrolyseofens [Dü16, S. 4]. Dies hat wiederum Auswirkungen auf dessen Elektrolyttemperatur (Badtemperatur, Temperatur der Schmelze) [STC+08, S. 309]. Im weiteren Verlauf der Arbeit wird die Elektrolyttemperatur als Badtemperatur bezeichnet.

Die Badtemperatur eines Elektrolyseofens stellt in der Aluminiumelektrolyse eine wesentliche Prozessvariable dar, die regelmäßig kontrolliert werden muss, um eine stabile Produktion von Aluminium zu gewährleisten [GK93, S. 215; Dü16, S. 42]. Die Badtemperatur kann jedoch nicht mit einer in den Öfen angebrachten Sensorik dauerhaft gemessen werden, da die korrosive Umgebung diese innerhalb kurzer Zeit zerstören würde [GK93, S. 215; Dü16, S. 41]. Die Messungen werden daher in bestimmten zeitlichen Abständen von geschultem Personal manuell an jedem Ofen durchgeführt [GK93, S. 215]. Im Anschluss werden die Messungen gespeichert und stehen für die Prozessregelung und weitere Analysen zur Verfügung. Darüber hinaus werden weitere Prozessdaten, wie beispielsweise die Ofenspannung und die Stromstärke, kontinuierlich und automatisiert aufgezeichnet. Die Durchführung manueller Messungen ist jedoch mit Kosten verbunden, die wegen der nicht zerstörungsfreien Messmethode zustande kommen [SO10, S. 1; MAW01, S. 297].

Es stellt sich die Frage, inwieweit die gespeicherten Prozessdaten der Öfen dazu verwendet werden können, um die Badtemperatur eines Ofens zu prognostizieren. Ansätze sind Methoden aus dem Bereich der Zeitreihenanalyse und des maschinellen Lernens, die es ermöglichen, anhand der gesammelten Daten Vorhersagen über ausgewählte Zielgrößen zu treffen. Eine Literaturübersicht über den Einsatz datengetriebener Methoden in der Aluminiumelektrolyse wird in unseren Arbeiten [GKD+18; GKD+23] dargestellt und im weiteren Verlauf dieser Arbeit aufgegriffen.

Der Einsatz datengetriebener Methoden ermöglicht einen schnellen Überblick über den aktuellen Zustand eines Elektrolyseofens. Dabei können potenzielle Ursachen für bestimmte Prozessänderungen, die den Elektrolyseprozess beeinträchtigen, aufgezeigt werden [Maj11, S. 191; MTC+11, S. 377]. Diese Analysen erlangen insbesondere in Zeiten der Energiewende eine hohe Relevanz, da die Primärproduktion von Aluminium nicht mehr mit einer konstanten, sondern mit einer variablen Energiezufuhr betrieben wird [Dü16, S. 3]. Um weiterhin einen stabilen Produktionsprozess zu gewährleisten, sind neue Analysen und Methoden erforderlich, um den Zustand eines Ofens auch bei variabler Energiezufuhr überwachen zu können [Dü16, S. 163–164]. Die Badtemperatur spielt bei der Zustandsüberwachung eines Ofens eine wesentliche Rolle. Der Fokus dieser Arbeit liegt daher auf der Badtemperatur, die mithilfe der Zeitreihenanalyse und dem Bereich des maschinellen Lernens untersucht und vorhergesagt

wird. Es ist nicht auszuschließen, dass sich die in dieser Arbeit vorgestellten Lösungsansätze auch auf die Liquidustemperatur übertragen lassen, die ebenfalls eine wesentliche Prozessvariable in der Aluminiumelektrolyse darstellt. Die Liquidustemperatur ist die Mindesttemperatur, bei der die Aluminiumproduktion betrieben werden muss, ohne dass der Elektrolyt zu erstarren beginnt [Hau16c, S. 804].

Im Rahmen dieser Arbeit werden die Wechselwirkungen der wesentlichen Prozessparameter der Aluminiumelektrolyse detailliert dargestellt und die verschiedenen Einflüsse auf die Badtemperatur aufgezeigt. Die daraus resultierenden Erkenntnisse fließen in die Entwicklung von Vorhersagemodellen ein. Für die Vorhersage der Badtemperatur werden in dieser Arbeit zwei Modelle global eingesetzt und untersucht, das autoregressive Modell (AR-Modell) und Random Forest (RF) Modell. Das globale AR-Modell fungiert als Basismodell und prognostiziert die Badtemperatur für den nächsten Tag anhand historischer Badtemperaturwerte. Eine solche Tagesprognose wird in dieser Arbeit als Forecasting bezeichnet. Im Rahmen eines Vergleichs mit lokalen Zeitreihenmodellen wird das globale AR-Modell ebenfalls evaluiert. Aufbauend auf den Ergebnissen aus unserer Arbeit [GKD+18] wird das globale RF-Modell für eine Gegenwartswartevorhersage (Nowcasting) der Badtemperatur eingesetzt. Als Nowcasting kann die Vorhersage bezeichnet werden, die für den aktuellen Zeitpunkt, für die nahe Zukunft und nahe Vergangenheit getätigt wird [BGR10, S. 5].

Die initiale Motivation für die vorliegende Arbeit ist aus dem Drittmittelprojekt „Thermische Flexibilisierung der Aluminiumelektrolyse (FlexTherm)“ mit der Förderkennzeichnung EFRE-0200490 entstanden, an dem der Aluminiumhersteller und Projektpartner TRIMET Aluminium SE (TRIMET) beteiligt war. Im Rahmen des Projekts wurden von der TRIMET Aluminium SE Essen (TAE) Anpassungen an den Elektrolyseöfen durchgeführt, um einen Produktionsbetrieb mit einer variablen Stromstärke zu ermöglichen. Das Projekt wurde vom Europäischen Fonds für regionale Entwicklung und der Europäischen Union mit einer regulären Laufzeit von drei Jahren (2017 – 2020) gefördert. An diesem Projekt war der Lehrstuhl für Automatisierungstechnik/Informatik (LfA) der Bergischen Universität Wuppertal beteiligt, der die TAE bei der Auswertung der Ofenprozessdaten sowie die Erstellung von Tagesprognosen unterstützte.

1.1 Aufbau der Arbeit

In dieser Arbeit wird Fachwissen aus den Bereichen Zeitreihenanalyse und Aluminiumelektrolyse kombiniert, um eine Vorhersage der Badtemperatur zu ermöglichen. Die vorliegende Arbeit richtet sich daher an Personen mit unterschiedlicher Expertise in den genannten Bereichen. Erfahrene Leser/-innen aus dem Bereich der Aluminiumelektrolyse können Kapitel 3 überspringen und sich stattdessen mit den Grundlagen der Zeitreihenanalyse in Kapitel 4 sowie dem praktischen Teil der Arbeit in Kapitel 5 befassen. Für Personen mit Fachwissen aus dem Bereich der Zeitreihenanalyse sind das Kapitel 3 sowie der praktische Teil in Kapitel 5 relevant.

Den ergebnisorientierten Leser/-innen wird die Zusammenfassung des praktischen Teils in Kapitel 5.5 sowie das Fazit in Kapitel 6 nahegelegt. Kapitel 6 richtet sich zudem an Personen, die weitere Informationen über noch anstehende Arbeiten erhalten wollen. Für das Verständnis des Ziels und der Lösungsansätze dieser Arbeit ist für alle Leser/-innen Kapitel 2 wesentlich. Weitere Informationen über den genauen Aufbau dieser Arbeit und über den Projektpartner erhalten interessierte Personen im vorliegenden Kapitel.

Der Aufbau dieser Arbeit orientiert sich am Vorgehensmodell „Machine Learning for Production (ML4P)“ [BP20]. Das Vorgehensmodell ML4P wurde unter der Projektleitung des Fraunhofer-Instituts für Optronik, Systemtechnik und Bildauswertung (IOSB) entwickelt und ermöglicht eine strukturierte Vorgehensweise für industrielle Projekte im Bereich des maschinellen Lernens. Im Unterschied zu bereits bestehenden Vorgehensmodellen liegt der Schwerpunkt des ML4P-Vorgehensmodells auf der Integration und dem Betrieb maschineller Lernverfahren in der industriellen Produktion [BP20, S. 2]. Das Vorgehensmodell umfasst insgesamt sechs Phasen, die unter anderem die Zielsetzung, Datenaufbereitung, Entwicklung und Evaluation von Lösungsansätzen sowie die Inbetriebnahme der entwickelten Lösungen beinhalten. Die sechs Phasen sind in Abbildung 1.1 dargestellt.

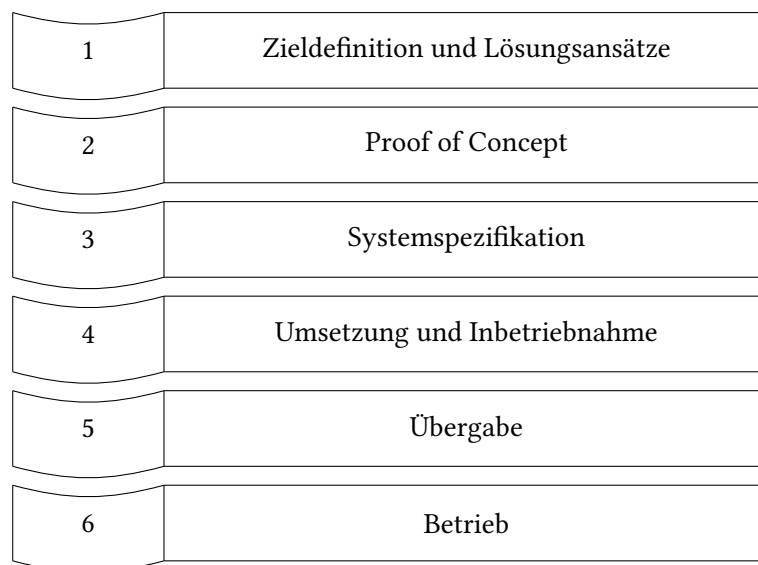


Abbildung 1.1: Die sechs Phasen des Vorgehensmodells „Machine Learning for Production (ML4P)“. Abbildung abgeändert nach [BP20, S. 3].

Die vorliegende Arbeit umfasst die ersten drei Phasen des Vorgehensmodells. In der ersten Phase werden die Ziele und Lösungsansätze definiert, die in Kapitel 2 vorgestellt werden. In diesem werden der Ist-Zustand der TAE und der Ziel-Zustand betrachtet. Zudem wird ein initiales Machine Learning Pipeline Diagramm (ML-Pipeline-Diagramm) präsentiert, um den Ist-Zustand vom Ziel-Zustand abzugrenzen. Des Weiteren werden zwei Messgeräte für die Bad-

temperaturmessung vorgestellt. Eine Literaturübersicht zeigt den Einsatz von datengetriebenen Methoden in der Aluminiumelektrolyse auf. In Kapitel 3 werden die Grundlagen des Hall-Héroult-Prozesses vorgestellt. Der Schwerpunkt des Kapitels liegt auf der Badtemperatur. Die Wechselwirkungen verschiedener Prozessvariablen während des Elektrolysebetriebs werden vorgestellt und deren Auswirkungen auf die Badtemperatur beschrieben. Eine experimentelle Auswertung wird gezeigt, die den Einfluss einer Hallenschaltung (kontrollierte Stromabschaltung einer Halle) auf die Badtemperatur von vier Aluminiumelektrolyseöfen veranschaulicht.

In Kapitel 4 werden grundlegende Konzepte der Zeitreihenanalyse vorgestellt, die in der zweiten Phase (Proof of Concept) des ML4P-Vorgehensmodells zur Analyse der Badtemperaturdaten der TAE herangezogen werden. Ein Schwerpunkt liegt auf der Berechnung von Zeitreihenmerkmalen, die eine Untersuchung der Eigenschaften von Zeitreihen ermöglichen. Darüber hinaus werden das AR-Modell und RF-Modell vorgestellt, die in dieser Arbeit als globale Modelle für eine Badtemperaturvorhersage eingesetzt werden. Die Begründung für die Wahl dieser Modelle und der Unterschied zwischen lokalen und globalen Vorhersagemodellen werden aufgezeigt. Außerdem werden Methoden erläutert, die im Kontext der Zeitreihenanalyse zum Training und zur Validierung von Vorhersagemodellen zum Einsatz kommen.

Das Vorgehen der zweiten Phase (Proof of Concept) wird in Kapitel 5 beschrieben. Hierbei werden zwei Lösungsansätze vorgestellt, die sich hinsichtlich der Art der Vorhersage unterscheiden. Im ersten Lösungsansatz wird mit dem AR-Modell eine Tagesprognose (Forecasting) auf Basis von historischen Badtemperaturdaten erstellt. Im zweiten Lösungsansatz wird mit dem RF-Modell eine Gegenwartsvorhersage (Nowcasting) der Badtemperatur erzeugt. Im Vorfeld werden die Badtemperaturdaten aufbereitet und mit den im Kapitel 4 vorgestellten Methoden untersucht. Im Anschluss werden die Daten für das Training, Validieren und Testen der vorgestellten Vorhersagemodelle eingesetzt und die jeweils erzielte Performanz verglichen. Außerdem wird eine Möglichkeit aufgezeigt, um die Gegenwartsvorhersagen des RF-Modells zu interpretieren.

In Kapitel 6, das der dritten Phase (Systemspezifikation) des Vorgehensmodells zugeordnet wird, werden die Ergebnisse aus der zweiten Phase reflektiert. Außerdem wird eine Möglichkeit vorgestellt, die eine Integration der entwickelten Lösungen in den Produktivbetrieb ermöglicht. Dabei wird ein interaktives Dashboard aufgezeigt, das am LfA entwickelt wurde. Im Ausblick dieser Arbeit werden die weiteren Arbeitsschritte dargelegt, die über diese Arbeit hinausgehen.

1.2 Projektpartner

Der Aluminiumhersteller TRIMET ist in der Produktion und Entwicklung von Aluminiumprodukten tätig und ist eine Tochtergesellschaft der TRIMET SE, die sich in Familienbesitz befindet. Der Sitz der TRIMET und der Muttergesellschaft ist in Essen (Deutschland). Das Geschäft der

Abschnitt 1.2. Projektpartner

TRIMET gliedert sich in drei Bereiche: Marketing & Sales (Essen, Berlin), Primary Products (Essen, Hamburg, Voerde) und Recycling (Essen, Gelsenkirchen, Harzgerode). Darüber hinaus ist die TRIMET mit 65 % an der TRIMET France SAS in Saint-Jean-de-Maurienne (Frankreich) beteiligt [TRI21, S. 7, 45]. Im Geschäftsjahr 2020/2021 wurden 76 % des Umsatzes im Bereich Primary Products erwirtschaftet [TRI21, S. 13].

Die industrielle Herstellung von Aluminium ist sehr energieintensiv [DDPR16, S. 571]. Laut dem Nachhaltigkeitsbericht 2021 der TRIMET wurde im Jahr 2021 am Standort Essen elektrische Energie in Höhe von 2,456 TWh umgesetzt [TRI23, S. 21]. Der spezifische Energiebedarf (Gleichstrom) für die Produktion von 1 Tonne Aluminium am Standort Essen wird mit 13,85 MWh angegeben [TRI23, S. 21]. Nach [Rhe22] ist der elektrische Energiebedarf für das Werk in Essen genauso hoch wie für die gesamte Stadt Essen.

Die TRIMET betreibt an den drei Standorten Essen, Hamburg und Voerde in Deutschland insgesamt 818 Elektrolyseöfen, um Primäraluminium zu produzieren [TRI24]. Von diesen befinden sich 360 Öfen in Essen, die sich auf drei Hallen aufteilen. In einer Halle befinden sich 120 PreBaked Point Feeder (PBPF) Öfen in einer Ende-zu-Ende-Aufstellung [KGD+20, S. 2]. Abbildung 1.2 zeigt die Anordnung der Öfen in einer der drei Hallen bei der TAE.



Abbildung 1.2: Ende-zu-Ende-Anordnung der PBPF-Öfen, die von der TRIMET in einer der drei Hallen in Essen betrieben werden. Das Bild wurde von der TRIMET zur Verfügung gestellt.

Im Rahmen der Flexibilisierungsmaßnahmen wurden am Standort Essen Anpassungen an den Aluminiumelektrolyseöfen vorgenommen. Die TRIMET hat die Ofenprozesssteuerung überarbeitet, jeweils eine Magnetfeldkompensation an den Öfen installiert und Wärmetauscher

angebracht, um weiterhin einen stabilen und effizienten Produktionsbetrieb bei einer variablen Stromstärke zu gewährleisten [Dü16, S. 3, 12; DMB19]. Während die Magnetfeldkompensation eine Aufwölbung des flüssigen Aluminiums in den Öfen bei hohen Stromstärken verhindern soll, sollen die Wärmetauscher die Wärmeverluste über die Ofenwand regulieren [GKD+23, S. 2; DDPR16; DMB19, S. 535]. Die Außenluft wird mithilfe von Ventilatoren durch die Wärmetauscher und somit an den Ofenwänden vorbeigesaugt. Auf diese Weise kann bei einer erhöhten Energiezufuhr die zusätzliche Wärme abgeführt und somit die Wärmebilanz eines Ofens beibehalten werden. Außerdem kann bei einem geringen Energieeintrag Luft im Wärmetauscher angestaut werden, um eine zusätzliche Wärmeisolation für einen Ofen zu erzeugen [DMB19, S. 535; Dü16, S. 4]. Abbildung 1.3 zeigt die an einem Ofen angebrachten Wärmetauscher, die über flexible Lüftungsschläuche mit davor installierten Sammelboxen verbunden sind. Die Installation der Wärmetauscher erfolgte zunächst zu Testzwecken in einer kleinen Gruppe an Elektrolyseöfen und wurde später innerhalb des FlexTherm-Projekts auf die restlichen Öfen in Halle 1 ausgeweitet.



Abbildung 1.3: Arbeiten an den Wärmetauschern, die an einem Ofen installiert sind. Die Wärmetauscher sind direkt an der Ofenwand angebracht und mit davor befestigten Sammelboxen über flexible Lüftungsschläuche verbunden. Das Bild wurde von der TRIMET zur Verfügung gestellt.

KAPITEL 2

Ziel und Lösungsansätze

Anhand des ML4P-Vorgehensmodells nach [BP20], das in Kapitel 1 vorgestellt wurde, werden in diesem Kapitel die Ziele und Lösungsansätze aufgezeigt. Der Fokus liegt auf der Badtemperatur, die in der Aluminiumelektrolyse eine wesentliche Prozessvariable ist. In Abbildung 2.1 ist ein initiales ML-Pipeline-Diagramm zu sehen, das den Ist-Zustand und den gewünschten Ziel-Zustand sowie den Datenfluss aufzeigt.

Der Ist-Zustand umschreibt den aktuellen Zustand bei der TAE hinsichtlich der manuellen Badtemperaturmessung. Hierbei werden zwei Messgeräte vorgestellt, mit der die Badtemperatur bei der TAE gemessen wird. Der Ziel-Zustand zeigt eine mögliche Integration von Badtemperaturvorhersagen in der Aluminiumelektrolyse auf und umfasst die dafür nötigen Schritte. Beide Zustände werden in den nachfolgenden beiden Kapiteln beschrieben. Im Anschluss wird ein Literaturüberblick gegeben, der den Einsatz von datengetriebenen Methoden in der Aluminiumelektrolyse beleuchtet. Dieser Überblick dient als Grundlage für die Entwicklung von potenziellen Lösungsansätzen, die eine Vorhersage der Badtemperatur ermöglichen sollen.

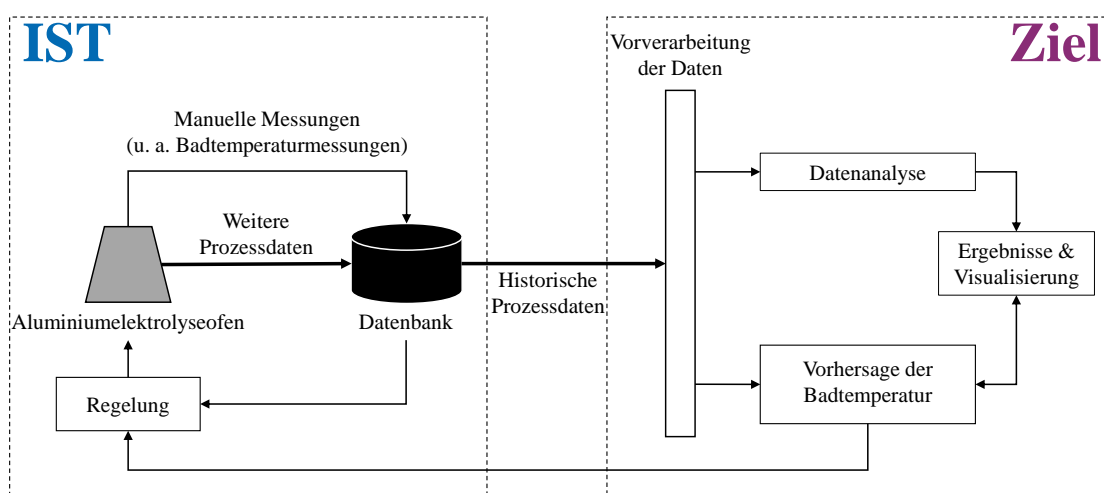


Abbildung 2.1: Initiales ML-Pipeline-Diagramm.

2.1 Ist-Zustand

Bei der TAE werden die Prozessdaten von insgesamt 360 Aluminiumelektrolyseöfen sowohl manuell als auch automatisiert erfasst und innerhalb einer Datenbank gespeichert. Der Datenfluss des Ist-Zustands ist in Abbildung 2.1 abgebildet. Die von den Aluminiumelektrolyseöfen erfassten Daten weisen eine unterschiedliche zeitliche Auflösung auf. Die Badtemperatur wird alle 24 bis 48 Stunden manuell gemessen und der entsprechende Messwert in eine Datenbank gespeichert. Die Liquidustemperatur (Schmelztemperatur) und die chemische Zusammensetzung des Elektrolyten werden in der Regel alle 48 Stunden an jedem Ofen manuell überprüft. Neben den manuellen Messungen werden weitere Prozessvariablen, wie beispielsweise die Stromstärke und Ofenspannung, automatisiert aufgezeichnet und in eine Datenbank mit einem entsprechenden Zeitstempel abgespeichert. Die zeitliche Auflösung der Stromstärke und der Ofenspannung beträgt fünf Minuten. Des Weiteren sind an den Außenseiten von ausgewählten Öfen Thermoelemente installiert [GKD+23]. Die aufgezeichneten Temperaturdaten der Thermoelemente dienen der Berechnung und Überwachung der Wärmebilanz dieser Öfen. Insgesamt können mit einem Zugriff auf die Datenbank historische Daten von Öfen abgerufen und für weitere Analysen verwendet werden.

Zur Regelung der Aluminiumelektrolyseöfen setzt die TAE die 9-Box-Matrix-Regelung ein, um die Badtemperatur durch Anpassungen der Ofenspannung in einem gewählten Kontrollband zu halten [RIW+16, S. 818]. Ein weiteres Ziel der 9-Box-Matrix-Regelung ist die kontrollierte Zugabe von Aluminiumfluorid (AlF_3) anhand der Liquidustemperatur [RIW+16, S. 818].

Die Aluminiumelektrolyseöfen der TAE werden insgesamt von drei Schichten überwacht. Die erste Schicht beginnt um sechs Uhr morgens, während die dritte Schicht um diese Uhrzeit endet. Die Arbeitszeit jeder Schicht beträgt acht Stunden. Jede Schicht übernimmt für eine Gruppe an Öfen die Badtemperaturmessung.

Für die Messung der Badtemperatur verwendet die TAE Messgeräte der Firma Heraeus ElectroNite. In Abbildung 2.2a ist das FiberLab®-Messgerät zu sehen. Mit diesem Gerät können neben der Badtemperatur mit einer Genauigkeit von $\pm 1^\circ\text{C}$ auch noch die Liquidustemperatur sowie die chemische Zusammensetzung des Elektrolyten ermittelt werden [Fiba]. Das in Abbildung 2.2b dargestellte C-V-Therm-Messgerät (Digilance) erlaubt neben der Badtemperaturmessung, die eine Genauigkeit von $\pm 0,5^\circ\text{C}$ aufweist, auch die Messung der Kathodenspannung (Cathode Voltage Drop) [Dig]. Wie bereits erläutert, ist die kontinuierliche Messung der Badtemperatur mithilfe einer in den Öfen angebrachten Sensorik nicht zerstörungsfrei möglich, da die korrosive Produktionsumgebung diese innerhalb kurzer Zeit auflösen würde [Dü16, S. 41; GK93, S. 215]. Eine kontinuierliche Badtemperaturmessung würde jedoch nach [ZJX+94, S. 28–29] den Vorteil bieten, dass sie als zusätzlicher Eingang für die Ofenregelung dienen kann. Des Weiteren kann eine kontinuierliche Badtemperaturmessung nach [ZJX+94, S. 29] dazu beitragen, spezifische

Effekte oder Eingriffe, wie beispielsweise Anodeneffekte oder Anodenwechsel, genauer zu untersuchen. Dieser Hintergrund führt zum Ziel dieser Arbeit, das im nachfolgenden Kapitel dargestellt wird.



(a) FiberLab® [Fibb].



(b) C-V-Therm (Digilance) [Dig].

Abbildung 2.2: Zwei Messgeräte der Firma Heraeus Electro-Nite, mit denen die Badtemperatur der Öfen bei der TAE gemessen wird. Die Verwendung dieser Bilder wurde von der Firma Heraeus Electro-Nite für diese Arbeit genehmigt.

2.2 Ziel-Zustand

In Anbetracht der Problematik einer kontinuierlichen Messung der Badtemperatur [Dü16, S. 41; GK93, S. 215] erscheint es zielführend, die aufgezeichneten Prozessdaten der TAE für eine Badtemperaturvorhersage zu nutzen. Um den angestrebten Ziel-Zustand zu erreichen (vgl. Abbildung 2.1), sollen die historischen Prozessvariablen aus einer zentralen Datenbank abgefragt werden. Eine anschließende Aufbereitung der Daten wird notwendig sein, um eine Datenbasis für die weitere Durchführung von Analysen zu schaffen. Die Datenanalysen zielen darauf ab, neue Erkenntnisse zu schaffen, die für die Entwicklung von Vorhersagemodellen von Nutzen sein können. Es kann nicht ausgeschlossen werden, dass die Erkenntnisse weitere Schritte in der Datenaufbereitung erforderlich machen.

Im Anschluss an die Datenvorverarbeitung und Datenanalysen ist die Entwicklung möglicher Lösungsansätze für eine Badtemperaturvorhersage durch datengetriebene Methoden vorgesehen. Die datengetriebenen Methoden sollen als Bestandteil des Betriebs zur Optimierung der Ofenreglung eingesetzt werden. Auch die Mitarbeitenden des Betriebs können die Vorhersagen nutzen, um Effekte innerhalb des Aluminiumelektrolyseofens genauer zu untersuchen oder etwaige Abnormalitäten frühzeitig zu erkennen. Ferner wird in [SBMW22, S. 378] der Einsatz

von Badtemperaturvorhersagen beschrieben, um fehlerhafte Badtemperaturmessungen zu identifizieren. Bei einer wesentlichen Abweichung zwischen der Vorhersage und der gemessenen Badtemperatur ist eine Wiederholung der Messung erforderlich.

Insgesamt erschließt die vorliegende Arbeit den in Abbildung 2.1 dargestellten Ziel-Zustand, um datengetriebene Methoden in den Produktivbetrieb der TAE zu überführen. Im Rahmen dieser Erschließung werden die einzelnen Schritte, die Vorverarbeitung der Daten, die Durchführung von entsprechenden Analysen und das Training, Validieren und Testen von ausgewählten Vorhersagemethoden, vorgestellt. In den entsprechenden Schritten werden jeweils wesentliche Schwierigkeiten aufgezeigt und Lösungen präsentiert, um eine Badtemperaturvorhersage in der Aluminiumelektrolyse zu erreichen. Darüber hinaus wird eine Möglichkeit für die Visualisierung der Vorhersagen präsentiert, die im Produktivbetrieb von den Mitarbeitenden der TAE eingesetzt werden kann.

2.3 Literaturüberblick

Der Einsatz datengetriebener Methoden ist in der Aluminiumelektrolyse vielfältig. Für unsere Arbeit [GKD+18] wurde bereits Literatur über den Einsatz von datengetriebenen Methoden im Bereich der Aluminiumelektrolyse recherchiert. Diese Recherche wird in diesem Kapitel aufgegriffen und um weitere relevante Literatur ergänzt. Für die Recherche wurde hauptsächlich der Literaturbereich der Aluminiumelektrolyse mit dem Fokus auf datengetriebenen Methoden durchsucht. Die Recherche folgte keiner bestimmten Vorgehensweise.

Als Pseudo-Modell wird in diesem Kapitel ein Modell bezeichnet, das als Grundlage für einen Vergleich mit weiteren Vorhersagemodellen dient. Beispielsweise stellen ein naives Modell, das die letzte Beobachtung als Vorhersage verwendet, oder die Berechnung des Mittelwerts über die Daten, der anschließend als Vorhersage eingesetzt wird, jeweils ein Pseudo-Modell dar. Mit dem Einsatz von Pseudo-Modellen kann der Nutzen von komplexeren Vorhersagemodellen überprüft werden. Diese sollten eine wesentlich bessere Vorhersageperformanz als Pseudo-Modelle aufweisen, um einen praktischen Einsatz begründen zu können.

In [Maj11; MTC+11] wird ein Framework präsentiert, das eine zeitnahe Erkennung von Anodenefekten, wie Anodeneffekte und Ansätze, während des Ofenbetriebs ermöglicht. In [KGD+20] wird eine Möglichkeit beschrieben, um Anodeneffekte vorherzusagen. Mehrere Vorhersagemodelle werden in der Arbeit gegenübergestellt, wobei das RF-Klassifikationsmodell die beste Performanz auf dem Testdatensatz erzielt. In [ZXWZ18] wird zudem ein Lösungsansatz basierend auf eXtreme Gradient Boosting (XGBoost) für die Vorhersage von Anodeneffekten vorgestellt. Anodeneffekte sind in der Aluminiumelektrolyse problematisch, da diese zu einem hohen Energieeintrag führen und klimaschädliche Gase freisetzen [KGD+20, S. 1]. In Kapitel 3.2 wird die Entstehung von Anodeneffekten genauer aufgezeigt. Die Erkennung von Ansät-

zen, die sich an den Anoden bilden können, ist in der Aluminiumelektrolyse wichtig, da diese die Effizienz der Aluminiumproduktion negativ beeinflussen [KGD+21]. In [KGD+21] werden maschinelle Lernverfahren miteinander verglichen, um Ansätze automatisiert erkennen zu können. Das XGBoost-Modell erreicht die beste Performanz auf den Testdaten. In Kapitel 3.2 werden Ansätze an Anoden und deren Einfluss auf die Badtemperatur erklärt.

In [MYTC12] wird der k-Means-Algorithmus eingesetzt, um anhand von Produktionsdaten Abnormalitäten voneinander abzugrenzen. Um das Verhalten von Aluminiumelektrolyseöfen leichter analysieren zu können, werden in [LSS+17] der Fuzzy-c-Means-Algorithmus und k-Means-Algorithmus verwendet. Damit werden Elektrolyseöfen mit ähnlichem Verhalten gruppiert, um eine leichtere Analyse zu ermöglichen. Auch in [HV12] werden Aluminiumelektrolyseöfen mit ähnlichem Verhalten in Gruppen eingeteilt, wobei das hierarchische Clustern eingesetzt wird.

Mittels Berechnung von 20 Zeitreihenmerkmalen und der Principal Component Analysis (PCA)¹ werden in unserer Arbeit [GKD+23] nicht plausible Temperaturverläufe von Thermoelementen identifiziert, die an Aluminiumelektrolyseöfen der TAE platziert sind. Mehrere zweidimensionale Merkmalsräume werden dazu erstellt, um diese hinsichtlich auffälliger Temperaturverläufe zu untersuchen. Außerdem wird ein interaktives Dashboard gezeigt, um einen praktischen Einsatz der vorgestellten Lösung in der industriellen Aluminiumherstellung zu ermöglichen.

In [CSO16] wird ein neuronales Netz implementiert, um chemische Prozessvariablen und die Badtemperatur vorherzusagen. Das neuronale Netz wird mit weiteren linearen und nicht linearen Modellen bezüglich der Vorhersagefehler verglichen. Anschließend wird das neuronale Netz für eine Simulation eingesetzt, um den Einfluss der Aluminiumfluoridzugabe auf die Badtemperatur und die Aluminiumfluoridkonzentration zu untersuchen. Dennoch fehlt ein Vergleich mit einem Pseudo-Modell, um die Vorhersageperformanz des neuronalen Netzes einschätzen zu können. Eine Analyse der Modellzusammenhänge sowie eine Interpretation der Badtemperaturvorhersagen ist zudem nicht vorhanden, die im Bereich der Aluminiumelektrolyse jedoch von Vorteil sein können, wie in unserer Veröffentlichung [GKD+18] dargelegt wird.

In [JGJ21] wird für die Vorhersage der Badtemperatur ein Modell in MATLAB® und Simulink® verwendet, mit dem die gegenwärtige und die zukünftige Badtemperatur eines Ofens unter den Bedingungen einer Strommodulation prognostiziert wird. Mit einem Optimierungsalgorithmus wird zunächst das Modell an den aktuellen Zustände eines Ofens angepasst, für den eine Badtemperaturvorhersage erzeugt werden soll. Anschließend werden mit dem Modell gegenwärtige und zukünftige Vorhersagen der Badtemperatur erstellt. Die Abweichung zwischen der gegenwärtigen Vorhersage und der gemessenen Badtemperatur wird mit weniger als 2 °C

¹Die PCA ist ein unüberwachtes lineares Transformationsverfahren, um die Merkmale eines Datensatzes in einen kleineren Merkmalsraum zu projizieren. Sie wird genutzt, um Muster in Daten zu identifizieren. Hierzu werden die Merkmale eines Datensatzes durch eine Linearkombination ersetzt. Ziel ist es, eine Linearkombination zu finden, die zu einer Maximierung der Varianz führt [Fro18, S. 271; RM17, S. 142; HWL15, S. 1617; HK17, S. 124–125, 130].

angegeben. Als Nachteil der präsentierten Lösung wird die lange Simulationszeit von bis zu 40 Minuten erwähnt, die bis zum Vorliegen einer Vorhersage vergeht.

In [WBSK+24] wird ein dynamisches Modell vorgestellt, mit dem unter anderem die Badtemperatur unter dem Einfluss einer Strommodulation kontinuierlich vorhergesagt wird. Hierbei werden die Daten von Thermoelementen für die Vorhersage berücksichtigt, die an der Ofenwand und in der Ofenverkleidung angebracht sind. Die vorhergesagte Badtemperatur wird mit regelmäßigen Badtemperaturmessungen an verschiedenen Stellen im Ofen verglichen. Es wird gezeigt, dass die kontinuierliche Vorhersage den generellen Trend der gemessenen Badtemperaturen nachbildet. Die Verwendung dieses dynamischen Modells erfordert jedoch den Einsatz von Thermoelementen an den Öfen, die aufgrund der Produktionsumgebung regelmäßig gewartet werden müssen, wie in unserer Arbeit [GKD+23] gezeigt wird.

Ein Multilayer Perceptron (MLP) wird in [SO10] als Soft-Sensor eingesetzt, um die Badtemperatur von Aluminiumelektrolyseöfen zu bestimmen. Es wird herausgestellt, dass das dort präsentierte Modell Temperaturmessungen nicht vollständig ersetzen kann. Im Bereich der Aluminiumelektrolyse wird in [FK00a] eine Vorgehensweise basierend auf der Predictive (Feature) Importance vorgestellt, um wesentliche Merkmale eines neuronalen Netzes identifizieren zu können. In [FK00b] werden ein neuronales Netz zur Vorhersage der Badtemperatur verwendet und die Merkmale mithilfe der Predictive (Feature) Importance untersucht. In der vorliegenden Arbeit wird eine vergleichbare Vorgehensweise gewählt, um wesentliche Merkmale für die Gegenwartsvorhersage zu identifizieren. In [SSC+19] wird ein Lösungsansatz vorgestellt, bei dem die Badtemperatur, der Aluminiumfluoridgehalt im Elektrolyt und die Höhe des flüssigen Aluminiums im Ofen mithilfe von neuronalen Netzen vorhergesagt wird. Dazu werden die Aluminiumelektrolyseöfen zunächst in Gruppen eingeteilt, für die anschließend jeweils ein neuronales Netz trainiert wird. Die Autor/-innen begründen ihre Vorgehensweise damit, dass ein einzelnes neuronales Netz nicht das Verhalten aller Öfen abbilden könne. Auf der anderen Seite sei jeweils ein neuronales Netz für jeden einzelnen Ofen zu aufwendig in der Anwendung [SSC+19, S. 2–3]. Entsprechende Ergebnisse für diese Aussagen werden in der Arbeit jedoch nicht aufgezeigt.

In den erläuterten Arbeiten [SO10], [FK00b] und [SSC+19] fehlt jeweils ein Vergleich mit einem Pseudo-Modell, um die Performanz der Vorhersagemodelle ins Verhältnis setzen zu können. Der Nutzen der verwendeten Modelle wird dadurch nicht ersichtlich. Darüber hinaus werden die erstellten Vorhersagen nicht weiter interpretiert. Lediglich die erreichten Performanzen der verwendeten Modelle werden in den Arbeiten dargestellt. Zudem weisen neuronale Netze Eigenschaften auf, die eine Praxisimplementierung erschweren können. In Kapitel 4.2.1 wird darauf genauer Bezug genommen.

In unserer Veröffentlichung [GKD+18] wird ein RF-Regressionsmodell implementiert, um eine Gegenwartsvorhersage für die Badtemperatur von Aluminiumelektrolyseöfen der TAE zu

erstellen. Mithilfe einer Permutation Feature Importance (PFI) werden Prozessvariablen analysiert, die einen wesentlichen Einfluss auf die Vorhersage haben. Darüber hinaus wird der Modellzusammenhang zwischen den Prozessvariablen und der Badtemperatur durch die Verwendung von Partial Dependence Plots analysiert. Ein Vergleich mit einem Pseudo-Modell wird jedoch nicht durchgeführt. In [SBMW22, S. 375–378] werden weitere Vorhersagemodelle aus dem Bereich des maschinellen Lernens sowie ein Pseudo-Modell (naives Modell) für eine Gegenwartsvorhersage der Badtemperatur eingesetzt. Dabei wird innerhalb der Datenvorverarbeitung die Aggregation von Zeitreihendaten aus unserer Arbeit [GKD+18] aufgegriffen. Das Gradient Boosting Modell erreicht auf den Testdaten die beste Performanz [SBMW22, S. 377]. Eine Interpretation der erstellten Gegenwartsvorhersagen wird jedoch nicht aufgezeigt.

In [MAW01] wird eine Zeitreihenanalyse durchgeführt und ein ARMAX-Zeitreihenmodell verwendet, um die Badtemperatur von Aluminiumelektrolyseöfen vorherzusagen. Das ARMAX-Zeitreihenmodell wird als generisches Modell bezeichnet, da es mit Daten von insgesamt 13 Aluminiumelektrolyseöfen trainiert und validiert wird. Als Begründung für diese Vorgehensweise wird in der Arbeit angeführt, dass dieses ARMAX-Zeitreihenmodell eine bessere Performanz auf den Validierungsdaten erreicht als ein Modell mit Daten eines einzelnen Ofens. Die Arbeit präsentiert jedoch keine empirischen Ergebnisse für diese Erkenntnis.

Im Rahmen der vorliegenden Arbeit werden zwei Lösungsansätze analysiert, die die Erstellung einer 1-Schritt-Badtemperaturvorhersage (Forecasting) sowie einer Gegenwartsvorhersage (Nowcasting) der Badtemperatur ermöglichen. Zunächst wird eine umfassende Analyse der Badtemperatur durchgeführt, die die Ergebnisse aus [MAW01] ergänzen. Für die 1-Schritt-Badtemperaturvorhersage erfolgt ein Vergleich zwischen einem globalen AR-Modell und lokalen klassischen Zeitreihenmodellen. Unter den lokalen Modellen befinden sich zudem zwei Pseudo-Modelle. Das globale AR-Modell dient als Basismodell und wird ähnlich wie in [MAW01] mit allen Badtemperaturdaten trainiert, während die lokalen Zeitreihenmodelle lediglich mit den Daten eines Ofens trainiert werden. Im Anschluss wird die Performanz aller Vorhersagemodelle gegenübergestellt, um empirisch den Nutzen eines globalen AR-Modells für die Badtemperaturvorhersage aufzuzeigen.

Der zweite Lösungsansatz wird basierend auf den Ergebnissen aus unserer Arbeit [GKD+18] mit aktuelleren Daten umgesetzt. Hierbei werden ein RF-Regressionsmodell für eine Gegenwartsvorhersage der Badtemperatur eingesetzt und die Performanz mit den Modellen aus dem ersten Lösungsansatz verglichen. Außerdem wird eine Möglichkeit für die visuelle Darstellung und Interpretation der Vorhersagen des RF-Regressionsmodells für den praktischen Einsatz vorgestellt. Eine derartige Interpretation wurde bislang im Literaturbereich der Aluminiumelektrolyse nach bestem Wissen noch nicht aufgezeigt.



KAPITEL 3

Grundlagen der Aluminiumelektrolyse

In diesem Kapitel werden die Grundlagen der industriellen Aluminiumherstellung vorgestellt. Zunächst wird auf den Hall-Héroult-Prozess und die wesentlichen Bestandteile eines Aluminiumelektrolyseofens eingegangen. Bei der industriellen Herstellung von Aluminium stellt die Badtemperatur eine wesentliche Prozessvariable dar, die in diesem Kapitel beschrieben wird. In diesem Zusammenhang werden zudem die Wechselwirkungen der unterschiedlichen Prozessvariablen innerhalb der Aluminiumelektrolyse erläutert.

3.1 Hall-Héroult-Prozess

Für die industrielle Herstellung von Aluminium wird der Hall-Héroult-Prozess genutzt, bei dem eine elektrolytische Reduktion von Aluminiumoxid (Al_2O_3) durchgeführt wird [GK93, S. 1; Dü16, S. 15]. Aluminiumoxid besitzt eine hohe Schmelztemperatur von knapp über 2000°C [Wib08, S. 1140]. Aus diesem Grund wird Aluminiumoxid in flüssigem Kryolith (Na_3AlF_6) gelöst und mit Aluminiumfluorid (AlF_3) sowie Calciumfluorid (CaF_2) vermischt, um die Schmelztemperatur in einen Bereich von 920°C bis 970°C zu senken [Hau16a, S. 4; Wib08, S. 1140; GK93, S. 50]. Die Schmelztemperatur wird auch Liquidustemperatur genannt und ist abhängig von der chemischen Zusammensetzung des Elektrolyten [Ris12, S. 17; RIW+16, S. 818]. Die Liquidustemperatur ist die Mindesttemperatur, bei der der Prozess betrieben werden muss, ohne dass der Elektrolyt zu erstarren beginnt [Hau16c, S. 804].

Die Badtemperatur ist die Temperatur des Elektrolyten und liegt üblicherweise 5°C bis 10°C oberhalb der Liquidustemperatur [Ree15, S. 8]. Die Differenz zwischen der Badtemperatur und der Liquidustemperatur wird Superheat genannt. Der Superheat ist ein entscheidender Faktor für die Lösungsfähigkeit von Aluminiumoxid dar und steht zudem im Zusammenhang mit der Stärke der Seitenkruste, die aus erstarrtem Elektrolyt besteht und den Ofen vor Beschädigungen schützt [Dü16, S. 10, 18; RIW+16, S. 818; GB19, S. 2137].

In Abbildung 3.1 ist eine Querschnittansicht eines Ofens zu sehen. Diese Art des Ofens kommt auch bei der TAE zum Einsatz. Die Anoden, die in den flüssigen Elektrolyt eintauchen, bestehen

aus Petrolkoks (Kohlenstoff, C) und Steinkohlenteerpech als Bindemittel [GK93, S. 4]. Die elektrolytische Reaktion mit dem gelösten Aluminiumoxid führt zur Bildung von flüssigem Aluminium [GK93, S. 1], das sich auf dem Kathodenblock absetzt. Des Weiteren entsteht durch die Reaktion mit dem Sauerstoff aus dem Aluminiumoxid und dem Kohlenstoff der Anoden gasförmiges Kohlenstoffdioxid (CO₂) [GK93, S. 1; Dü16, S. 26]. Diese chemische Reaktion lässt sich mit Gleichung 3.1 abbilden. Die Anoden brauchen sich auf, sodass sie nachgeführt werden müssen [GK93, S. 208]. Alle 22 bis 26 Tage werden die aufgebrauchten Anoden durch neue Anoden ersetzt, um eine kontinuierliche Produktion an Aluminium zu ermöglichen [GK93, S. 2, 5]. Das flüssige Aluminium im Ofen wird in der Regel alle 24 bis 48 Stunden abgesaugt [GK93, S. 213]. Um einen Luftabbrand der Anoden zu vermeiden und eine thermische Isolation zu erzeugen, werden neu eingesetzte Anoden mit Aluminiumoxid oder zerkleinertem Elektrolyt eingedeckt [GK93, S. 202].

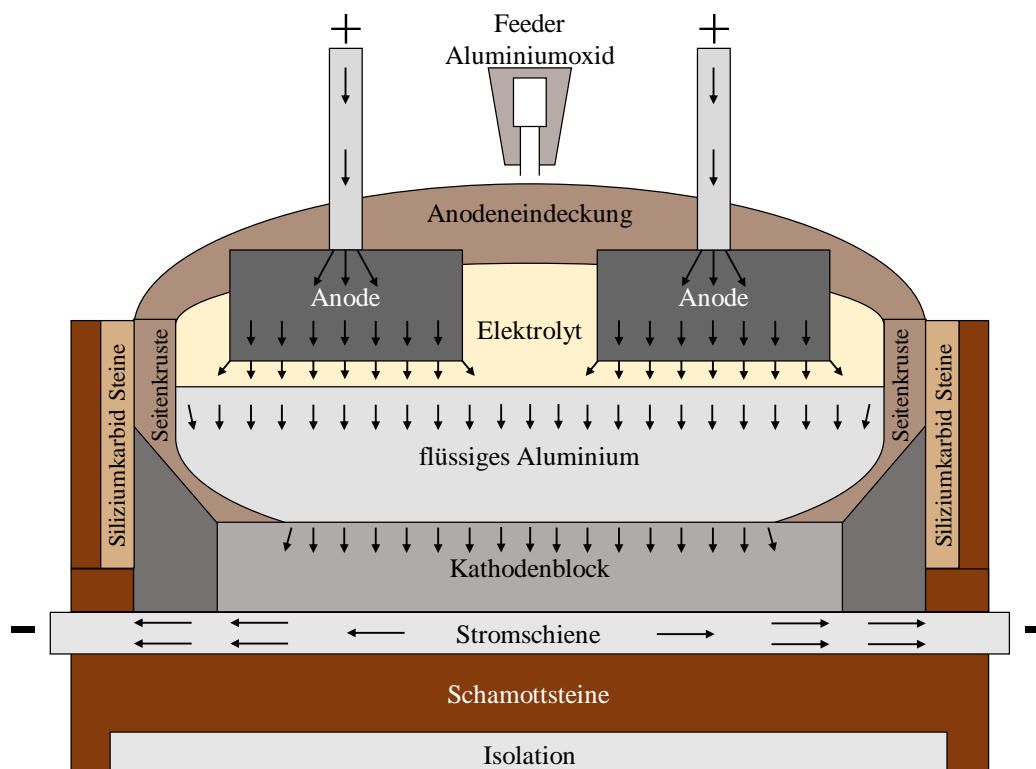
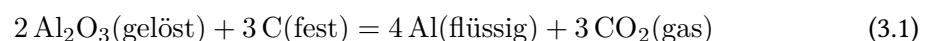


Abbildung 3.1: Querschnittsansicht eines Aluminiumelektrolyseofens. Abgeändert nach [GB19, S. 2137].



Aus 1,89 kg Aluminiumoxid kann theoretisch eine Menge von 1 kg Aluminium (Al) gewonnen werden [GK93, S. 3]. Die theoretische Produktionsmenge an Aluminium ist bei einem

konstanten Wirkungsgrad proportional zur Stromstärke [Dü16, S. 4; GK93, S. 148]. Mit einer Stromstärke von 175 kA kann in einer Zeit von 24 Stunden eine Menge von ungefähr 1409 kg Aluminium produziert werden [GK93, S. 148]. Eine Erhöhung der Stromstärke erhöht zwar die Produktionsmenge an Aluminium, wirkt sich aber auch auf die sensible Energiebilanz eines Ofens aus [Dü16, S. 4].

Das Hinzufügen von Aluminiumoxid wird als Fütterung bezeichnet. Dazu wird die obere Kruste aufgebrochen, sodass durch die Öffnung Aluminiumoxid dem flüssigen Elektrolyt hinzugefügt werden kann [GK93, S. 79–80]. Für das Aufbrechen der Kruste sind in den Öfen der TAE Krustentreiber installiert, die ein Loch in die obere Kruste brechen. Das Aluminiumoxid befindet sich in einem Behälter, sodass das Aluminiumoxid durch die geöffnete Kruste in Schüben in den flüssigen Elektrolyt gelangt. Diese Technik der Fütterung wird als “Point Feeding” [GK93, S. 80–81] bezeichnet und erfolgt bei der TAE vollständig automatisiert.

Die Zugabe von Aluminiumoxid wird anhand der Ofenspannung bzw. eines berechneten Ofenwiderstandes (Pseudo-Widerstand) durchgeführt. In Abbildung 3.2 ist der Zusammenhang zwischen der Aluminiumoxidkonzentration des Elektrolyten und der Ofenspannung bei unterschiedlichen Anode-Kathode-Abständen (Anode-Cathode Distance (ACD)) zu sehen. Üblicherweise wird der Ofenwiderstand für die Ofenregelung verwendet [KMSS13, S. 760]. Der Ofenwiderstand hat gegenüber der Ofenspannung den Vorteil, dass dieser sich bei geringen Stromstärkeänderungen nicht ändert und damit ein stabileres Eingangssignal für die Ofenregelung darstellt [KMSS13, S. 760]. Die Berechnung des Ofenwiderstandes kann über die in [KMSS13, S. 760; GK93, S. 221] angegebene Gleichung erfolgen.

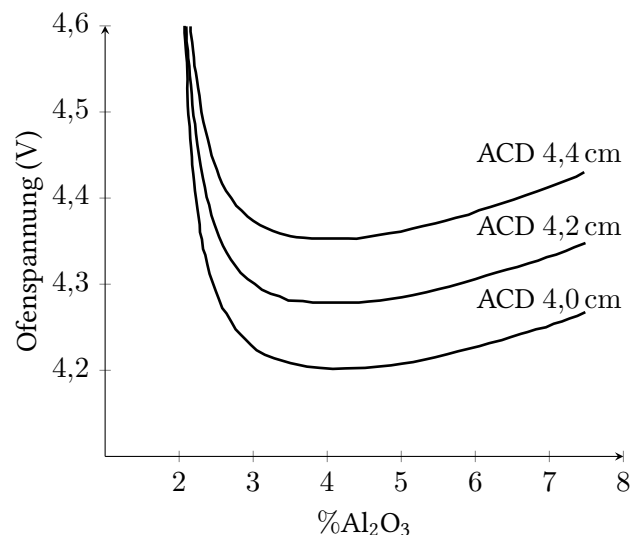


Abbildung 3.2: Zusammenhang zwischen der Aluminiumoxidkonzentration und der Ofenspannung bei unterschiedlichen Anode-Kathode-Abständen (ACD). Abgeändert nach [Hau16b, S. 158].

Das Ziel der Ofenregelung ist es unter anderem, die Aluminiumoxidkonzentration des Elektrolyten zu überwachen und anzupassen. Der ideale Wert der Aluminiumoxidkonzentration liegt nach [Dü16, S. 18] bei 3 %. Aus Abbildung 3.2 geht hervor, dass bei einer zu geringen Aluminiumoxidkonzentration der Ofenwiderstand rapide ansteigt. Dieses Verhalten wird als Anodeneffekt bezeichnet, bei dem klimaschädliche Gase freigesetzt werden (vgl. Kapitel 3.2). Dagegen steigt der Ofenwiderstand bei einer hohen Aluminiumoxidkonzentration aufgrund der Widerstandserhöhung des Elektrolyten an [Dü16, S. 21].

Die Badtemperatur eines Aluminiumelektrolyseofens ist eine wichtige Prozessvariable [GK93, S. 215], die bei der TAE alle 24 bis 48 Stunden (Stand 08/2022) manuell gemessen und so angepasst wird, dass energieeffizient produziert, Aluminiumoxid optimal gelöst und eine schützende Randkruste erzeugt wird [Dü16, S. 42]. Die Badtemperatur wird zudem als Maß für die Stabilität des Ofenbetriebs und als Warnsignal für Ofenabnormalitäten verwendet [GK93, S. 215]. Eine niedrige Badtemperatur kann zudem zu einer höheren Stromausbeute und einem niedrigen Energieverbrauch beitragen, wobei eine bestimmte Differenz zwischen der Badtemperatur und Liquidustemperatur bestehen sollte, die für eine optimale Lösung von Aluminiumoxid im Elektrolyt und eine schützende Randkrustenbildung sorgt [Mad92, S. 453; GK93, S. 26; Dü16, S. 42]. Eine konstante Badtemperatur ist in der Praxis nur schwierig möglich, da unterschiedliche Faktoren die Badtemperatur eines Ofens beeinflussen, die in Kapitel 3.2 aufgezeigt werden. Nach [TCY13, S. 58] schwankt die Badtemperatur in den besten Aluminiumhütten der Welt mit einer Standardabweichung von 5 °C bis 6 °C.

3.2 Elektrolyttemperatur (Badtemperatur)

Während des Ofenbetriebs kommt es zu unterschiedlich starken Schwankungen der Elektrolyttemperatur (Badtemperatur), die von verschiedenen Einflussfaktoren bedingt sind. Abbildung 3.3 zeigt die Wechselwirkungen zwischen den Prozessvariablen eines Ofens, wie Badtemperatur, Superheat, Badchemie und Seitenkruste. Die dargestellten Wechselwirkungen sind dabei nur gültig, wenn eine Seitenkruste im Ofen vorhanden ist [STC+08, S. 309]. Eine Änderung der Ofenspannung hat einen Einfluss auf die Energiebilanz eines Ofens, die sich über die ändernde Wärmeverluste über die Seitenwand letztendlich selbst beeinflusst. In der Literatur lassen sich Versuchsauswertungen finden, die die Auswirkungen von unterschiedlichen Prozessparametern, Eingriffen und Störungen auf die Badtemperatur vermitteln.

Im Folgenden werden die wesentlichen Einflussfaktoren auf die Badtemperatur mithilfe der Literatur aufgezeigt, die für die Entwicklung eines Modells für die Vorhersage der Badtemperatur relevant sein könnten. Des Weiteren wird ein Versuch präsentiert, dessen Auswertung im Rahmen dieser Arbeit erfolgte und den Einfluss einer kontrollierten Hallenschaltung auf die Badtemperatur aufzeigt.

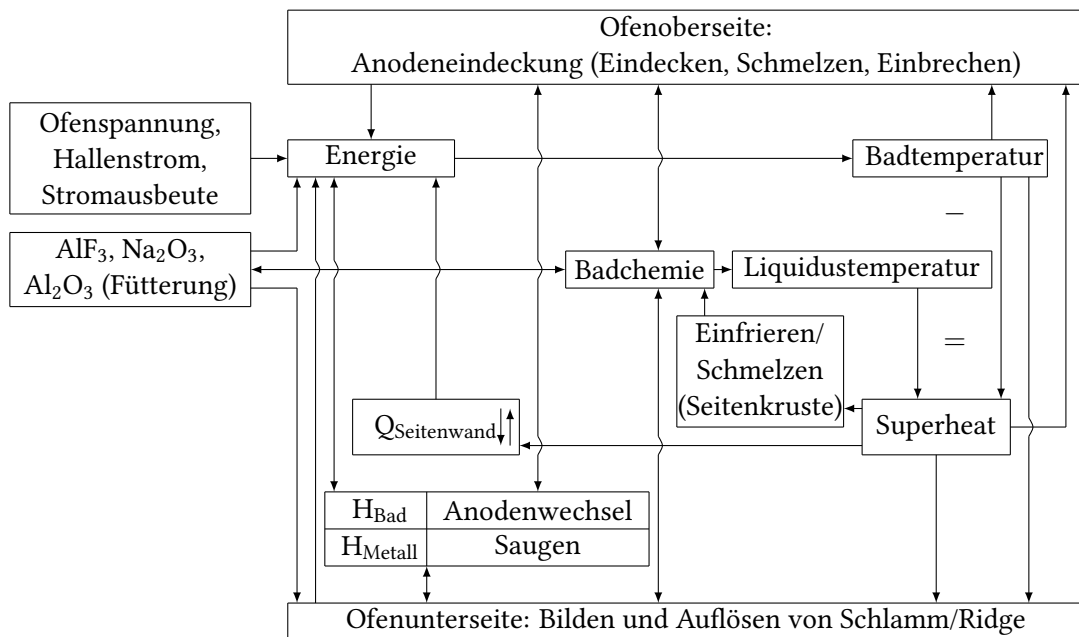


Abbildung 3.3: Wechselwirkungen zwischen den einzelnen Prozessvariablen eines Aluminiumelektrolyseofens. Abgeändert nach [STC+09, S. 312].

Ofenspannung

Für die Regelung der Badtemperatur setzt die TAE die 9-Box-Matrix-Regelung ein, die das primäre Ziel verfolgt, die Badtemperatur mittels Anpassungen der Ofenspannung in einem vorgegebenen Kontrollband zu halten [RIW+16, S. 817]. Rieck et al. stellten bei einer Änderung der Ofenspannung um 100 mV (–100 mV) für 24 Stunden eine Badtemperaturänderung um etwa 7 °C (–7 °C) fest [RIW+16, S. 820]. Die Ofenspannung ist folglich eine essentielle Größe in der Aluminiumelektrolyse, die unmittelbar die Stabilität des Ofenbetriebs beeinflussen kann. Die ohmschen Spannungsabfälle sind bis auf die Spannungsverluste über die Stromschienen maßgeblich für die Wärmeentwicklung eines Ofens [Dü16, S. 20; HBGH03, S. 269]. Dabei trägt der ohmsche Spannungsabfall über den Elektrolyt wesentlich zur Wärmeerzeugung im Ofen bei. Der ohmsche Spannungsabfall über den Elektrolyt liegt nach [GK93, S. 22–23] zwischen 1,3 und 2,0 V.

Abbildung 3.4 zeigt den Einfluss einer Erhöhung der Ofenspannung auf die Wärme- und Materialbilanz eines Ofens. Mit der Erhöhung des Anode-Kathode-Abstands (ACD) erhöht sich die Ofenspannung, die zu einer Erhöhung der Badtemperatur und des Superheats führt. Das führt anschließend zum Schmelzen der Seitenkruste, wodurch weiterer Kryolith in das Bad gelangt, zu einer Verdünnung von Aluminiumfluorid (AlF₃) führt und einen Anstieg der Liquidustemperatur verursacht. Das Aufschmelzen der Seitenkruste bewirkt außerdem, dass sich ein zusätzlicher

Wärmeverlust des Ofens einstellt. Insgesamt stellt sich durch eine Spannungserhöhung ein neue Wärme- und Materialbilanz des Ofens ein.

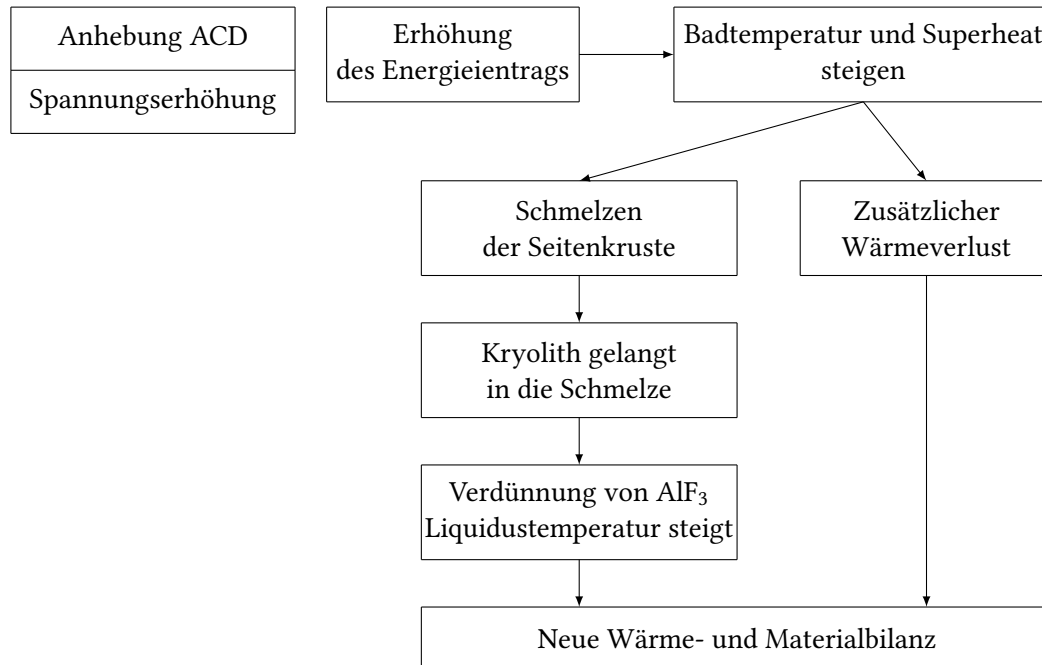


Abbildung 3.4: Mit der Erhöhung des Energieeintrags stellt sich eine neue Wärme- und Materialbilanz eines Ofens ein. Abgeändert nach [RIW+16, S. 818].

Hallenstrom

Eine Veränderung der Stromstärke verursacht eine Änderung der Wärmeentwicklung im Ofen, die die TAE durch eine Anpassung der Ofenspannung bis zu einem gewissen Grad kompensiert [Dü16, S. 141]. Eine höhere Wärmeentwicklung im Ofen, die nicht mehr von der Verringerung der Ofenspannung kompensiert werden kann, wird unter anderem mittels an den Ofenwänden angebrachter Wärmetauscher kompensiert [Dü16, S. 4, 10, 141].

Als Hallenschaltung wird bei der TAE eine kontrollierte Stromabschaltung einer gesamten Halle für einen vorgegebenen Zeitraum bezeichnet, um bestimmte Arbeiten an Öfen durchführen zu können. Vor allem bei der Installation der Magnetfeldkompensation und Wärmetauscher in Halle 1 im Jahr 2018 kam es zu einer hohen Anzahl an Hallenschaltungen, die die Wärmebilanz der Öfen beeinflussten [DMB19]. Die Funktion der Magnetfeldkompensation und der Wärmetauscher wurde bereits in Kapitel 1.2 erläutert.

Eine durchgeführte Simulation in [DMB19, S. 537–539] zeigt, dass eine Hallenschaltung für 1,5 Stunden eine Badtemperaturänderung von etwa -30 °C verursacht. Die Autoren weisen jedoch darauf hin, dass die ermittelten Badtemperaturänderungen im Vergleich zum realen Prozess zu

Abschnitt 3.2. Elektrolyttemperatur (Badtemperatur)

hoch sind, da bestimmte Effekte, die die Wärmeverluste beeinflussen, in der Simulation nicht berücksichtigt wurden.

Am 15.03.2021 wurde bei der TAE in Halle 1 gegen 12:00 Uhr eine Hallenschaltung mit einer Dauer von einer Stunde durchgeführt, die anschließend im Rahmen dieser Arbeit ausgewertet wurde. Vor, während und nach der Hallenschaltung wurden die Badtemperaturen der Öfen 1099, 1103, 1107 und 1109 zu bestimmten Messzeitpunkten direkt nacheinander zweimal gemessen, sodass für jeden Zeitpunkt ein Messwertpaar vorhanden ist. Die Messwertpaare sind für jeden Ofen in Abbildung 3.5 zeitlich aufgetragen. Des Weiteren ist die gemittelte Badtemperatur, die sich aus den jeweiligen Messwertpaaren ergibt, für jeden Zeitpunkt angegeben. Die Standardabweichung der Differenzen, die jeweils aus den Messwertpaaren berechnet wird, beträgt 2,5 °C. Die Messung um 12:45 Uhr am Ofen 1107 ist ein Ausreißer. Zu diesem Zeitpunkt wurde eine nicht plausible Badtemperatur gemessen, die in Abbildung 3.5 als Ausreißer markiert ist und nicht in die hier durchgeführten Berechnungen eingeflossen ist. Gegen 13:00 Uhr wurde der Hallenstrom wieder auf das ursprüngliche Niveau angehoben.

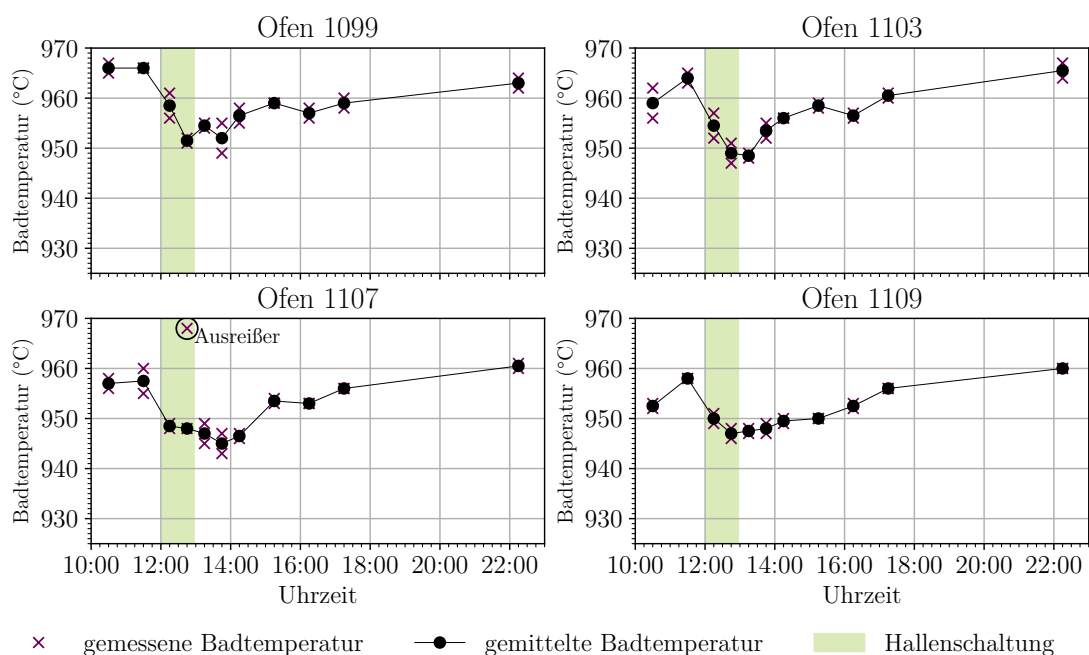


Abbildung 3.5: Zeitlicher Badtemperaturverlauf der Öfen 1099, 1103, 1107 und 1109 vor, während und nach einer Hallenschaltung, die am 15.03.2021 gegen 12 Uhr stattgefunden hat. Für eine bessere Lesbarkeit der Verläufe sind die gemittelten Badtemperaturen durch Linien miteinander verbunden.

Aus Abbildung 3.5 geht hervor, dass die Badtemperatur für jeden Ofen während der Hallenschaltung unterschiedlich stark abfällt. Um die Unterschiede der einzelnen Öfen aufzuzeigen, sind die Differenzen der gemittelten Badtemperatur zwischen der letzten Messung vor der Hallenschaltung und der letzten Messung während der Hallenschaltung in Tabelle 3.1 zu finden.

Abschnitt 3.2. Elektrolyttemperatur (Badtemperatur)

Die mittlere Badtemperaturänderung liegt für die einstündige Hallenschaltung bei $-12,5\text{ °C}$ mit einer Standardabweichung von $2,7\text{ °C}$. Auch das Verhalten der Badtemperatur nach der Hallenschaltung ist unterschiedlich. Für den Ofen 1109 ist ein langsamer Anstieg der Badtemperatur direkt nach der Hallenschaltung erkennbar. Bei den übrigen Öfen setzte der Anstieg der Badtemperaturen erst nach einer Verzögerung von ungefähr einer Stunde nach der Hallenschaltung ein.

Ofen	Differenz (°C)
1099	-14,5
1103	-15,0
1107	-9,5
1109	-11,0
Mittelwert und Standardabweichung	$-12,5 \pm 2,7$

Tabelle 3.1: Badtemperaturdifferenz für jeden Ofen zwischen der letzten Messung vor der Hallenschaltung und der letzten Messung während der einstündigen Hallenschaltung am 15.03.2021. Die mittlere Differenz, die aus den einzelnen Differenzen berechnet wurde, liegt bei $-12,5\text{ °C}$ mit einer Standardabweichung von $2,7\text{ °C}$.

Anodenwechsel

Das thermische Verhalten eines Ofens wird maßgeblich von einem Anodenwechsel beeinflusst [Dü16, S. 44]. Bei einem Anodenwechsel kann die Badtemperatur am Messloch um bis zu 30 °C fallen [Mad92, S. 454]. Zudem kann es bis zu zwölf Stunden dauern, bis sich die Badtemperatur auf das Temperaturniveau vor dem Anodenwechsel zurückbewegt hat [Mad92, S. 454; RIW+16, S. 819].

Eine Anode verweilt bei der TAE in der Regel 28 Tage bis 32 Tage im Ofen [TRI13, S. 58–59] bis der Stumpf der abgebrannten Anode innerhalb eines Anodenwechsels aus dem Ofen gezogen wird. Alle zwei Tage findet bei der TAE an einem Ofen ein Anodenwechsel statt. Bei einem Anodenwechsel werden bis zu zwei Anoden eines Ofens gewechselt [Dü16, S. 44]. Die neuen Anoden werden vor dem Einsetzen nicht vorgeheizt und weisen daher eine Temperatur auf, die der Umgebungstemperatur entspricht [Dü16, S. 44]. Unmittelbar nach dem Einsetzen der neuen Anode erstarrt der Elektrolyt um die Anode, was zu einer ungleichmäßigen Stromverteilung im Ofen führt [Dü16, S. 44].

Metallsaugen

Das in den Öfen produzierte Aluminium wird bei der TAE in der Regel alle zwei Tage abgesaugt [TRI13, S. 49]. Dabei kann sich die Badtemperatur im Saugloch für einen Zeitraum von bis zu zehn Minuten reduzieren [Mad92, S. 454]. Nach dem Saugen besteht aufgrund des geringeren Metallstands eine kleinere Kontaktfläche zwischen dem flüssigen Aluminium und der Seiten-

wand des Ofens, wodurch sich dieser über die Zeit aufheizt [Dü16, S. 44–45]. Das Saugen von Metall hat einen Einfluss auf das thermische Verhalten eines Ofens.

Anhand von Metallstandmessungen wird die Menge von flüssigem Metall bestimmt, die aus einem Ofen abgesaugt werden muss, um einen optimalen Betriebszustand zu erhalten. Dabei wird ein Eisenstab für einige Sekunden in das flüssige Aluminium getaucht. Anschließend wird dieser wieder entnommen, sodass sich eine dünne Schicht an Bad und Aluminium auf dem Eisenstab bildet. Anhand der Länge der Metallschicht, die sich entlang des Eisenstabs gebildet hat, kann der Metallstand eines Ofens ermittelt werden. Der Metallstand sollte nicht zu niedrig oder zu hoch sein, um einen stabilen Betrieb zu gewährleisten. In der Literatur wird ein Sollwert für den Metallstand von 20 cm angegeben [GK93, S. 213].

Flusssaugen

Das Saugen von Fluss (Elektrolyt) wird unter anderem durchgeführt, wenn Öfen einen zu niedrigen Flusstand besitzen. Dabei wird Fluss von einem Ofen abgesaugt und einem Ofen mit niedrigem Flusstand hinzugefügt [GK93, S. 214]. Die Kontrolle des Flusstands erfolgt analog zur Kontrolle des Metallstands. Wie bei einem zu niedrigen Metallstand heizt sich ein Ofen bei einem zu niedrigen Flusstand auf. Demgegenüber führt ein zu hoher Flusstand zu einer Abkühlung des Ofens. Der Flusstand hat daher einen Einfluss auf das thermische Verhalten eines Ofens und sollte sich stets im optimalen Bereich befinden, um Probleme im Ofenbetrieb zu vermeiden [TRI13, S. 46–47]. Nach [GK93, S. 214] ist ein Flusstand von 20 cm ein typischer Sollwert, der in der Aluminiumelektrolyse verwendet wird.

Zugabe von Natriumcarbonat (Na_2CO_3)

Mittels Zugabe von Natriumcarbonat (Na_2CO_3), auch Soda genannt, kann der Aluminiumfluoridgehalt eines Ofens beeinflusst werden. Eine Zugabe erfolgt zum Beispiel, wenn der Aluminiumfluoridgehalt eines Ofens aufgrund einer falschen Dosierung zu hoch ist [WTBW11, S. 491; GSWSK12, S. 930]. Die Zugabe von Natriumcarbonat in einen Ofen bewirkt eine Erhöhung der Liquidustemperatur und eine Reduzierung des Superheats. Die Reduzierung des Superheats führt zu einer Verringerung der Wärmeverluste über die Seitenwand und die obere Seitenkruste des Ofens wird dicker, sodass die Badtemperatur des Ofens ansteigt [Kva15, S. 32].

Zugabe von Aluminiumoxid (Al_2O_3)

Die Zugabe von Aluminiumoxid erfolgt bei der TAE anhand der kontinuierlichen Überwachung der Ofenspannung [Dü16, S. 10, 43]. In [Mad92, S. 454] wird von einer Reduzierung der Badtemperatur von bis zu 12 °C nach der Zugabe von Aluminiumoxid berichtet. Auch in [RIW+16, S. 819] wird eine Verringerung der Badtemperatur beobachtet, nachdem kaltes Aluminiumoxid in den Ofen gegeben wurde. Des Weiteren zeigen die Autoren in [RIW+16, S. 819] auf, dass die

Badtemperatur innerhalb eines Fütterungszyklus Schwankungen von bis zu 5 °C unterliegen kann.

Zugabe von Aluminiumfluorid (AlF_3)

Nach Abbildung 3.6 führt eine Erhöhung der Zugabe an Aluminiumfluorid (AlF_3) zunächst zu einer Verringerung der Liquidustemperatur [RIW+16, S. 818]. Die Änderung der Liquidustemperatur bewirkt eine Änderung des Superheats, der den Wärmestrom durch die Seitenwände eines Ofens aufgrund der schmelzenden Seitenkruste beeinflusst. Die Änderung des Wärmestroms über die Seitenwand hat einen unmittelbaren Einfluss auf die Wärmebilanz des Ofens.

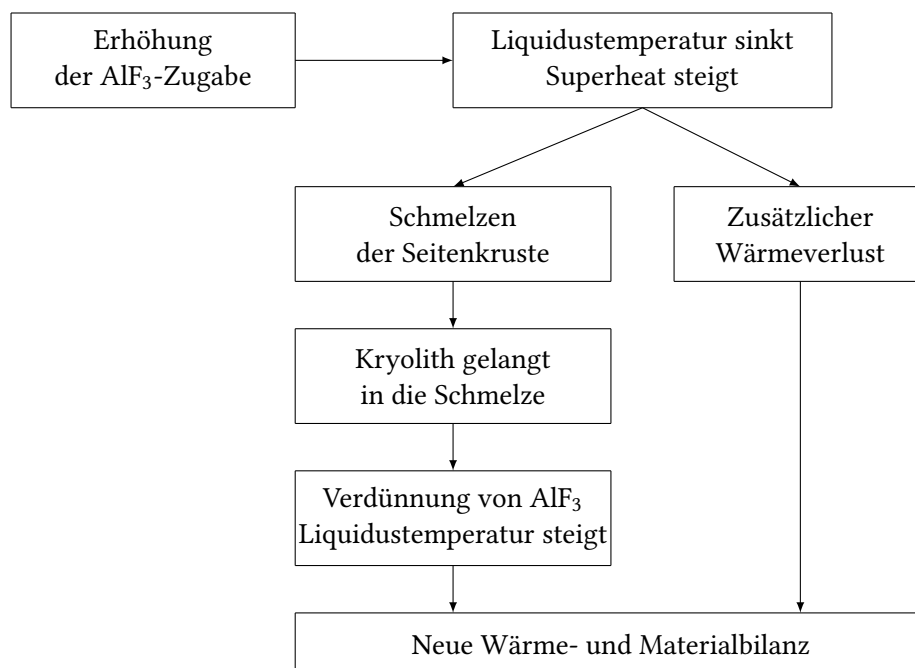


Abbildung 3.6: Mit der Erhöhung der Zugabe an Aluminiumfluorid (AlF_3) stellt sich eine neue Wärme- und Materialbilanz des Ofens ein. Abgeändert nach [RIW+16, S. 818].

Rieck et al. stellten in einem Versuch fest, dass die doppelte Menge an Aluminiumfluorid eine Verringerung der Liquidustemperatur um 3 °C pro Tag bewirkte, während keine Zugabe eine Erhöhung der Liquidustemperatur um 4 °C pro Tag zur Folge hatte [RIW+16, S. 820]. Die Badtemperatur folgte in dem Versuch langsam der Liquidustemperatur.

Anodeneffekte

Anodeneffekte resultieren aus einer zu geringen Konzentration an gelöstem Aluminiumoxid im Elektrolyt, sodass die Ofenspannung kurzfristig auf bis zu 80 V ansteigen kann [KGD+20,

S. 1; GK93, S. 210; TUV16, S. 131]. Aufgrund der geringen Konzentration an Aluminiumoxid verschlechtert sich das Benetzen der Anode mit dem Elektrolyt, da sich Gasblasen unter der Anode bilden und den elektrischen Widerstand des Ofens erhöhen [GK93, S. 210]. Eine solche Widerstandserhöhung resultiert in einen rapiden Anstieg der Ofenspannung, wobei klimaschädliche Gase (Perfluorcarbone) freigesetzt werden [GK93, S. 201; WTL14, S. 529]. Der erhöhte Energieeintrag innerhalb des Ofens kann zu einem Aufheizen des Ofens und Schmelzen der Seitenkruste führen [KGD+20, S. 1; TUV16, S. 131].

Ansätze

Das ungleichmäßige Abbrennen einer Anode kann zu einem Ansatz unter der Anode führen, der einen Kurzschluss verursachen kann [TCY13, S. 186]. Die Entstehung eines Ansatzes ist in Abbildung 3.7 zu erkennen. Ablagerungen unter der Anode können eine isolierende Fläche erzeugen, die schlechter abbrennt als der restliche Bereich der Anode. Im Laufe der Zeit bildet sich ein zunehmend größerer Ansatz, der zu kurzzeitigen Kurzschlüssen mit dem flüssigen Aluminium führen kann. Dies ist auf die wellenförmige Bewegung der Aluminiumoberfläche aufgrund des Magnetfelds und der Gasentwicklung an den Anoden zurückzuführen [GK93, S. 147]. Falls der Ansatz bis zu diesem Zeitpunkt nicht erkannt wird, kann dieser mit dem weiteren Absenken der Anoden in das flüssige Metall ragen und einen langanhaltenden Kurzschluss auslösen. Ein solcher Kurzschluss beeinträchtigt die Stromausbeute und führt zu einem zusätzlichen Wärmeeintrag [Mar17, S. 817; Mar18, S. 486–487; RWP00].

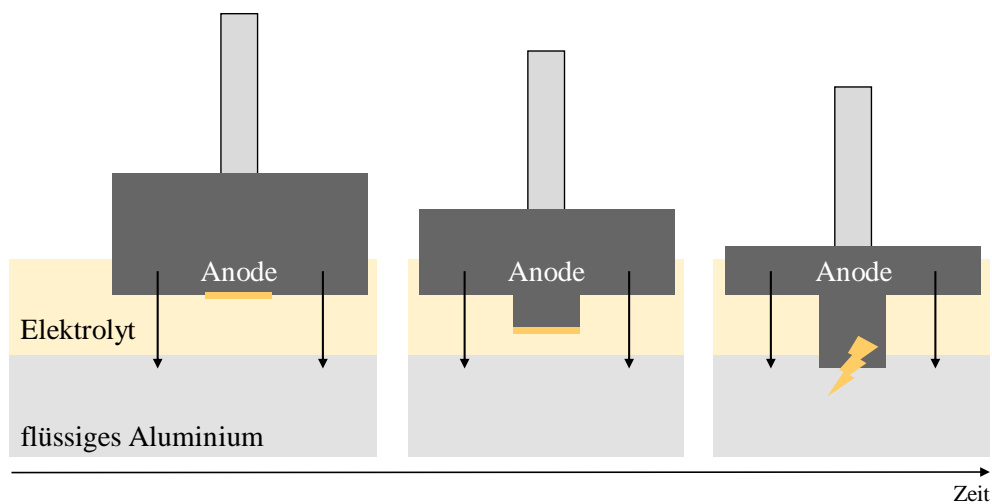


Abbildung 3.7: Entstehung eines Ansatzes unter einer Anode. Abgeändert nach [KGD+21, S. 766; Mar17, S. 817; Mar18, S. 487].

Rolofs et al. [RWP00] zeigen auf, dass die Badtemperatur von Öfen mit einer hohen Anzahl an Ansätzen und einer niedrigen durchschnittlichen Ofenspannung um durchschnittlich 4 °C höher lag als bei unauffälligen Öfen. Die Erhöhung der Badtemperatur erklären Rolofs et al. durch

den Kurzschluss, der einen zusätzlichen Wärmeeintrag bedeutet. Der zusätzliche Wärmeeintrag beeinflusst somit unmittelbar die Wärmebilanz eines Ofens und wirkt sich folglich auf dessen Badtemperatur aus.

Schlammbildung

Eine Schlammbildung kann in einem Ofen auftreten, wenn das zugeführte Aluminiumoxid nicht vollständig im Elektrolyt aufgelöst werden kann. Das nicht aufgelöste Aluminiumoxid setzt sich auf dem Boden des Ofens ab und führt zu einer Verschlammung. Aus Abbildung 3.8 geht hervor, dass der Schlamm wie ein zusätzlicher elektrischer Widerstand wirkt und eine ungleichmäßige Stromverteilung im Ofen erzeugt [GK93, S. 210; Tan10, S. 27; Ree15, S. 55].

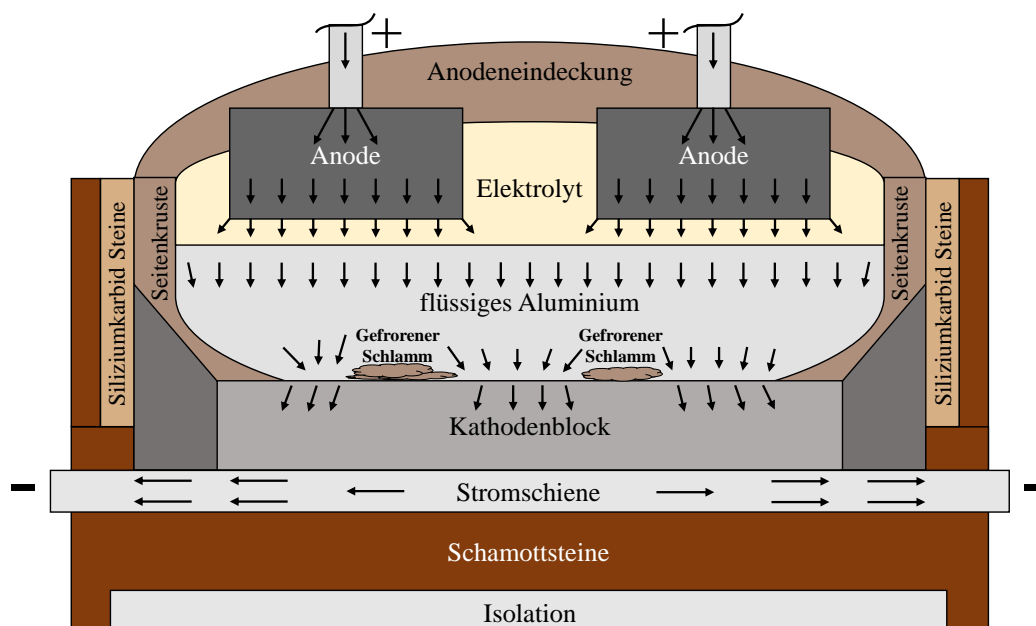


Abbildung 3.8: Der Schlamm auf den Boden eines Ofens wirkt wie ein zusätzlicher elektrischer Widerstand und führt zu einer ungleichmäßigen Stromverteilung. Abgeändert nach [Tan10, S. 27].

Anodeneindeckung

Zwischen 40 % und 60 % der gesamten Wärmeverluste werden über die Oberseite eines Ofens an die Umgebung abgegeben [GK93, S. 29–30]. Die isolierende Anodeneindeckung spielt eine wesentliche Rolle bei der Anpassung der Wärmeverluste über die Ofenoberseite [GK93, S. 29–30]. Für die Anodeneindeckung wird Aluminiumoxid oder zerkleinerter Elektrolyt verwendet [GK93, S. 202]. Durch die wärmeisolierende Wirkung der Anodeneindeckung ist eine Reduzierung der Ofenspannung möglich, da weniger Energie für die Wärmeentwicklung umgesetzt werden muss [TRI13, S. 32], um die Badtemperatur im Sollwertbereich zu halten. Eine zu dick aufgetragene Eindeckung kann jedoch dazu führen, dass die Anoden überhitzen, Ansätze

entstehen und sich die Badtemperatur des Ofens erhöht [TRI13, S. 36]. Die Anodeneindeckung hat daher einen wesentlichen Einfluss auf die Wärmebilanz eines Ofens und beeinflusst somit dessen Badtemperatur.

3.3 Zusammenfassung

In diesem Kapitel wurden die Grundlagen der industriellen Aluminiumherstellung aufgezeigt, für die der Hall-Héroult-Prozess genutzt wird. Die Wechselwirkungen eines Aluminiumelektrolyseofens wurden dargestellt. Der Aufbau eines Aluminiumelektrolyseofens wurde gezeigt, der in dieser Form auch bei der TAE zum Einsatz kommt. Die wesentlichen Einflüsse auf Badtemperatur wurden mithilfe der Literatur präsentiert. In diesem Zuge wurden auch die Auswirkungen einer einstündigen Hallenschaltung, die bei der TAE stattgefunden hat, auf die Badtemperatur von vier ausgewählten Öfen dargelegt. Die Ergebnisse zeigen, dass sich bei dieser Hallenschaltung die Badtemperatur der Öfen im Mittel um $-12,5\text{ °C}$ mit einer Standardabweichung von $2,7\text{ °C}$ geändert hat.

Insgesamt können die in diesem Kapitel dargestellten Einflüsse, die sich auf die Badtemperatur auswirken, für die Entwicklung von Vorhersagemodellen hilfreich sein. Bevor die ersten Lösungsansätze für die Vorhersage der Badtemperatur entwickelt werden, werden im nächsten Kapitel wesentliche Grundlagen der Zeitreihenanalyse vorgestellt. Diese Grundlagen sind für das weitere Verständnis nötig, da die Badtemperaturdaten der jeweiligen Öfen als Zeitreihe aufgefasst werden und in Kapitel 5 mit Methoden der Zeitreihenanalyse untersucht werden.



KAPITEL 4

Grundlagen der Zeitreihenanalyse

Das vorliegende Kapitel gliedert sich in zwei Teile: Im ersten Teil werden Grundlagen der Zeitreihenanalyse vorgestellt. Zunächst werden die Definitionen einer Zeitreihe sowie einer Zeitreihenvorhersage dargestellt. Anschließend werden die Komponenten beschrieben, aus denen eine Zeitreihe besteht. Damit wird ein Übergang zu dem Bereich der Zeitreihenmerkmale geschaffen, mit denen die Eigenschaften von Zeitreihen quantifiziert werden können. Zeitreihenmerkmale werden in unterschiedlichen Anwendungsbereichen verwendet, die in diesem Kapitel vorgestellt werden. Sie sind auch Bestandteil für die Gegenwartsvorhersage der Badtemperatur, die in dieser Arbeit durchgeführt wird. Des Weiteren wird in diesem Kapitel der Unterschied zwischen stationären und nicht stationären Zeitreihen aufgezeigt. Zudem werden die Grundlagen mehrerer Analysemethoden vorgestellt, wie die der (partiellen) Autokorrelation und des Periodogramms. Die aufgezeigten Methoden sind wesentlich für das Verständnis des praktischen Teils dieser Arbeit, der nach diesem Kapitel folgt.

Der zweite Teil dieses Kapitels fokussiert sich auf die Zeitreihenvorhersage. Es werden die Unterschiede zwischen lokalen und globalen Vorhersagemethoden aufgezeigt und die in dieser Arbeit verwendeten Vorhersagemethoden, das AR-Modell und RF-Modell, vorgestellt. Schließlich werden verschiedene Strategien zur sinnvollen Validierung und zum Testen von Vorhersagemethoden im Kontext der Zeitreihenanalyse präsentiert.

4.1 Zeitreihe

Die Erfassung der Dynamiken eines Prozesses kann mittels Zeitreihen erfolgen. Hierbei werden wiederholt Messungen über die Zeit an dem zu untersuchenden Prozess durchgeführt. Durch die anschließende Analyse der zeitlich angeordneten Messungen kann ein Verständnis über die Dynamik des Prozesses gewonnen werden [Ful18, S. 88]. Beispielsweise stellen die Badtemperaturmessungen eines Ofens, die täglich durchgeführt und zeitlich angeordnet werden, eine Zeitreihe dar.

Eine Zeitreihe

$$\{y_t : t = 1, \dots, n\} = \{y_1, y_2, \dots, y_n\} \quad (4.1)$$

besteht aus n Beobachtungen, die zu diskreten Zeitpunkten $1, 2, \dots, n$ aufgezeichnet werden. Wenn die Anzahl n keine Rolle spielt, dann wird Gleichung 4.1 mit $\{y_t\}$ abgekürzt [CM09, S. 19; HA21, S. 23–25].

Eine Vorhersage zum Zeitpunkt t für die Zeitreihe $\{y_t : t = 1, \dots, n\}$ für den Zeitpunkt $t + k$ wird mit $\hat{y}_{t+k|t}$ angegeben. So stellt $\hat{y}_{t+1|t}$ eine Vorhersage für y_{t+1} dar, für die alle verfügbaren Beobachtungen $\{y_1, y_2, \dots, y_t\}$ genutzt werden [CM09, S. 19; HA21, S. 23–25].

4.1.1 Komponenten einer Zeitreihe

Eine Zeitreihe kann in eine Trend-, Zyklus-, Saison- und Restkomponente zerlegt werden. Die Trend-, Zyklus- und Saisonkomponente sind die systematischen Komponenten einer Zeitreihe, während die Restkomponente als irreguläre Komponente angegeben wird. Trend- und Zykluskomponente werden in der Praxis zu einer Komponente zusammengefasst, die als Trend-Zyklus-Komponente oder Trendkomponente bezeichnet wird. Die Trendkomponente berücksichtigt langfristige Niveauänderungen einer Zeitreihe, während die Zykluskomponente Bewegungen einer Zeitreihe abbildet, die nicht mit einer festen Periode auftreten und im wirtschaftlichen Kontext als Konjunkturzyklen aufgefasst werden. Die Saisonkomponente hingegen bildet die Muster einer Zeitreihe ab, die mit einer festen Periode auftreten. Dies kann beispielsweise ein tägliches oder wöchentliches Muster sein. Die Restkomponente umfasst die verbleibenden Anteile einer Zeitreihe, die nicht von der Trend-Zyklus-Komponente und der Saisonkomponente beschrieben werden können [FHK+16, S. 509–511; HA21, S. 37–39, 59].

Die Zerlegung einer Zeitreihe in eine Trend-Zyklus-Komponente, Saison- und Restkomponente erfordert zunächst Annahmen über den Zusammenhang der einzelnen Komponenten. Häufig wird ein additiver oder multiplikativer Zusammenhang angenommen, wobei auch eine Mischform aus beiden Zusammenhängen möglich ist. Gleichung 4.2 zeigt den additiven Zusammenhang, während Gleichung 4.3 den multiplikativen Zusammenhang darstellt [FHK+16, S. 509–511; HA21, S. 64–69].

$$y_t = S_t + T_t + R_t \quad (4.2)$$

$$y_t = S_t \cdot T_t \cdot R_t \quad (4.3)$$

In den Gleichungen 4.2 und 4.3 stellen S_t die Saisonkomponente, T_t die Trend-Zyklus-Komponente und R_t die Restkomponente mit $t = 1, \dots, n$ dar. Der multiplikative Zusammenhang in Gleichung 4.3 sollte gewählt werden, wenn sich die Schwan-

kungen proportional mit dem Niveau der Zeitreihe verändern. Sofern keine proportionale Änderung vorliegt, ist der additive Zusammenhang für eine Zeitreihenzerlegung zu verwenden [FHK+16, S. 509–511; HA21, S. 64–69].

Für die Zerlegung einer Zeitreihe und die einzelnen Komponenten können unterschiedliche Methoden eingesetzt werden, die in [HA21; FHK+16; CM09] beschrieben werden. Die klassische Zeitreihenzerlegung bildet die Ausgangsbasis für weitere Zerlegungsmethoden [HA21, S. 76]. Bei der klassischen Zerlegung wird zunächst mittels eines gleitenden Durchschnitts (Moving-Average) die Trend-Zyklus-Komponente einer Zeitreihe geschätzt [HA21, S. 69; FHK+16, S. 516]. Mit dem gleitenden Durchschnitt wird die Zeitreihe geglättet, der sich mittels Gleichung 4.4 berechnen [HA21, S. 69] lässt.

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j} \quad (4.4)$$

\hat{T}_t ist der geschätzte Trend-Zyklus zum Zeitpunkt t und m stellt mit $m = 2k + 1$ die Ordnung des gleitenden Durchschnitts (m -Moving-Average (MA)) dar. m wird in der Regel abhängig von der Saisonalität der Zeitreihe gewählt. Beispielsweise wird bei monatlichen Daten eine jährliche Saisonalität angenommen. In diesem Fall beträgt $m = 12$. Wenn m gerade ist, wird nach der ersten Glättung ein weiteres Mal der gleitende Durchschnitt mit der Ordnung 2 über die bereits geglätteten Beobachtungen berechnet ($2 \times m$ -MA), um einen zentrierten gleitenden Durchschnitt zu erhalten [HA21, S. 72–73, 76].

Bei einer additiven Zerlegung wird anschließend eine Subtraktion $y_t - \hat{T}_t$ durchgeführt, aus der sich die trendbereinigte Zeitreihe ergibt. Anschließend werden von der trendbereinigten Zeitreihe die Mittelwerte anhand der entsprechenden Saisonalität gebildet. Für eine Zeitreihe mit einer monatlichen Auflösung werden beispielsweise alle Werte des Monats Januar gemittelt. Die Bildung des Mittelwerts wird für alle weiteren Monate durchgeführt, woraus zwölf gemittelte Werte resultieren. Die gemittelten Werte werden anschließend so angepasst, dass die Summe 0 ergibt. Abschließend werden die angepassten Werte abhängig von der Länge der ursprünglichen Zeitreihe repliziert. Die replizierten Werte ergeben die geschätzte Saisonkomponente \hat{S}_t . Durch die weitere Subtraktion $y_t - \hat{T}_t - \hat{S}_t$ wird die geschätzte Restkomponente \hat{R}_t ermittelt [HA21, S. 76].

Für die multiplikative Zerlegung wird nach erfolgter Glättung der Zeitreihe eine Division y_t / \hat{T}_t durchgeführt. Die weitere Vorgehensweise erfolgt analog zur additiven Zerlegung. Jedoch muss bei der Anpassung die Summe der gemittelten Werte m ergeben. Die geschätzte Restkomponente \hat{R}_t ergibt sich durch die Berechnung von $y_t / (\hat{T}_t \cdot \hat{S}_t)$ [HA21, S. 77].

Bedingt vom gleitenden Durchschnitt sind die ersten und letzten Werte der geschätzten Trend-Zyklus-Komponente \hat{T}_t und der Restkomponente \hat{R}_t bei der klassischen Zeitreihen-

zerlegung nicht vorhanden [HA21, S. 78]. Dieser und weitere Nachteile [HA21, S. 78] führen dazu, dass in der Praxis anderweitige Zerlegungsmethoden eingesetzt werden, wie zum Beispiel das Census X11-Verfahren, das Berliner Verfahren und das Seasonal and Trend decomposition using Loess (STL) Verfahren [FHK+16, S. 521; HA21, S. 78–84]. Die Zerlegung einer Zeitreihe in ihre einzelnen Komponenten erlaubt unter anderem die Berechnung von Zeitreihenmerkmalen, die die Eigenschaften einer Zeitreihe beschreiben.

4.1.2 Zeitreihenmerkmale/Anwendungsbereiche

Unter Zeitreihenmerkmalen werden numerische Beschreibungen verstanden, die die Eigenschaften von Zeitreihen quantifizieren [HA21, S. 89; Ful18, S. 89–90]. Einfache Zeitreihenmerkmale umfassen beispielsweise das empirische Mittel oder den Median, die die Lage von Zeitreihen beschreiben. Weitere Zeitreihenmerkmale, wie die Stärke der saisonalen und Trend-Zyklus-Komponente, können mithilfe einer zuvor durchgeführten additiven Zeitreihenzerlegung bestimmt werden. Die Stärke der saisonalen Komponente F_S wird über Gleichung 4.5 und die Stärke der Trend-Zyklus-Komponente F_T über Gleichung 4.6 berechnet [HA21, S. 92–97].

$$F_S = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right) \quad (4.5)$$

$$F_T = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)}\right) \quad (4.6)$$

Die Berechnung von F_S und F_T erlaubt eine Analyse von Zeitreihendatensätzen hinsichtlich der saisonalen und Trend-Zyklus-Komponente. Die Werte, die F_S und F_T annehmen können, liegen in dem Intervall $[0, 1]$. Die Berechnung von F_S ist hilfreich, um nicht saisonale Zeitreihen zu identifizieren. Nach [HA21, S. 273] ist eine Zeitreihe nicht saisonal, wenn $F_S < 0,64$ ist. Abbildung 4.1 zeigt zwölf Zeitreihen aus dem M1- und M3-Wettbewerb [Hyn18], die unterschiedliche Stärken bezüglich der saisonalen Komponente und der Trend-Zyklus Komponente aufweisen.

In Abbildung 4.1 sind in der ersten Reihe vier Zeitreihen dargestellt, die eine dominante Trend-Zyklus-Komponente aufweisen, während die saisonale Komponente in diesen Zeitreihen schwach ausgeprägt ist. Die mittlere Reihe zeigt vier Zeitreihen, die sich durch eine ausgeprägte Saisonalität auszeichnen. Dies wird anhand des hohen Werts von F_S für jede Zeitreihe deutlich. Die Trend-Zyklus-Komponente ist in diesen Zeitreihen hingegen nur in geringem Maße vorhanden. Die dritte Reihe der Abbildung 4.1 weist Zeitreihen mit einer dominanten Trend-Zyklus- und saisonalen Komponente auf.

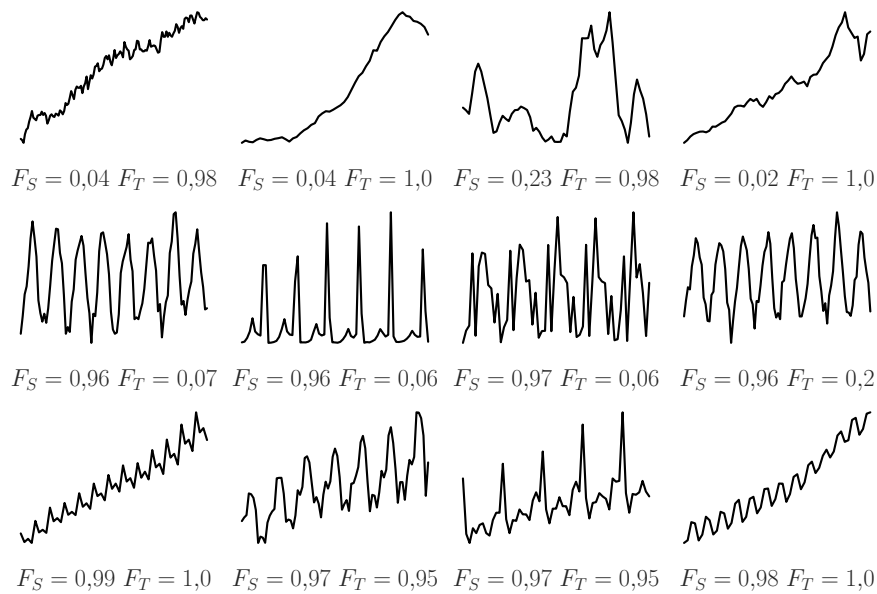


Abbildung 4.1: Zwölf Zeitreihen aus dem M1- und M3-Wettbewerb [Hyn18] besitzen unterschiedliche Stärken bezüglich der saisonalen und Trend-Zyklus Komponente. Abbildung abgeändert nach [GKD+23, S. 2].

Insgesamt kann die Berechnung von Zeitreihenmerkmalen auch als Dimensionsreduktionsverfahren verstanden werden, da die Zeitreihe auf eine bestimmte Anzahl an Merkmalen reduziert wird [WSH06, S. 339]. Die unterschiedlichen Einsatzmöglichkeiten von Zeitreihenmerkmalen, die im Folgenden erörtert werden, legen nahe, dass die Auswahl der Zeitreihenmerkmale anhand der spezifischen Problemstellung sowie der Eigenschaften der zu untersuchenden Zeitreihe getroffen werden sollte [KHL20, S. 358]. Somit wird sichergestellt, dass ein „globales Bild“ [WSH06, S. 339] von einer Zeitreihe erzeugt wird. Bei der Wahl von Zeitreihenmerkmalen kann Expertenwissen von Mitarbeitenden aus der zu untersuchenden Domäne hilfreich sein [GKD+23, S. 2]. Die Berechnung von Zeitreihenmerkmalen, wie dem empirischen Mittel, dem minimalen und dem maximalen Wert, ist sinnvoll, um ungewöhnliche Sensor-Zeitreihen von Thermoelementen im Bereich der Aluminiumelektrolyse zu identifizieren (vgl. unsere Arbeit [GKD+23]). Im Falle eines Vergleichs von Zeitreihen mit unterschiedlichen physikalischen Größen kann sich die Berechnung des empirischen Mittels, des minimalen und des maximalen Werts als nicht aussagekräftig erweisen. In diesem Fall sollten Zeitreihenmerkmale gewählt werden, die unabhängig von physikalischen Größen sind, um sinnvolle Aussagen bezüglich der Eigenschaften von unterschiedlichen Zeitreihen treffen zu können. Solche Zeitreihenmerkmale sind beispielsweise die Stärke der Saisonalität und die Stärke der Trend-Zyklus-Komponente.

Anwendungsbereiche

Beschreibungen von Zeitreihen werden zusammen mit Ähnlichkeitsmaßen in unterschiedlichen Bereichen eingesetzt, um nach Zeitreihen mit bestimmten Eigenschaften zu suchen, Anomalien in Zeitreihen zu erkennen, wiederkehrende Verläufe in Zeitreihen zu ermitteln und Zeitreihen zu gruppieren oder zu klassifizieren [Ful18, S. 91]. In unserer Arbeit [GKD+23, S. 3–4] werden sechs unterschiedliche Anwendungsbereiche aufgezeigt, in denen Zeitreihenmerkmale verwendet werden. Diese Anwendungsbereiche werden anschließend mit entsprechenden Literaturangaben dargestellt. Auch wenn in einzelnen Arbeiten Zeitreihenmerkmale nicht explizit erwähnt werden, lässt sich anhand der Vorgehensweise eine Verknüpfung zum diesem Bereich herstellen.

Regression: Im Rahmen des FlexTherm Projekts wird in unserer Arbeit [GKD+18] für verschiedene Prozessvariablen das arithmetische Mittel oder die Summe innerhalb gleitender Fenster berechnet. Diese berechneten Zeitreihenmerkmale werden anschließend für das Training eines RF-Regressionsmodells verwendet, um die Badtemperatur von Aluminiumelektrolyseöfen vorherzusagen.

Klassifikation: In [KGD+20] verwenden die Autoren Zeitreihenmerkmale aus [LSK+19] und [GGC+], die innerhalb gleitender Fenster von verschiedenen Prozessvariablen berechnet werden. Die berechneten Zeitreihenmerkmale werden für das Training von Vorhersagemodellen verwendet, um anschließend Anodeneffekte in Aluminiumelektrolyseöfen vorherzusagen. Des Weiteren wird in [NAM01] ein MLP zusammen mit acht Zeitreihenmerkmalen verwendet, um Muster in Control Charts zu klassifizieren.

Detektion von Anomalien: Um ungewöhnliche Server-Lasten bei der Internet-Firma Yahoo zu identifizieren, werden in [HWL15] 18 Merkmale von Zeitreihen berechnet. Die Merkmale werden anschließend in einer PCA verwendet, um ungewöhnlich Zeitreihen zu identifizieren. Das Ziel der PCA wurde bereits in Kapitel 2.3 erläutert. In [THSM+20] wird ein Framework vorgestellt, um abnormale Zeitreihen in einem Datenstrom zu identifizieren. Das Framework besteht aus einer Offline- und Online-Phase. In der Offline-Phase wird ein Modell anhand von Zeitreihen trainiert, die das gewöhnliche Verhalten des betrachteten Systems repräsentieren. In der Online-Phase wird das trainierte Modell unter Verwendung eines gleitenden Fensters auf neue Daten angewendet, um abnormale Zeitreihen zu erkennen. Das Framework beinhaltet unter anderem die Berechnung von 14 Zeitreihenmerkmalen, mit denen eine PCA durchgeführt wird. Außerdem wird ein Algorithmus vorgestellt, um nicht stationäre Systemzustände zu erkennen und das Modell auf einen neuen Zustand einzustimmen.

Clustering: In [HV12] verwenden Horvath et al. das hierarchische Clustern, um Aluminiumelektrolyseöfen in Gruppen mit gleichartigem Verhalten einzuteilen. Dazu werden für jede Prozessvariable sieben Zeitreihenmerkmale berechnet, um den Umfang der Daten zu reduzieren.

Visualisierung: In [KHSM17] werden sechs Zeitreihenmerkmale berechnet, um den M3-Datensatz [MH00] zu visualisieren und zu analysieren. Anhand der berechneten Zeitreihenmerkmale wird eine PCA durchgeführt und die ersten beiden Hauptkomponenten visualisiert, um einen Überblick über die Eigenschaften des M3-Datensatzes zu erhalten. Mithilfe eines genetischen Algorithmus werden zudem neue Zeitreihen generiert, die nicht typisch für den M3-Datensatz sind und den Merkmalsraum des M3-Datensatzes erweitern. Darüber hinaus vergleichen Kang et al. den Merkmalsraum mit der Performanz von ausgewählten Vorhersagemethoden, um daraus gezielt Vorhersagemethoden abzuleiten, die geeignet für die Vorhersage einer Zeitreihe erscheinen.

Meta-Learning: In der Arbeit [THA23] wird das Framework FFORMS (Feature-based FOREcast Model Selection) vorgestellt, um eine geeignete Methode für eine Zeitreihenvorhersage zu finden. Dabei werden insgesamt 37 Zeitreihenmerkmale mit vorgegebenen Vorhersagemethoden über ein RF-Modell verknüpft. In [PL04] greifen Prudêncio et al. auf den Einsatz von Algorithmen aus dem Bereich des maschinellen Lernens zurück, um die Auswahl an Vorhersagemethoden für die Zeitreihenvorhersage anhand von Zeitreihenmerkmalen zu automatisieren. Im ersten Fallbeispiel wird ein Entscheidungsbaum als Meta-Learner mit zehn Zeitreihenmerkmalen von stationären Zeitreihen für die Klassifizierung von zwei Vorhersagemethoden verwendet. Für das zweite Fallbeispiel werden fünf Zeitreihenmerkmale von den jährlichen Daten aus dem M3-Datensatz [MH00] berechnet und als Eingang für drei Klassifizierer vom Typ MLP als Meta-Learner verwendet.

Aus dieser Übersicht geht hervor, dass Zeitreihenmerkmale in unterschiedlichen Kontexten eine Anwendung finden. Unter anderem in Kombination mit gleitenden Fenstern werden Zeitreihenmerkmale eingesetzt, um anschließend eine Regression, Klassifikation oder Anomaliedetektion durchzuführen. Diese Kombination wird auch für die Gegenwartsvorhersage (Nowcasting) der Badtemperatur eingesetzt, die in Kapitel 5.4.2 dieser Arbeit durchgeführt wird.

4.1.3 Stationäre Zeitreihen

Eine Zeitreihe wird als stationär bezeichnet, wenn ihre statistischen Eigenschaften nicht von der Zeit abhängig sind [HA21, S. 265; Nel20, S. 1175]. Stationäre Zeitreihen weisen weder einen Trend noch eine Saisonalität auf und zeigen einen annähernd waagerechten Verlauf mit konstanter Varianz. Zyklisches Verhalten kann jedoch enthalten sein [HA21, S. 266]. Stationäre Zeitreihen spielen eine wichtige Rolle bei der Entwicklung von Autoregressive Integrated Moving Average (ARIMA) Modellen. Des Weiteren hilft die Unterscheidung zwischen stationären und nicht stationären Zeitreihen bei der Wahl einer geeigneten Validierungsmethode, um eine realistische Modellperformanz abschätzen zu können (vgl. Kapitel 4.3).

Durch die Bildung von Differenzen einer Zeitreihe kann aus einer nicht stationären Zeitreihe eine stationäre Zeitreihe erzeugt werden. Die Differenzierung trägt dazu bei, den Mittelwert einer

Zeitreihe zu stabilisieren. Bei der gewöhnlichen Differenzierung wird die Differenz zwischen der aktuellen Beobachtung und der vorherigen Beobachtung gebildet. Die Berechnung der gewöhnlich differenzierten Zeitreihe erfolgt über Gleichung 4.7. In der Literatur werden diese differenzierten Beobachtungen auch als „erste Differenzen“ bezeichnet [HA21, S. 267–269].

$$y'_t = y_t - y_{t-1} \quad (4.7)$$

y'_t stellt die differenzierte Beobachtung der Zeitreihe zum Zeitpunkt t dar. Bei Zeitreihen, die saisonalen Schwankungen unterliegen, erfolgt in der Regel eine saisonale Differenzierung, um den Mittelwert der Zeitreihe zu stabilisieren. Bei der saisonalen Differenzierung wird die Differenz zwischen der aktuellen Beobachtung und der Beobachtung aus der vorherigen Saison gebildet. Eine saisonale Differenzierung mit der Periode m wird mittels Gleichung 4.8 erreicht [HA21, S. 269].

$$y'_t = y_t - y_{t-m} \quad (4.8)$$

Auch hier stellt y'_t die differenzierte Beobachtung der Zeitreihe zum Zeitpunkt t dar. Eine Zeitreihe mit einer monatlichen Auflösung kann beispielsweise eine jährliche Saisonalität beinhalten. In diesem Fall kann eine saisonale Differenzierung mit der Periode $m = 12$ durchgeführt werden, da sich das Muster bei einer jährlichen Saisonalität alle zwölf Monate wiederholt.

Da saisonale Zeitreihen nicht stationär sind, ist es daher sinnvoll, diese zunächst auf enthaltene Saisonalitäten zu untersuchen. Dazu kann nach [HA21, S. 273] das in Kapitel 4.1.2 vorgestellte Zeitreihenmerkmal, die Stärke der Saisonalität, berechnet werden. Falls die Berechnung der Stärke der Saisonalität ergibt, dass die Zeitreihe saisonal ist, so ist eine saisonale Differenzierung empfehlenswert.

Anschließend kann durch den Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test [KPSS92] überprüft werden, ob eine gewöhnliche Differenzierung durchgeführt werden sollte. Im Rahmen dieses Tests wird als Nullhypothese angenommen, dass die Zeitreihe stationär ist. Anhand eines gewählten Signifikanzniveaus wird anschließend darüber entschieden, ob die Nullhypothese abgelehnt wird. Bei Ablehnung der Nullhypothese ist die Zeitreihe nicht stationär. In diesem Fall ist eine gewöhnliche Differenzierung empfehlenswert. Der KPSS-Test sollte danach erneut auf die differenzierte Zeitreihe angewendet werden [HA21, S. 272–273].

4.1.4 Autokorrelation

Die Autokorrelation beschreibt den linearen Zusammenhang zwischen einer Zeitreihe und einer von sich selbst zeitlich verschobenen Version [Nel20, S. 1182]. Die Zeitreihe wird mit sich selbst korreliert. Dazu wird die Zeitreihe mehrfach um einen Schritt verschoben und jeweils der

Korrelationskoeffizient zwischen der ursprünglichen Zeitreihe und ihrer verschobenen Version berechnet. Die Autokorrelationskoeffizienten einer Zeitreihe können durch die Autokorrelationsfunktion (AKF) in Gleichung 4.9 berechnet werden und liegen zwischen -1 und 1 [HA21, S. 52; CM09, S. 33–34].

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (4.9)$$

r_k stellt den berechneten Korrelationskoeffizient an der Verzögerungsstelle k dar. n ist die Anzahl der Beobachtungen der Zeitreihe. \bar{y} ist der Mittelwert der Zeitreihe, der mit Gleichung 4.10 berechnet wird [CM09, S. 31].

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t \quad (4.10)$$

In der Regel wird der berechnete Autokorrelationskoeffizient r_k gegen die Verzögerungsstelle k aufgetragen. Die resultierende Darstellung wird als Korrelogramm bezeichnet [HA21, S. 52]. In Abbildung 4.2 ist ein Beispiel für ein Korrelogramm mit einem Konfidenzintervall von 95 % zu sehen, das die Autokorrelationskoeffizienten zeigt, die mit Beobachtungen eines autoregressiven Prozesses berechnet wurden [CM09, S. 81–83]. Wie aus Abbildung 4.2 ersichtlich, weisen die Koeffizienten einen sinusförmigen Verlauf auf. Das deutet darauf hin, dass die Beobachtungen aus einem autoregressiven Prozess stammen [HA21, S. 283].

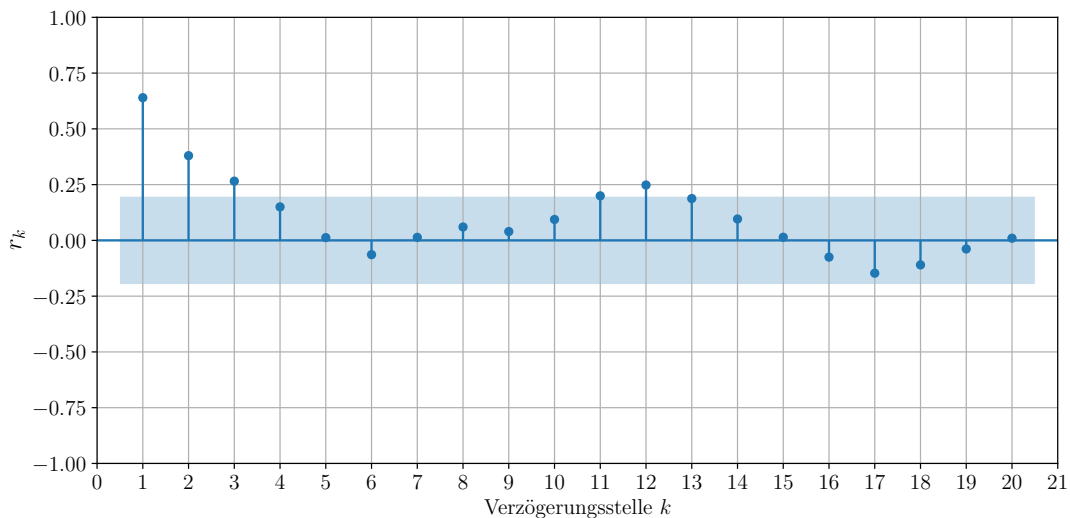


Abbildung 4.2: Beispiel für ein Korrelogramm der AKF. Entnommen aus [CM09, S. 81–83].

Neben der AKF wird üblicherweise auch die partielle Autokorrelationsfunktion (PAKF) berechnet. Mit ihr wird der Effekt von übergreifenden Korrelationen aufgrund vorheriger Verzögerungsstellen entfernt. Anschließend wird die Korrelation an der Stelle k berechnet [CM09, S. 81]. Dazu kann ein autoregressives Modell, $AR(k)$, verwendet werden, das mit den Beobachtungen der Zeitreihe trainiert wird. Der k -te Modellparameter des trainierten $AR(k)$ -Modells entspricht dann dem partiellen Autokorrelationskoeffizient α_k [CM09, S. 81; HA21, S. 282]. Das $AR(k)$ -Modell wird in Kapitel 4.2.2 vorgestellt.

In Abbildung 4.3 ist das Korrelogramm der PAKF mit einem Konfidenzintervall von 95 % zu sehen. Die Koeffizienten der PAKF wurden anhand der gleichen Beobachtungen ermittelt, die bereits für das Korrelogramm in Abbildung 4.2 verwendet wurden. Mit Ausnahme des Korrelationskoeffizient α_1 sind alle weiteren Koeffizienten in Abbildung 4.3 nicht signifikant.

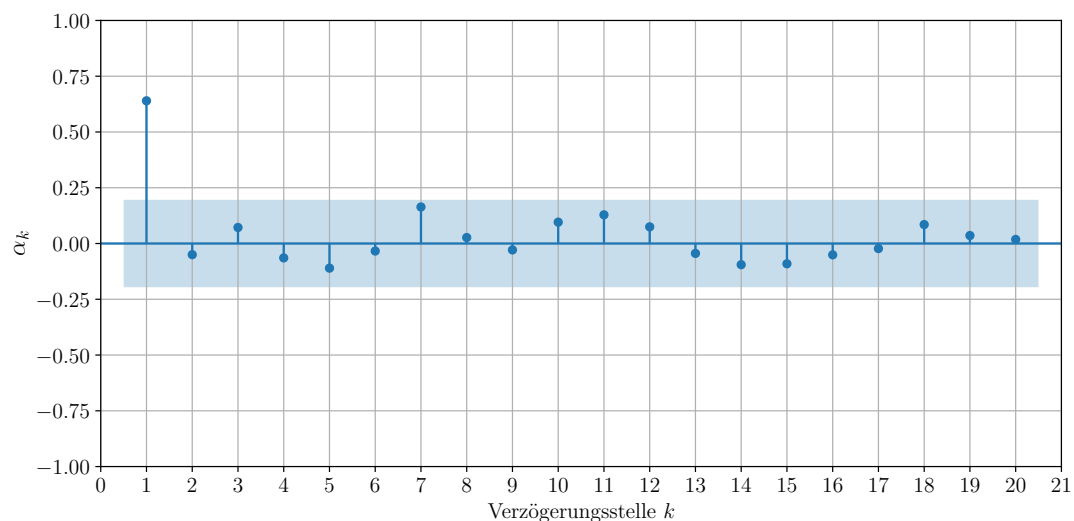


Abbildung 4.3: Beispiel für ein Korrelogramm der PAKF. Entnommen aus [CM09, S. 81–83].

In Kombination mit den Erkenntnissen aus Abbildung 4.2 lässt sich Folgendes ableiten: Die in Abbildung 4.2 dargestellten Koeffizienten weisen einen sinusförmigen Verlauf auf, während lediglich der Korrelationskoeffizient α_1 der PAKF signifikant ist. Diese Ergebnisse lassen den Schluss zu, dass es sich um Beobachtungen handelt, die aus einem autoregressiven Prozess erster Ordnung $AR(1)$ stammen [HA21, S. 283; CM09, S. 81]. Da die AKF und PAKF an der Verzögerungsstelle $k = 0$ immer 1 ist, wird dieser Wert in den Abbildungen 4.2 und 4.3 nicht angegeben.

Insgesamt kann mit der AKF und PAKF die Modellordnung k eines $AR(k)$ -Modells ermittelt werden, wenn die verwendeten Beobachtungen auch aus einem autoregressiven Prozess stammen. Des Weiteren lässt sich mithilfe der AKF und PAKF die Ordnung eines Moving-Average-Modells

(MA-Modells) bestimmen. Das MA-Modell sollte nicht mit dem gleitenden Durchschnitt bei der Zeitreihenzerlegung in Kapitel 4.1.1 verwechselt werden. Im Rahmen dieser Arbeit erfolgt keine Betrachtung des MA-Modells. Für weiterführende Informationen zum Thema MA-Modell sei auf die Literatur [HA21; CM09; Nel20] verwiesen.

4.1.5 Periodogramm

Die Identifikation der in der Zeitreihe enthaltenen Frequenzen mittels eines Periodogramms erlaubt Rückschlüsse auf mögliche Periodizitäten. Dabei wird eine Zeitreihe als eine Summe an Sinusschwingungen aufgefasst. Für eine Zeitreihe $\{y_t : t = 1, \dots, n\}$ mit einer geraden Anzahl an Beobachtungen kann dazu Gleichung 4.11 angegeben werden [CM09, S. 171, 173; SS17, S. 169].

$$y_t = a_0 + a_1 \cos\left(\frac{2\pi t}{n}\right) + b_1 \sin\left(\frac{2\pi t}{n}\right) + \dots \\ + a_{n/2-1} \cos\left(\frac{2(n/2-1)\pi t}{n}\right) + b_{n/2-1} \sin\left(\frac{2(n/2-1)\pi t}{n}\right) + a_{n/2} \cos(\pi t) \quad (4.11)$$

Gleichung 4.11 ist ein Regressionsmodell, bei dem die Koeffizienten $a_1, b_1, a_2, b_2, a_3, b_3, \dots, a_{n/2-1}, b_{n/2-1}, a_{n/2}$ geschätzt werden. Der Mittelwert der Zeitreihe ist a_0 . Mit Gleichung 4.12 kann die Amplitude für die m -te Harmonische berechnet werden, wobei m zwischen 1 und $n/2$ liegt und eine ganze Zahl ist. Die berechnete Amplitude gibt die Stärke der jeweiligen Frequenz in der Zeitreihe an [CM09, S. 173–174; SS17, S. 169].

$$A_m = \sqrt{a_m^2 + b_m^2} \quad (4.12)$$

Die quadrierten Amplituden A_m^2 werden anschließend gegen m/n aufgetragen, wodurch sich das skalierte Periodogramm für die Zeitreihe ergibt. In der Praxis wird für die Berechnung der Amplituden die Fast Fourier Transform (FFT) [CT65] verwendet [SS17, S. 169; CM09, S. 174].

4.2 Vorhersagemethoden

Im Bereich der Zeitreihenanalyse kann der lokale oder globale Ansatz gewählt werden, um eine Zeitreihenvorhersage durchzuführen. Beide Ansätze werden im nachfolgenden Kapitel erörtert. Anschließend werden das AR-Modell und RF-Modell vorgestellt, die in dieser Arbeit für die Vorhersage der Badtemperatur verwendet werden.

4.2.1 Lokale und globale Vorhersagemodelle

Eine übliche Vorgehensweise in der Zeitreihenvorhersage ist es, anhand einer Zeitreihe ein Vorhersagemodell zu trainieren, das anschließend für die Vorhersage der Zeitreihe verwendet wird. Für jede neue Zeitreihe wird ein eigenständiges Modell trainiert. Eine solche Vorgehensweise wird als lokal bezeichnet, da angenommen wird, dass jede Zeitreihe jeweils aus einem unterschiedlichen Erzeugungsprozess stammt [SFGJ20; MMH21, S. 1632]. Der lokale Ansatz kann bei Zeitreihen mit wenigen Beobachtungen zu einer Überanpassung führen, sofern nicht mittels entsprechender Gegenmaßnahmen, wie der Berücksichtigung von weiteren Informationen oder der Einschränkung der Modellanpassung, gegengesteuert wird. Diese Gegenmaßnahmen erfordern jedoch eine Einzelbetrachtung jeder Zeitreihe, woraus ein hoher zeitlicher Aufwand entstehen kann [MMH21, S. 1632].

Im Vergleich zur lokalen Vorgehensweise erfolgt beim globalen Ansatz zunächst eine Zusammenführung aller Zeitreihen, die anschließend für das Training eines einzelnen (globalen) Modells verwendet werden. Die größere Anzahl an verfügbaren Beobachtungen ermöglicht es beim globalen Ansatz, eine Überanpassung zu verhindern. Der globale Ansatz basiert auf der Annahme, dass die betrachteten Zeitreihen aus demselben Erzeugungsprozess stammen [MMH21, S. 1633]. So werden in [BBS20] zunächst ähnliche Zeitreihen anhand von Zeitreihenmerkmalen (vgl. Kapitel 4.1.2) gruppiert. Anschließend wird für jede Gruppe ein globales Modell trainiert. In [SFGJ20] wird zudem ein globales neuronales Netz, DeepAR, vorgestellt, das auf ähnlichen Zeitreihen basiert.

In [MMH21] wird aufgezeigt, dass beim globalen Ansatz eine Ähnlichkeit von Zeitreihen nicht erforderlich ist. Die Ergebnisse legen unter anderem nahe, dass globale Modelle auch bei heterogenen Zeitreihen eine wettbewerbsfähige Performanz gegenüber den lokalen Modellen aufweisen. Ein lineares AR-Modell dient dabei als globales Basismodell, das einfach zu implementieren ist und sich hinsichtlich der Modellstruktur mit dem klassischen ARIMA-Zeitreihenmodell überschneidet [MMH21, S. 1639].

Die einfache Handhabung des globalen AR-Modells sowie dessen wettbewerbsfähigen Modellperformanz, die in [MMH21] aufgezeigt wird, sind Argumente für die Verwendung eines globalen AR-Modells für die Tagesprognose (Forecasting) der Badtemperatur. Zwar handelt es sich bei den aufgezeichneten Badtemperaturen um Zeitreihen, die als ähnlich angesehen werden können, dies sollte jedoch keine Einschränkung für das globale AR-Modell darstellen. Aus dem globalen AR-Modell ergibt sich zudem eine Ausgangsbasis für weitere Forschungsarbeiten, in deren Rahmen zusätzliche (globale) Modelle für die Vorhersage der Badtemperatur entwickelt und anschließend mit dem Basismodell verglichen werden können. Das globale AR-Modell wird im nachfolgenden Kapitel 4.2.2 beschrieben.

Die Gegenwartsvorhersage (Nowcasting) der Badtemperatur wird mit einem globalen RF-Modell durchgeführt, das im Kapitel 4.2.3 vorgestellt wird. Die Wahl des RF-Modells wird mit den in

[BLS+24; HG20] dargelegten Ergebnissen und Argumenten begründet. In [BLS+24] werden klassische maschinelle Lernmodelle mit Deep Learning Modellen anhand von tabellarischen Datensätze gegenübergestellt. Vier Datensätze werden für die Klassifikation und ein Datensatz für die Regression verwendet. Die Gegenüberstellung belegt, dass Entscheidungsbäume, die auf dem Gradient Boosting Verfahren basieren, die meisten Deep Learning Modelle hinsichtlich der Vorhersageperformanz übertreffen. Des Weiteren zeigen die Ergebnisse, dass das RF-Modell eine bessere Performanz auf dem Regressionsdatensatz aufweist als die meisten in der Arbeit verwendeten Deep Learning Ansätze. Zwar erreicht das RF-Modell nicht die Performanz der Gradient Boosting Modelle. Dennoch liegt es im Ergebnis direkt hinter diesen [BLS+24, S. 7511]. Auch wenn in [BLS+24] keine Zeitreihen für die Untersuchung verwendet werden, zeigen die Ergebnisse den effektiven Einsatz des RF-Modells in unterschiedlichen Szenarien auf. Jedoch können auch Zeitreihendaten tabellarisch strukturiert werden, wie im späteren Kapitel 5.4.2 für die Gegenwartsvorhersage (Nowcasting) aufgezeigt wird.

In [HG20, S. 325] wird vorgeschlagen, dass für eine Vorhersage mit tabellarischen Daten zunächst ein RF-Modell verwendet werden soll. Gradient Boosting Modelle sind nach [HG20, S. 325] zwar etwas genauer in der Vorhersage, gehen jedoch mit einer Vielzahl an Hyperparametern einher, die erst anhand der Daten angepasst werden müssen. Des Weiteren neigt ein RF-Modell in der Regel nicht zu einer Überanpassung, was bei Gradient Boosting Modellen und neuronalen Netzen der Fall ist [HG20, S. 325]. Eine weitere Eigenschaft des RF-Modells ist, dass die Daten nicht vorab normalisiert werden müssen, was bei einem Einsatz von neuronalen Netzen in der Regel erforderlich ist [HG20, S. 325]. Das hat den Vorteil, dass die Praxisimplementierung und die Wartung des RF-Modells vereinfacht wird.

Das RF-Regressionsmodell ist jedoch nicht in der Lage, eine Zielvariable zu extrapolieren [HG20, S. 325]. Da die Badtemperatur der Aluminiumelektrolyseöfen in einem bestimmten Wertebereich liegt, wie im späteren Kapitel 5.3.2 aufgezeigt wird, stellt die Schwierigkeit der Extrapolation in dieser Arbeit allerdings kein Problem dar. Daher ist das RF-Modell in dieser Arbeit ein Lösungsansatz für die Gegenwartsvorhersage der Badtemperatur, der aufbauend auf unserer Arbeit [GKD+18] weiter untersucht wurde.

4.2.2 Autoregressives Modell

Das autoregressive Modell mit der Ordnung k , $AR(k)$, setzt sich aus einer Linearkombination vergangener Beobachtungen einer Zeitreihe zusammen [HA21, S. 275]. Die Vorhersage des $AR(k)$ -Modells für den Zeitpunkt t ist mit Gleichung 4.13 angegeben.

$$\hat{y}_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_k y_{t-k} \quad (4.13)$$

Gleichung 4.13 kann auch als ein multiples lineares Regressionsmodell aufgefasst werden, mit dem Unterschied, dass vergangene Beobachtungen als Merkmale anstatt externer Prädiktoren eingesetzt werden [HA21, S. 275]. ϕ_1, \dots, ϕ_k stellen die zu schätzenden Parameter und c die zu schätzende Konstante des $\text{AR}(k)$ -Modells dar. Die Schätzung der Modellparameter und der Konstante kann mittels Minimierung der Summe der quadratischen Fehler erfolgen [KND15, S. 333; CM09, S. 79].

Das $\text{AR}(k)$ -Modell lässt sich sowohl lokal als auch global einsetzen. Um ein globales $\text{AR}(k)$ -Modell zu erzeugen, wird ein Trainingsvorgang durchgeführt, bei dem die Modellparameter und die Konstante in Gleichung 4.13 anhand der Beobachtungen aller vorliegenden Zeitreihen geschätzt werden. Ziel des globalen $\text{AR}(k)$ -Modells ist es, anhand der vergangenen Beobachtungen den nächsten Wert jeder Zeitreihe vorherzusagen [CTM20, S. 2004]. Dafür werden die Beobachtungen zunächst in Abhängigkeit von der Ordnung k in eine Matrix eingebettet. Dieser Vorgang wird in der Literatur auch als Lag Embedding bzw. Time Delay Embedding bezeichnet [MMH21, S. 1639; CTM20, S. 2004]. Die folgende Matrix resultiert, wenn die Beobachtungen einer Zeitreihe $\{y_1, y_2, \dots, y_n\}$ eingebettet werden [BHK18, S. 71].

$$\begin{bmatrix} y_1 & y_2 & \dots & y_k & y_{k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{t-k} & y_{t-k+1} & \dots & y_{t-1} & y_t \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{n-k} & y_{n-k+1} & \dots & y_{n-1} & y_n \end{bmatrix}$$

Die letzte Spalte der resultierenden Matrix ist die Zielvariable, während die restlichen Spalten als Merkmale für die Vorhersage der Zielvariable eingesetzt werden. Anhand der aufgezeigten Matrix ist erkennbar, dass eine 1-Schritt-Vorhersage durchgeführt wird. Für das globale $\text{AR}(k)$ -Modell wird für jede Zeitreihe eine solche Einbettung durchgeführt. Die resultierenden Matrizen werden anschließend zu einer großen Matrix zusammengeführt, mit der das globale $\text{AR}(k)$ -Modell trainiert, validiert und getestet wird [MMH21, S. 1639]. In dieser Arbeit wird das globale $\text{AR}(k)$ -Modell durch die LinearRegression-Klasse aus der scikit-learn Bibliothek [PVG+11] in der Version 1.3.2 umgesetzt.

4.2.3 Random Forest

Der Random Forest (RF) besteht aus einer Ansammlung an Entscheidungsbäumen, die jeweils eine Vorhersage über eine Zielvariable tätigen. Die einzelnen Vorhersagen werden anschließend zu einer Gesamtvorhersage aggregiert. Ein Entscheidungsbaum ist ein algorithmisches Modell, bei dem der Datensatz in Abhängigkeit von der Zielvariable in möglichst homogene Gruppen eingeteilt wird [Bre01b, S. 199; KND15, S. 136]. Die Basis für die Entwicklung eines Entscheidungsbaums ist der Iterative Dichotomizer 3 (ID3) Algorithmus nach [Qui86], bei dem

der Informationsgewinn als Metrik für die Einteilung von kategorialen Daten verwendet wird [KND15, S. 134–137]. Darüber hinaus existieren weitere Algorithmen wie C4.5, J48 und CART (Classification And Regression Trees), die auf dem ID3-Algorithmus aufbauen [KND15, S. 167]. Der ID3-Algorithmus erzeugt einen möglichst flachen Entscheidungsbaum mittels rekursiver Vorgehensweise [KND15, S. 135, 144].

Grundlegend besteht ein Entscheidungsbaum aus einem Wurzelknoten, inneren Knoten und Blättern. Die Knoten und Blätter sind über Äste miteinander verbunden [KND15, S. 121]. Die Knoten enthalten jeweils Entscheidungsregeln, die ausgehend vom Wurzelknoten nacheinander abgearbeitet werden. Abhängig vom Ergebnis der Entscheidungsregel in einem Knoten, wird ein bestimmter Ast gewählt, um zum nächsten Knoten zu gelangen. Dieser Ablauf wird solange fortgesetzt, bis ein Blatt erreicht wird. Das Blatt enthält die endgültige Vorhersage der Zielvariable [KND15, S. 122]. In Abbildung 4.4 sind drei Entscheidungsbäume zu sehen, die unterschiedlich aufgebaut sind.

Die Entscheidungsregeln in den Knoten werden mit einem geeigneten Algorithmus aus den Daten gelernt. Für kategoriale Daten berechnet der ID3-Algorithmus den Informationsgewinn [KND15, S. 134–135, 144]. Beginnend beim Wurzelknoten, wird das kategoriale Merkmal aus den Daten ausgewählt, das den höchsten Informationsgewinn erzeugt. Je höher der Informationsgewinn eines kategorialen Merkmals ist, desto besser ist das Merkmal in der Lage, die Daten hinsichtlich der Zielvariable in homogene Gruppen einzuteilen. Das Merkmal mit dem höchsten Informationsgewinn bildet die erste Entscheidungsregel und somit den Wurzelknoten. Anhand der aufgestellten Entscheidungsregel wird der Datensatz in weitere Gruppen eingeteilt. Für jede dieser neuen Gruppen wird der ID3-Algorithmus erneut ausgeführt. Dieser Vorgang wird solange durchgeführt, bis jede Gruppe hinsichtlich der Zielvariable homogen ist oder eine weitere Einteilung der Daten nicht mehr möglich ist. In jedem Pfad, der vom Wurzelknoten bis zu einem bestimmten Blatt verläuft, wird ein Merkmal lediglich einmal für die Einteilung der Daten herangezogen [KND15, S. 134–137].

Bei einem kontinuierlichen Merkmal wird für die Einteilung der Daten ein Grenzwert ermittelt, der den größten Informationsgewinn liefert. Anschließend wird dieser Wert mit dem jeweiligen Informationsgewinn der restlichen Merkmale verglichen. Im Vergleich zu kategorialen Merkmalen kann ein kontinuierliches Merkmal mehrmals für die Einteilung der Daten entlang eines Pfades eingesetzt werden. Für eine Zielvariable, die kontinuierliche Werte enthält, wird die gewichtete Varianz berechnet. Das Merkmal, das die gewichtete Varianz der Zielvariable minimiert, teilt die Daten an einem Knoten weiter auf. Damit kann ein Entscheidungsbaum auch für die Regression eingesetzt werden [KND15, S. 150, 152–154].

Entscheidungsbäume besitzen in der Regel einen geringen Verzerrungsfehler [Lou14, S. 69; HTF09, S. 587–588]. Allerdings sind sie sensitiv gegenüber verrauschten Daten, was zu einem

hohen Varianzfehler führt [HTF09, S. 587; KND15, S. 158–159]. In Kapitel 4.3 werden die Begriffe Verzerrungs- und Varianzfehler erklärt.

Der RF verringert den Varianzfehler, indem mehrere Entscheidungsbäume trainiert werden, die untereinander eine möglichst geringe Korrelation aufweisen [HTF09, S. 587–588]. Dazu wird jeder Entscheidungsbaum auf zufällig gezogenen Daten aus dem Datensatz trainiert. Als Beispiel ist in Abbildung 4.4 ein RF-Modell zu sehen, das aus drei unterschiedlichen Entscheidungsbäumen besteht. Mit jedem Entscheidungsbaum wird eine Vorhersage über die Zielvariable erzeugt. Die einzelnen Vorhersagen werden bei einer Regression über die Berechnung des arithmetischen Mittels oder des Medians zu einer Gesamtvorhersage aggregiert. Bei einer Klassifikation wird die Gesamtvorhersage über einen Mehrheitsentscheid ermittelt [KND15, S. 165].

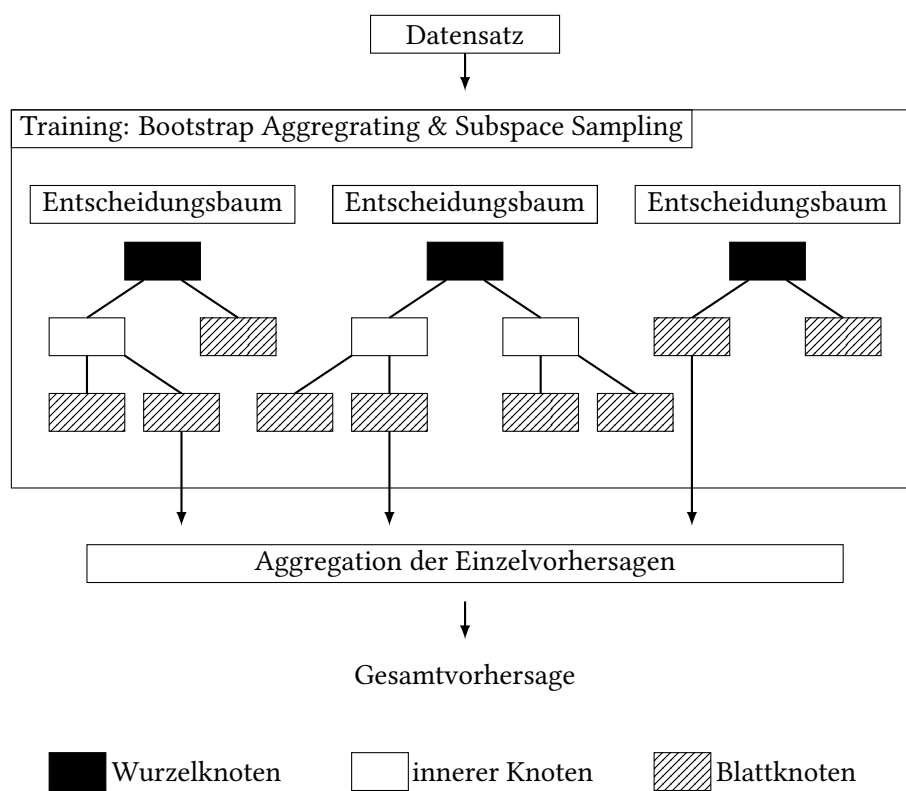


Abbildung 4.4: Aufbau eines RF-Modells, das aus drei Entscheidungsbäumen besteht. Abbildung abgeändert nach [KND15, S. 166].

Die Daten für das Training der einzelnen Entscheidungsbäume werden zufällig gewählt und können mehrmals gezogen werden (Ziehen mit Zurücklegen). Der Ziehvorgang wird so lange wiederholt, bis die Originalgröße des Trainingsdatensatzes erreicht wird. Dieses Verfahren wird als Bootstrap Aggregating (Bagging) bezeichnet. Aufgrund der zufällig gewählten Trainingsdaten ist jeder Entscheidungsbaum unterschiedlich aufgebaut. Neben Bagging wird beim RF das Verfahren Subspace Sampling angewendet, bei dem nur eine bestimmte Anzahl an zufällig

ausgewählten Merkmalen für das Training eines Entscheidungsbaums berücksichtigt wird. Die zufällige Wahl der Merkmale findet während der Aufteilung der Daten in einem Knoten eines Entscheidungsbaums statt [HTF09, S. 588]. Die Kombination von Bootstrap Aggregating und Subspace Sampling ist in Abbildung 4.5 dargestellt.

Der RF wird in [Bre01a] vorgestellt und greift nach [Lou14, S. 71–72] die in [Bre96; AGW97] vorgestellten Methoden auf. In dieser Arbeit wird ein RF-Regressionsmodell durch die RandomForestRegression-Klasse aus der scikit-learn Bibliothek [PVG+11] in der Version 1.3.2 umgesetzt.

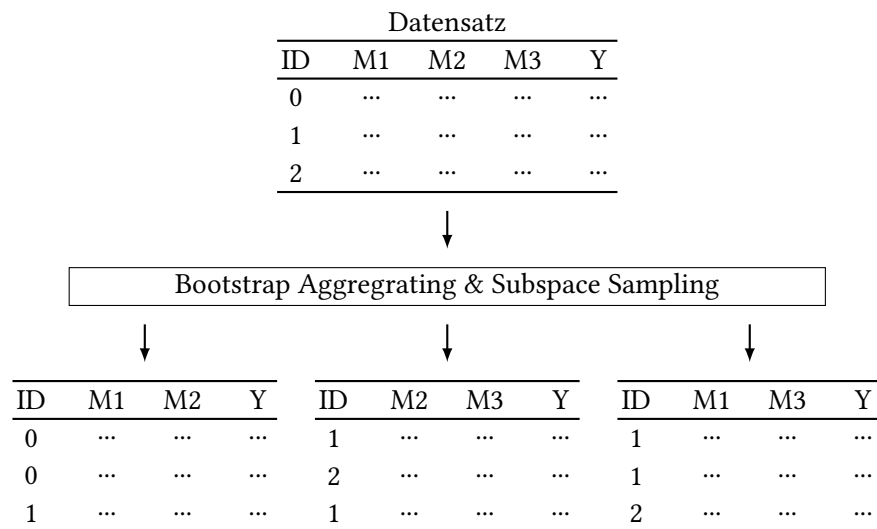


Abbildung 4.5: Bootstrap Aggregating (Bagging) und Subspace Sampling angewendet auf einen Datensatz, der die Merkmale M1, M2 und M3 und die Zielvariable Y beinhaltet. Abbildung abgeändert nach [KND15, S. 166].

4.3 Validieren und Testen

Das Training eines Vorhersagemodells erfolgt über einen Trainingsdatensatz, der ein Teil des gesamten Datensatzes ist. Die übrigen Daten sind der Testdatensatz, der für die abschließende Schätzung der tatsächlichen Modellperformanz dient [Nel20, S. 191]. Der Datensatz wird folglich in einen Trainings- und Testdatensatz eingeteilt.

Bei der Einteilung ist darauf zu achten, dass der Trainings- und Testdatensatz gleichermaßen die Betriebspunkte des Prozesses enthält. Wird der Trainingsdatensatz zu klein gewählt, wodurch Betriebspunkte des Prozesses im Trainingsdatensatz fehlen, so kann von einem Vorhersagemodell keine hohe Vorhersageperformanz erwartet werden. Bei einem zu klein gewählten Testdatensatz hingegen kann die Schätzung des Modellperformanz ungenau werden [Nel20, S. 191].

Des Weiteren ist zu beachten, dass der Trainingsdatensatz nicht für die Evaluation des Vorhersagemodells herangezogen wird, da dies zu einer zu optimistischen Schätzung der Modellperformanz führt [Ras18, S. 7]. Der Grund hierfür liegt darin, dass bei der Verwendung der Trainingsdaten lediglich der Verzerrungsfehler zu optimistisch geschätzt wird [Nel20, S. 184].

Der Verzerrungsfehler resultiert aus der Inflexibilität des Modells, das aufgrund der gewählten Modellstruktur nicht in der Lage ist, die Zusammenhänge in den Daten zu erlernen [Nel20, S. 178]. Ein Verzerrungsfehler liegt beispielsweise vor, wenn ein quadratischer Zusammenhang über ein lineares Modell beschrieben werden soll. Mit dem linearen Modell ist es nicht möglich, den quadratischen Zusammenhang in den Daten zu erlernen, auch wenn die Parameter des linearen Modells optimal ermittelt werden können. Folglich bleibt stets ein Verzerrungsfehler bestehen, der auf die gewählte Modellstruktur zurückzuführen ist. Inflexible Modelle erreichen daher auf den Trainings- und Testdaten eine schlechte Performanz. In diesem Fall wird von einer Unteranpassung (Underfitting) gesprochen [Nel20, S. 184; KND15, S. 11; RM17, S. 197].

Das Gegenteil einer Unteranpassung ist eine Überanpassung (Overfitting), die auf den Varianzfehler eines Modells zurückgeführt wird. In diesem Fall übersteigt der Varianzfehler den Verzerrungsfehler. Der Varianzfehler resultiert aus der Ungenauigkeit der geschätzten Modellparameter, die von den optimalen Modellparametern abweichen. Dies ist darauf zurückzuführen, dass die Trainingsdaten, aus denen die Modellparameter geschätzt werden, in der Regel verrauscht sind und nur in begrenztem Umfang zur Verfügung stehen [Nel20, S. 180]. Mit einer größeren Anzahl an Modellparametern wird das Modell flexibler, allerdings steigt auch die Ungenauigkeit der geschätzten Parameter, was zu einem höheren Varianzfehler führt.

Für überangepasste Modelle kann die Erhöhung der Trainingsdaten eine sinnvolle Vorgehensweise sein, um den Varianzfehler und damit die Überanpassung zu verringern [Nel20, S. 182]. In vielen Fällen ist eine Erhöhung an Trainingsdaten jedoch nicht realisierbar. Alternativ kann die Anzahl der Modellparameter verringert werden, um die Überanpassung zu reduzieren [Nel20, S. 181]. Es ist jedoch zu beachten, dass eine zu starke Verringerung der Parameteranzahl nicht in ein unterangepasstes Modell führt. An dieser Stelle ist der Konflikt zwischen einer Unteranpassung und Überanpassung ersichtlich. Eine simultane Reduzierung des Verzerrungs- und Varianzfehlers ist durch die Änderung der Parameteranzahl daher nicht möglich. Dieses Dilemma wird in der Literatur als Verzerrung/Varianz-Kompromiss (bias/variance tradeoff) bezeichnet [Nel20, S. 183].

4.3.1 Three-Way Holdout

Beim maschinellen Lernen kommen Algorithmen zum Einsatz, um Modelle anhand von Trainingsdaten anzupassen [Ras18, S. 6]. Solche Algorithmen besitzen Einstellungsparameter (Hyperparameter), um die Modellanpassung zu optimieren [Ras18, S. 20]. Bei der Anpassung stellt beispielsweise die maximale Baumtiefe eines Entscheidungsbaums einen Hyperparameter dar.

Die Anpassung der Hyperparameter erfolgt anhand einer Bewertung der Modellperformanz, die für unterschiedliche Hyperparameterkonfigurationen mit einem Teil des Trainingsdatensatzes ermittelt wird. Dieser Teil bildet den Validierungsdatensatz [KND15, S. 406].

Die Aufteilung in einen Trainings-, Validierungs- und Testdatensatz wird auch als Three-Way-Holdout-Methode bezeichnet [Ras18, S. 22]. In Abbildung 4.6 wird die Einteilung eines Datensatzes über die Three-Way-Holdout-Methode verdeutlicht, bei der zusätzlich die zeitliche Komponente der Daten (Out-of-Time Sampling) [KND15, S. 412] berücksichtigt wird.

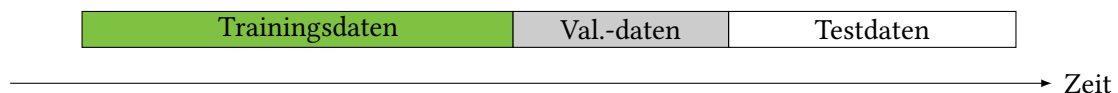


Abbildung 4.6: Einteilung der Daten mit der Three-Way-Holdout-Methode in einen Trainings-, Validierungs- und Testdatensatz. Die zeitliche Komponente der Daten (Out-of-Time Sampling) [KND15, S. 412] wird dabei berücksichtigt.

In Bezug auf das Größenverhältnis von Trainings-, Validierungs- und Testdaten bestehen keine festen Vorgaben. In der Literatur werden Verhältnisse wie 5:2:3, 4:2:4 [KND15, S. 406] und 6:2:2 [Fro18, S. 198] angegeben. Die Trainings- und Validierungsdaten dienen der Anpassung von Vorhersagemodellen, während der Testdatensatz für die abschließende Evaluation des besten Vorhersagemodells herangezogen wird. Ein abschließender Test ist weiterhin erforderlich, da die Validierungsdaten aufgrund der Anpassung der Hyperparameter in das Training „durchsickern“ können und somit nicht mehr für die Abschätzung einer realistischen Modellperformanz verwendet werden können [Ras18, S. 22].

Die Three-Way-Holdout-Methode findet Anwendung, wenn der Datensatz eine ausreichende Größe besitzt, sodass sowohl der Trainings- als auch der Validierungs- und Testdatensatz die Betriebspunkte eines Prozesses gleichermaßen beinhalten. Sofern eine begrenzte Datenmenge zur Verfügung steht, bieten sich andere Validierungsmethoden im Bereich der Zeitreihenanalyse an, um eine realistische Modellperformanz abschätzen zu können [BHK18, S. 70; CTM20, S. 1998]. In [CTM20] stellen Cerqueira et al. unterschiedliche Validierungsmethoden im Bereich der Zeitreihenvorhersage vor, die im Hinblick auf die Abschätzung der Modellperformanz empirisch untereinander verglichen werden. Cerqueira et al. weisen darauf hin, dass die Wahl einer Validierungsmethode nicht trivial ist, da die Beobachtungen einer Zeitreihe aufgrund ihrer zeitlichen Anordnung in der Regel nicht unabhängig voneinander sind [CTM20, S. 1997–1998].

Anhand von 174 realen Zeitreihen zeigen Cerqueira et al. auf, dass sich keine bestimmte Methode von den untersuchten Validierungsmethoden für alle betrachteten Zeitreihen eignet. Sie machen darauf aufmerksam, weitere Untersuchungen durchzuführen, in denen Zeitreihenmerkmale (siehe Kapitel 4.1.2) in Zusammenhang mit verschiedenen Validierungsmethoden gesetzt werden [CTM20, S. 2012]. Im Fazit empfehlen Cerqueira et al. jedoch für stationäre Zeitreihen die Methode Blocked Cross-Validation (CV-BI) und für nicht-stationäre Zeitreihen

die Methode Repeated Holdout (Rep-Holdout) [CTM20, S. 2019–2020]. Die folgenden Validierungsmethoden werden im Rahmen der Modellselektion vorgestellt. Dabei werden die Hyperparameter eines Modells innerhalb der entsprechenden Validierungsmethode solange variiert, bis die bestmögliche Modellperformanz ermittelt wurde.

4.3.2 Blocked Cross-Validation (CV-BI)

Die Blocked Cross-Validation (CV-BI) ähnelt der klassischen Cross-Validation (CV) [CTM20, S. 2002]. Bei beiden Methoden erfolgt eine Einteilung der Beobachtungen in k -Blöcke. Der wesentliche Unterschied zwischen der CV und der CV-BI besteht darin, dass bei der CV-BI die Beobachtungen nicht zufällig gemischt werden. Die zeitliche Anordnung der Beobachtungen einer Zeitreihe bleibt bei der CV-BI erhalten [CTM20, S. 2002]. Jeder Block wird anschließend einmalig als Validierungsdatensatz verwendet, während die restlichen $k - 1$ Blöcke als Trainingsdaten für das entsprechende Modell verwendet werden. Insgesamt werden k Durchläufe durchgeführt, in denen unterschiedliche Trainings- und Validierungsdaten verwendet werden. Abschließend werden die anhand der k Validierungsdatensätze ermittelten Performanzwerte aggregiert, um einen endgültigen Performanzwert zu erhalten [RM17, S. 191–192; KND15, S. 408–410; Nel20, S. 193]. Abbildung 4.7 zeigt den Ablauf der CV-BI mit $k = 10$.

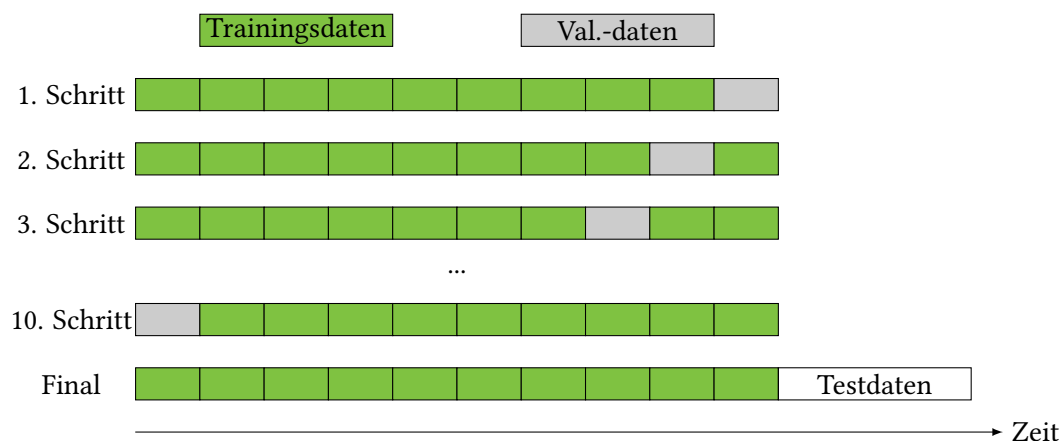


Abbildung 4.7: Ablauf der CV-BI mit $k = 10$. Die zeitliche Anordnung der Beobachtungen bleibt bei der CV-BI erhalten. In einem finalen Schritt werden alle Trainingsdaten für das Training des Modells mit der höchsten Performanz verwendet, das für einen abschließenden Test auf die Testdaten angewendet wird. Abbildung abgeändert nach [CTM20, S. 2001; RM17, S. 192].

Die Hyperparameter eines Modells werden solange variiert, bis die bestmögliche Modellperformanz innerhalb der CV-BI ermittelt wurde. Im finalen Schritt wird das Modell mit der höchsten Performanz mit den gesamten Trainingsdaten erneut trainiert. Der Vorteil für diesen abschließenden Trainingsvorgang ist, dass alle Trainingsdaten verwendet werden, was in der Regel zu

einer verbesserten Modellperformanz führt. Anschließend wird das Modell auf den separaten Testdatensatz angewendet, um eine abschließende Modellperformanz zu ermitteln [Ras18, S. 30; RM17, S. 192].

4.3.3 Repeated Holdout (Rep-Holdout)

Repeated Holdout (Rep-Holdout) basiert auf der wiederholten Anwendung von Out-of-Time Sampling [KND15, S. 412]. Die Methode ist nach [CTM20, S. 2018–2020] empfehlenswert, wenn nicht stationäre Zeitreihen vorliegen. Innerhalb eines vorher festgelegten Zeitfensters wird zufällig ein Zeitpunkt a gewählt, der die Zeitreihe in einen Trainings- und Validierungsdatensatz einteilt. Diese Vorgehensweise wird mehrmals durchgeführt. In dem Experiment von Cerqueira et al. wurden 10 Durchläufe durchgeführt [CTM20, S. 2009]. Abbildung 4.8 zeigt die Rep-Holdout-Methode für einen Durchlauf.

Die in den Durchläufen ermittelten Performanzwerte werden im Anschluss zu einem gesamten Wert aggregiert. Der Gesamtwert dient einem Vergleich mit weiteren Modellen, die sich durch abweichende Hyperparameter auszeichnen. Das Modell mit der bestmöglichen Performanz kann anschließend, wie bereits bei der CV-BI erwähnt, mit den gesamten Trainings-, und Validierungsdaten trainiert werden. Die finale Modellperformanz kann auf einem separaten Testdatensatz berechnet werden.

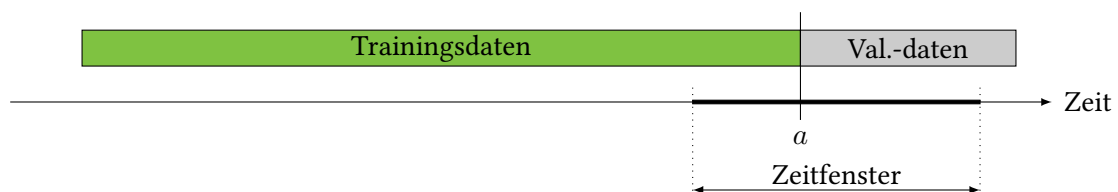


Abbildung 4.8: Vorgehensweise der Rep-Holdout-Methode. Aus einem vorher festgelegten Zeitfenster wird zufällig ein Zeitpunkt a gewählt, der die Zeitreihe in einen Trainings- und Validierungsdatensatz einteilt. Die Einteilung wird mehrmals durchgeführt. Die ermittelten Performanzwerte werden anschließend zu einem Gesamtwert aggregiert. Abbildung abgeändert nach [CTM20, S. 2000].

4.3.4 Prequential-Methode

Zur Vollständigkeit wird die in [CTM20, S. 2000–2001] genannte Prequential-Methode erwähnt, die bereits in unserer Arbeit [GKD+18] für die Gegenwartsvorhersage der Badtemperatur verwendet wurde. Bei der Prequential-Methode werden die Beobachtungen einer Zeitreihe in Blöcke unterteilt, wobei die zeitliche Anordnung der Beobachtungen erhalten bleibt. Anschließend werden die in den ersten Blöcken enthaltenen Beobachtungen als Trainingsdaten verwendet, während die Beobachtungen in dem darauffolgenden Block als Validierungsdaten dienen. Nach der erfolgten Validierung werden die Trainingsdaten um den nächsten Block

erweitert. Die Beobachtungen des darauffolgenden Blocks werden für die Validierung verwendet. Dieses Vorgehen wird so lange wiederholt, bis der letzte Block für die Validierung verwendet wurde [CTM20, S. 2000]. Die in den einzelnen Schritten erzielten Performanzwerte werden zu einem Gesamtwert aggregiert. Anschließend durchläuft ein weiteres Modell mit abweichenden Hyperparametern die einzelnen Schritte. Das Modell, das die bestmögliche Performanz erreicht, wird in einem finalen Schritt auf den gesamten Trainingsdaten trainiert, um eine abschließende Modellperformanz auf den Testdaten zu erhalten [Ras18, S. 30]. Abbildung 4.9 zeigt den Ablauf der Prequential-Methode. Des Weiteren existieren diverse Variationen der Prequential-Methode, die in [CTM20, S. 2000–2001] aufgezeigt werden.

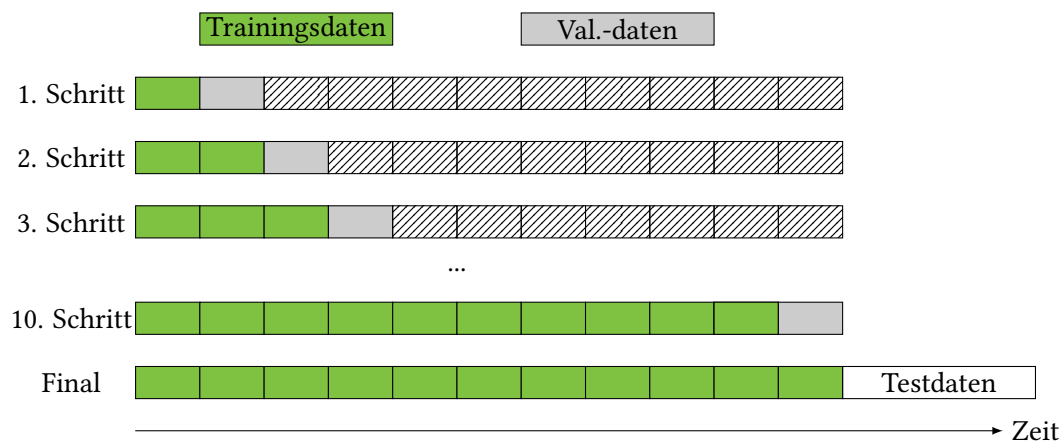


Abbildung 4.9: Vorgehensweise der Prequential-Methode mit elf Blöcken. Die zeitliche Anordnung der Beobachtungen bleibt bei dieser Methode erhalten. Die schraffierten Blöcke werden in dem jeweiligen Schritt nicht berücksichtigt. In einem finalen Schritt werden alle Trainingsdaten für das Training des Modells mit der höchsten Performanz verwendet, das für einen abschließenden Test auf die Testdaten angewendet wird. Abbildung geändert nach [CTM20, S. 2000; GKD+18, S. 5].

4.4 Zusammenfassung

In diesem Kapitel wurden Grundlagen der Zeitreihenanalyse dargestellt, die für die Proof-of-Concept-Phase in Kapitel 5 genutzt werden. Zunächst wurde die Zeitreihenzerlegung vorgestellt, mit der sich eine Zeitreihe in eine Trend-Zyklus-Komponente, Saison- und Restkomponente zerlegen lässt. Mithilfe dieser Komponenten können Eigenschaften einer Zeitreihe berechnet werden, wie beispielsweise die Stärke der Saisonalität einer Zeitreihe. Die numerische Beschreibung einer bestimmten Zeitreiheigenschaft wird auch als Zeitreihenmerkmal bezeichnet. Anhand einer Literaturübersicht wurde herausgestellt, dass Zeitreihenmerkmale unter anderem in Kombination mit gleitenden Fenstern in unterschiedlichen Anwendungsbereichen eingesetzt

werden. Die Berechnung von Zeitreihenmerkmalen in gleitenden Fenstern ist Bestandteil für die Gegenwartsvorhersage (Nowcasting) der Badtemperatur in dieser Arbeit.

Um eine Zeitreihe auf Stationarität zu überprüfen, können die Stärke der Saisonalität und der KPSS-Test angewendet werden, die in diesem Kapitel beschrieben wurden und im praktischen Teil dieser Arbeit auf die Badtemperaturverläufe angewendet werden. Darüber hinaus wurden die AKF und PAKF vorgestellt. Beide Funktionen werden im nachfolgenden Kapitel verwendet, um die Badtemperaturverläufe hinsichtlich ihrer Autokorrelationen näher zu untersuchen. Das vorgestellte Periodogramm unterstützt zudem bei der Analyse von Periodizitäten in den Badtemperaturverläufen.

Der Unterschied zwischen einem lokalen und globalen Vorhersagemodell wurde aufgezeigt. Während bei einer lokalen Vorhersage für jede Zeitreihe ein eigenständiges Vorhersagemodell trainiert wird, wird bei der globalen Vorhersage ein einziges Modell mit allen verfügbaren Zeitreihen trainiert. In diesem Zuge wurden das AR-Modell und RF-Modell vorgestellt und Begründungen für die Wahl dieser Modelle dargestellt. Beide Modelle werden in dieser Arbeit für eine globale Vorhersage der Badtemperatur eingesetzt. Abschließend wurden mehrere Methoden aufgezeigt, um Vorhersagemodelle im Kontext der Zeitreihenanalyse zu validieren. Die Wahl einer Validierungsmethode findet im praktischen Teil statt, nachdem die Badtemperaturverläufe auf Stationarität überprüft wurden.



KAPITEL 5

Proof of Concept

In diesem Kapitel werden die anhand der Proof-of-Concept-Phase [BP20, S. 9] durchgeführten Schritte vorgestellt, die für die Implementierung der Lösungsansätze abgearbeitet wurden. Die Proof-of-Concept-Phase ist von einer iterativen Vorgehensweise gekennzeichnet, mit der die zuvor definierten Lösungsansätze umgesetzt und überprüft werden. In Abbildung 5.1 ist diese Vorgehensweise nach [BP20, S. 9] dargestellt. Die Proof-of-Concept-Phase beinhaltet den Austausch mit Prozessexperten/-innen über Hypothesen und Lösungsansätze, die Datenbereitstellung, die Lösungsentwicklung sowie die Evaluation der entwickelten Lösungen.

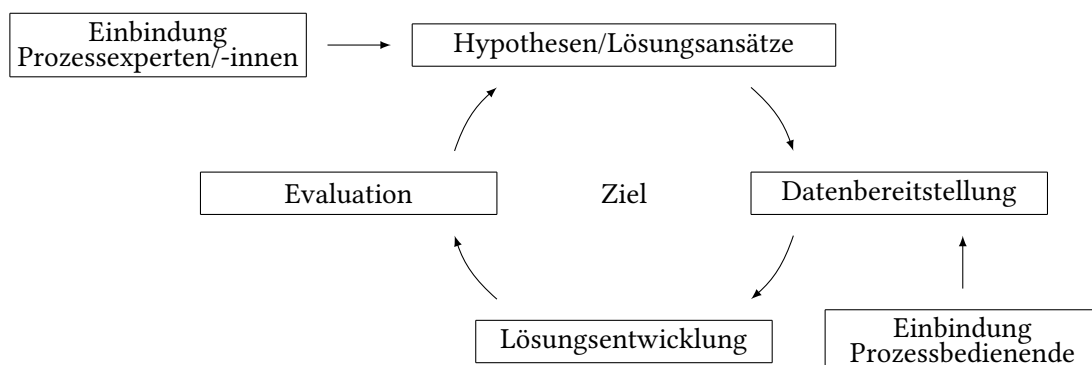


Abbildung 5.1: Iterative Vorgehensweise innerhalb der Proof-of-Concept-Phase. Abbildung abgeändert nach [BP20, S. 9].

Zunächst werden in diesem Kapitel der Datenfluss zwischen der TAE und dem LfA durch das angepasste initiale ML-Pipeline-Diagramm aufgezeigt und die Implementierung der Lösungsansätze erörtert. Im Rahmen der Datenbereitstellung wurden die von der TAE bereitgestellten Badtemperaturdaten aufbereitet und analysiert. Die aufbereiteten Daten dienten der anschließenden Entwicklung von Lösungen zur Vorhersage der Badtemperatur. Die Vorhersage wurde in eine Tagesprognose (Forecasting) und in eine Gegenwartsvorhersage (Nowcasting) aufgeteilt. Innerhalb der Evaluation erfolgte eine Gegenüberstellung der Performanz der Vorhersagemodelle. Insgesamt wurde die Proof-of-Concept-Phase in dieser Arbeit einmal durchlaufen.

Zeitangaben, die in der koordinierten Weltzeit (Coordinated Universal Time (UTC)) angegeben sind, haben in diesem Kapitel den Zusatz UTC. Alle anderen Zeitangaben sind in der Zeitzone Europa/Berlin angegeben.

5.1 Hypothesen/Lösungsansätze

Für die Vorhersage der Badtemperatur wurden für diese Arbeit zwei Lösungsansätze analysiert, deren Vorgehensweisen in Kapitel 2 beschrieben ist. Die Wahl der globalen Vorhersagemodelle wurde in Kapitel 4.2 begründet. Der erste Lösungsansatz ist ein Vergleich zwischen einem globalen AR-Modell und ausgewählten lokalen Vorhersagemodellen, mit denen eine Tagesprognose (Forecasting) der Badtemperatur erreicht wird.

Die Erstellung von Gegenwartsvorhersagen (Nowcasting) der Badtemperatur wird mit dem zweiten Lösungsansatz erzielt. Dazu wurde ein globales RF-Regressionsmodell trainiert und validiert. Für das Training des RF-Regressionsmodells wurden die in Kapitel 3.2 beschriebenen Prozessvariablen berücksichtigt, die einen wesentlichen Einfluss auf die Badtemperatur ausüben. In Abbildung 5.2 ist das angepasste ML-Pipeline-Diagramm abgebildet, das den Datenfluss zwischen der TAE und dem LfA darstellt. Im Vergleich zum initialen ML-Pipeline-Diagramm in Abbildung 2.1 ist die Berücksichtigung der Badtemperaturvorhersagen in die bestehenden Ofenregelung nicht eingezeichnet, da der Fokus dieser Arbeit auf der Entwicklung von Vorhersagemodellen liegt.

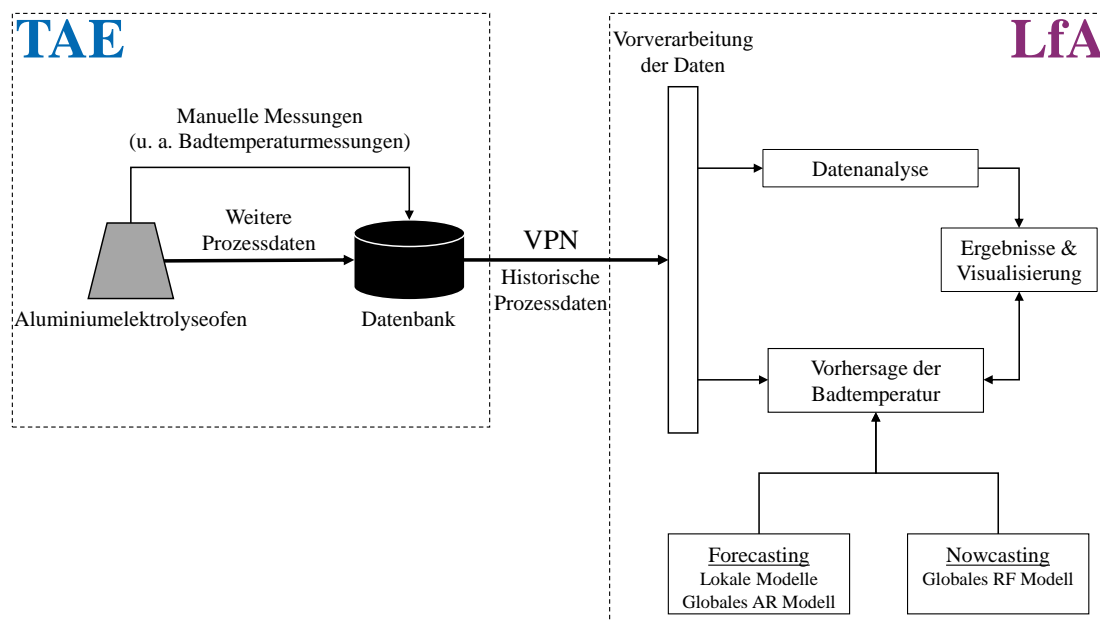


Abbildung 5.2: Angepasstes ML-Pipeline-Diagramm für die Vorhersage der Badtemperatur.

Auf die historischen Prozessdaten wurde über ein virtuelles privates Netzwerk (VPN) zwischen der TAE und dem LfA zugegriffen. Wie in Kapitel 2 beschrieben, zeichnet die TAE die Betriebsdaten der einzelnen Aluminiumelektrolyseöfen auf und überträgt diese in eine Datenbank. Auch manuelle Messungen, wie die Messung der Badtemperatur eines Ofens, werden in einer Datenbank abgespeichert. Der Zugriff auf die historischen Prozessdaten der TAE erlaubte eine anschließende Weiterverarbeitung und Analyse der Daten. Die Analyseergebnisse wurden für die Entwicklung der zuvor beschriebenen Modelle eingesetzt, die eine Badtemperaturvorhersage ermöglichen.

5.2 Datenbereitstellung

Für die Badtemperaturvorhersage wurden die Badtemperaturdaten der TAE aus dem Zeitraum vom 01.01.2022 06:00 Uhr bis zum 24.04.2023 06:00 Uhr von den Öfen in Halle 1 aufbereitet und analysiert. Dabei wurden ausschließlich fehlerfreie Messungen berücksichtigt. Im ersten Schritt der Datenbereitstellung wurden die Zeitstempel der einzelnen Badtemperaturverläufe untersucht. Daraus geht hervor, dass Badtemperaturmessungen an einem Ofen nicht exakt zur gleichen Uhrzeit stattgefunden haben und auch mehrfach innerhalb einer Schicht aufgetreten sind.

Des Weiteren erfolgte eine Verknüpfung der Badtemperaturdaten mit weiteren Betriebsdaten, um ausgeschaltete Öfen zu identifizieren und Rückschlüsse auf das Ofenalter zu ziehen. Öfen, die innerhalb des Analysezeitraums erst kürzlich in Betrieb genommen wurden oder ausgeschaltet waren, wurden in dieser Arbeit nicht berücksichtigt.

5.2.1 Anpassung der Zeitstempel

Aufgrund des manuellen Messvorgangs der Badtemperatur finden die täglichen Messungen an einem Ofen nicht exakt zur gleichen Uhrzeit statt. Daher konnte nicht davon ausgegangen werden, dass die zeitlichen Abstände der Badtemperaturmessungen eines Ofens immer gleich sind. Der Arbeitstag der TAE besteht aus drei Schichten, wobei die erste Schicht regulär um 6 Uhr, die zweite Schicht um 14 Uhr und die dritte Schicht um 22 Uhr beginnt. In den Abbildungen 5.3 und A.1 ist die zeitliche Variation der Messzeitpunkte für die Öfen aus Halle 1 zu sehen. Dazu wurde jeweils die zeitliche Differenz zwischen dem Messzeitpunkt (UTC) der Badtemperatur und dem Beginn der ersten Schicht berechnet.

Die Abbildungen 5.3 und A.1 veranschaulichen die Zeiten der drei Schichten anhand der verschobenen Boxplots. Aus diesen geht hervor, dass im Untersuchungszeitraum die Badtemperaturmessungen für die Öfen 1081 bis 1120 von der ersten Schicht, für die Öfen 1041 bis 1080 von der zweiten Schicht und für die Öfen 1001 bis 1040 von der dritten Schicht durchgeführt wurden. Es gibt Messungen, die außerhalb der regulären Schicht stattgefunden haben und zum Teil als Ausreißer in den Boxplots dargestellt sind. Gründe für diese nicht regulären Messungen

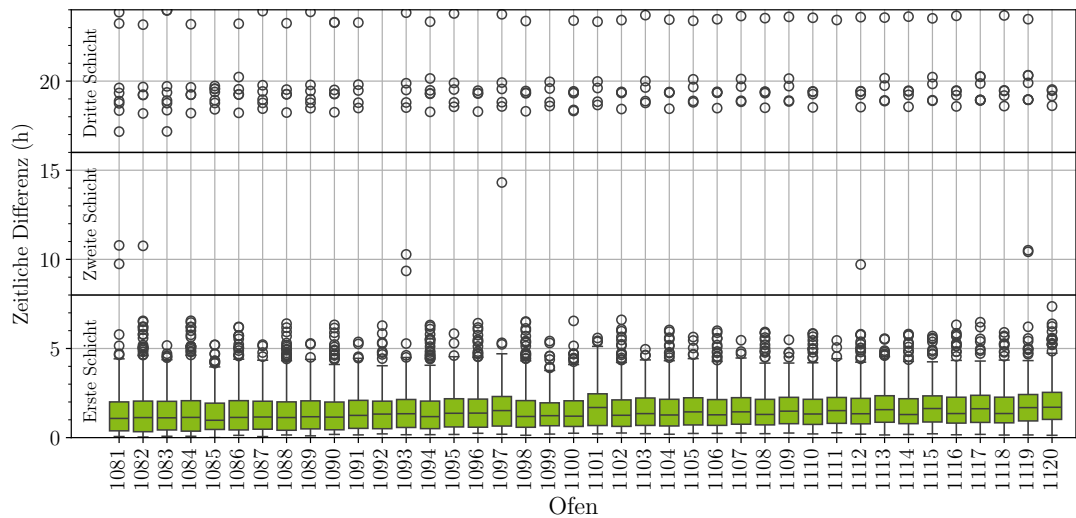


Abbildung 5.3: Die Boxplots zeigen die Verteilung der zeitlichen Differenzen für die Öfen der ersten Schicht. Die Differenzen wurden jeweils zwischen dem Messzeitpunkt (UTC) der Badtemperatur und dem Beginn der ersten Schicht berechnet.

könnten Kontrollen von auffälligen Öfen mit hoher Badtemperatur gewesen sein oder Versuchsdurchführungen, bei denen zusätzliche Messungen stattgefunden haben. Als Beispiel für eine Versuchsdurchführung kann die Untersuchung der Auswirkung einer Hallenschaltung auf die Badtemperatur genannt werden, die in Kapitel 3.2 beschrieben wurde. Des Weiteren kann der Ofen 1041 genannt werden, an dem am 08.03.2022 und 09.03.2022 zusätzliche Messungen durchgeführt wurden, die in Tabelle 5.1 ersichtlich sind.

Zeitstempel (UTC)	Badtemperatur (°C)
06.03.2022 17:44:00	980,4
07.03.2022 15:03:00	977,5
08.03.2022 16:27:00	978,5
08.03.2022 22:40:00	977,8
09.03.2022 00:25:00	979,6
09.03.2022 01:39:00	977,6
09.03.2022 02:23:00	978,5
09.03.2022 03:39:00	977,9
09.03.2022 04:06:00	980,2
09.03.2022 14:45:54	975,8

Tabelle 5.1: Zusätzliche Badtemperaturmessungen, die am 08.03.2022 und 09.03.2022 am Ofen 1041 durchgeführt wurden.

Anhand der Abbildungen 5.3 und A.1 sowie der Tabelle 5.1 ist ersichtlich, dass die Messungen an einem Ofen nicht zur gleichen Uhrzeit stattgefunden haben und somit die zeitlichen Abstände

zwischen den Messungen variieren. Zeitreihen, deren Beobachtungen mit ungleichmäßigen Zeitabständen aufgezeichnet werden, werden als nicht reguläre Zeitreihen bezeichnet. Sie können im Bereich der Zeitreihenanalyse zu Schwierigkeiten führen, da zahlreiche State-Of-The-Art-Methoden reguläre Zeitreihen voraussetzen [CCV21, S. 1]. Eine Möglichkeit, um mit dieser Problematik umzugehen, ist die Anpassung der Messzeitstempel, sodass diese zueinander einen gleichmäßigen zeitlichen Abstand aufweisen.

In einem ersten Schritt wurden außerordentliche Badtemperaturmessungen entfernt. Lediglich die Messungen, die von der zuständigen regulären Schicht durchgeführt wurden, wurden in die weiteren Untersuchungen einbezogen. Beispielsweise wurden nur die Messungen vom Ofen 1041 berücksichtigt, die innerhalb des zweiten Schichtzeitraums stattgefunden haben. Für die Messungen am 08.03.2022 und 09.03.2022 in Tabelle 5.1 bedeutet das, dass lediglich die Messungen am 08.03.2022 um 16:27 Uhr (UTC) und am 09.03.2022 um 14:45 Uhr (UTC) weiter analysiert wurden. Alle restlichen Messungen, die an diesen beiden Tagen stattgefunden haben, liegen außerhalb der zweiten Schicht (13:00 Uhr bis 21:00 Uhr (UTC)) und wurden von der weiteren Analyse ausgeschlossen.

Bei Mehrfachmessungen innerhalb einer Schicht wurde die letzte (späteste) Badtemperaturmessung betrachtet. Das lässt sich damit begründen, dass eine erneuerte Messung durchgeführt wird, wenn Zweifel an der ersten Messung bestehen. Daher wurde für die weitere Vorgehensweise lediglich die letzte (späteste) Messung in einer Schicht berücksichtigt.

Das Verwerfen von Messungen hatte jedoch zur Folge, dass fehlende Einträge im Datensatz erzeugt wurden. Fehlende Einträge treten auch dann auf, wenn Messungen an einem Ofen nicht durchgeführt oder nicht in die entsprechende Datenbank übertragen wurden. Zum Beispiel erzeugen Öfen, die nicht im Betrieb sind, fehlende Einträge im Datensatz, da an abgeschalteten Öfen keine Messungen stattfinden. Tabelle 5.2 beinhaltet Messeinträge vom Ofen 1090 aus dem Jahr 2022. Aus dieser geht hervor, dass für den 12.01.2022 und den 16.01.2022 jeweils kein Messeintrag vorhanden ist. Für die weitere Vorgehensweise wurden Messeinträge mit fehlendem Datum ergänzt. Der zugehörige Badtemperaturwert wurde als fehlend gekennzeichnet (n.a.).

Für die Tagesprognose der Badtemperatur spielte die Uhrzeit der jeweiligen Badtemperaturmessung keine Rolle. Lediglich das Datum einer Messung war von Interesse. Daher wurde nur das Datum der jeweiligen Zeitstempel berücksichtigt. Das Datum, an dem jeweils die erste Schicht begonnen hat, wurde als Referenzdatum verwendet und jeder Badtemperaturmessung in den folgenden drei Schichten zugeordnet. Das Referenzdatum wurde auch den Messungen aus der dritten Schicht zugeordnet, die nach Mitternacht stattfanden. Zum Beispiel wurde am 11.01.2022 um 21:35 Uhr (UTC) eine Badtemperaturmessung am Ofen 1001 durchgeführt. Die nächste reguläre Messung fand am 13.01.2022 um 00:43 Uhr (UTC) am Ofen 1001 statt. Die Messung am 11.01.2022 behielt das Datum bei, da der Beginn der vorherigen ersten Schicht am

Datum	Badtemperatur (°C)
10.01.2022	953,4
11.01.2022	958,4
13.01.2022	956,1
14.01.2022	948,8
15.01.2022	954,0
17.01.2022	952,5
18.01.2022	957,2

Tabelle 5.2: In der Tabelle liegen einige Badtemperaturmessungen aus dem Jahr 2022 vor, die am Ofen 1090 durchgeführt wurden. Für den 12.01.2022 und den 16.01.2022 ist jeweils kein Messeintrag vorhanden.

11.01.2022 um 05:00 Uhr (UTC) war. Die Messung am 13.01.2022 erhielt hingegen das Datum 12.01.2022, da die vorherige erste Schicht am 12.01.2022 um 05:00 Uhr (UTC) begann.

Aus Tabelle 5.3 ist ersichtlich, dass nun für den 12.01.2022 und 16.01.2022 jeweils ein Eintrag vorhanden ist. Der Badtemperaturwert an den beiden Tagen ist als fehlend mit n.a. (nicht angegeben) gekennzeichnet.

Datum	Badtemperatur (°C)
10.01.2022	953,4
11.01.2022	958,4
12.01.2022	n.a.
13.01.2022	956,1
14.01.2022	948,8
15.01.2022	954,0
16.01.2022	n.a.
17.01.2022	952,5
18.01.2022	957,2

Tabelle 5.3: Die zuvor fehlenden Einträge für den Ofen 1090 sind nun für den 12.01.2022 und den 16.01.2022 in der Tabelle vorhanden, wobei der jeweilige Badtemperaturwert als fehlend mit n.a. (nicht angegeben) gekennzeichnet wurde.

5.2.2 Verknüpfung mit weiteren Betriebsdaten

Um Rückschlüsse darüber zu erhalten, ob fehlende Badtemperaturmessungen aufgrund von abgeschalteten Öfen zustande gekommen sind, wurde der Betriebszustand der Öfen mit den entsprechenden Badtemperaturmessungen verknüpft. Außerdem wurde das Alter der einzelnen Öfen betrachtet. Neu angefahrne Öfen werden als Early Operation Pots bezeichnet und besitzen andere Sollwerteinstellungen als Öfen, die schon länger in Betrieb sind. Aus diesen Gründen ist es sinnvoll, für Analysen das Ofenalter zu berücksichtigen, da junge Öfen ein anderes Verhalten als alte aufweisen [KGD+21, S. 768].

In dieser Arbeit wurden Öfen ausgeschlossen, die im Analysezeitraum außer Betrieb waren. Das hatte den Vorteil, dass nahezu vollständige Badtemperaturverläufe analysiert und miteinander verglichen werden konnten. So wurde die weitere Vorverarbeitung der Daten vereinfacht, da die Tage, an denen die Öfen nicht in Betrieb waren, nicht mehr gesondert berücksichtigt werden mussten. Da sich nach [KGD+21, S. 768] das Verhalten älterer Öfen von jungen Öfen unterscheidet, wurden für die weitere Untersuchung nur Öfen betrachtet, die im Analysezeitraum älter als 100 Tage waren. Insgesamt führte dieser Ausschluss dazu, dass die Badtemperaturdaten von 60 Öfen in der vorliegenden Arbeit weiter untersucht wurden.

5.2.3 Zusammenfassung

Mit den in diesem Kapitel aufgezeigten Vorverarbeitungsschritten wurden aus den Badtemperaturdaten der einzelnen Öfen vollständige Zeitreihen erzeugt. Die Abbildung 5.4 zeigt jeweils einen Ausschnitt des Badtemperaturverlaufs von dem Ofen 1001 vor und nach der Anpassung der Daten über die in diesem Kapitel aufgezeigten Schritte. In dieser sind die Mehrfachmessungen zwischen dem 06.09.2022 und dem 09.09.2022 vor der Anpassung der Badtemperaturdaten ersichtlich, die nach der Anpassung nicht mehr vorhanden sind.

Für die Tagesprognose spielen die Uhrzeiten der einzelnen Messungen keine wesentliche Rolle. Daher wurden die Uhrzeiten der Messungen verworfen, da lediglich der Tag der Messung relevant war. Die originalen Messzeitpunkte blieben jedoch für weitere Analysen und die Gegenwartsvorhersage (Nowcasting) im Datensatz erhalten.

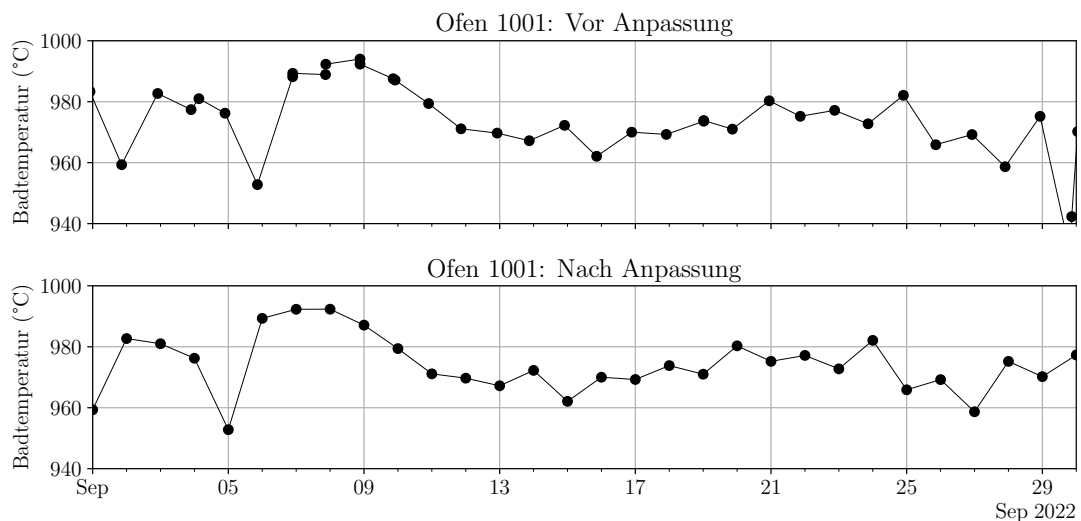


Abbildung 5.4: Der Badtemperaturverlauf des Ofens 1001 ab dem 01.09.2022 bis zum 30.09.2022 vor und nach der Anpassung der Badtemperaturdaten.

5.3 Zeitreihenanalyse der Badtemperaturdaten

Im weiteren Verlauf dieser Arbeit wurde der komplette Datensatz in einen Trainings- und Testdatensatz aufgeteilt. Anschließend wurden die Trainingsdaten näher untersucht. Zunächst wurde die Verteilung der Badtemperaturdaten mithilfe von Boxplots dargestellt und Ausreißer identifiziert. Bei den identifizierten Ausreißern wurde zwischen plausiblen und nicht plausiblen Ausreißern unterschieden. Der Unterschied wird in diesem Kapitel erläutert. Besonderer Fokus lag auf starken Badtemperaturänderungen, die hinsichtlich ihrer Plausibilität zusammen mit der TAE untersucht wurden.

Anschließend wurden vorhandene Datenlücken in den Badtemperaturdaten untersucht und mithilfe einer geeigneten Methode geschlossen. Zudem wurde die (partielle) Autokorrelationsfunktion für die Badtemperaturverläufe berechnet. Um Periodizitäten in den Badtemperaturverläufen zu identifizieren, wurde von jedem Badtemperaturverlauf jeweils ein Periodogramm berechnet. Durch die Berechnung der Stärke der Saisonalität und die Durchführung des KPSS-Tests wurden die Badtemperaturverläufe auf Stationarität überprüft.

5.3.1 Trainings- und Testdaten

Für die Vorhersage der Badtemperatur wurde zunächst der gesamte Datensatz in einen Trainings- und Testdatensatz unterteilt. Die Trainingsdaten wurden für die Datenanalyse und das Training der Vorhersagemodelle verwendet, während der Testdatensatz ausschließlich als abschließender Test für die entwickelten Vorhersagemodelle eingesetzt wurde. Alle auf Basis der Trainingsdaten getroffenen Entscheidungen wurden auf die Testdaten angewendet. Informationen über die Einteilung von Trainings- und Testdatensätzen wurden bereits in Kapitel 4.3 erläutert.

Aufgrund der zeitlichen Anordnung der Badtemperaturdaten wurde der in Kapitel 5.2 vorverarbeitete Datensatz mit einer zeitlichen Trennung in einen Trainings- und Testdatensatz aufgeteilt (vgl. Kapitel 4.3.1). Die in dieser Arbeit gewählte Aufteilung ist in Abbildung 5.5 dargestellt. Für die Trennung wurde der Zeitpunkt 01.01.2023 gewählt. Datenpunkte vor dem 01.01.2023 wurden dem Trainingsdatensatz zugeordnet. Alle bis zum 23.04.2023 übrigen Datenpunkte bildeten den Testdatensatz. Die Einteilung wurde so gewählt, um ein komplettes Jahr in den Trainingsdaten zu berücksichtigen. Für die Gegenwartsvorhersage (Nowcasting) wurde außerdem ein Teil der Trainingsdaten als Validierungsdaten eingesetzt. Details dazu folgen in Kapitel 5.4.2. Die nachfolgenden Auswertungen erfolgten ausschließlich auf den Trainingsdaten.

5.3.2 Verteilung und Ausreißer

Um einen Überblick über die Verteilung der Badtemperaturdaten zu erhalten, eignet sich ein Boxplot. In Abbildung 5.6 ist ein Boxplot dargestellt, dessen statistische Kennwerte anhand der

Abschnitt 5.3. Zeitreihenanalyse der Badtemperaturdaten

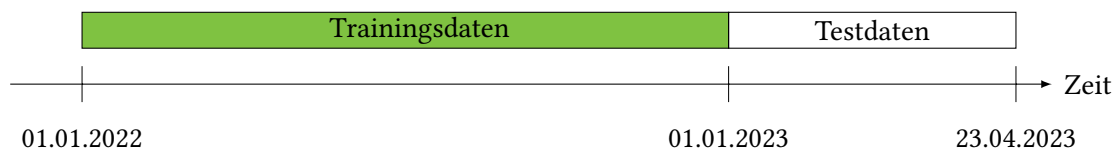


Abbildung 5.5: Zeitliche Aufteilung der Badtemperaturdaten in einen Trainings- und Testdatensatz.

Trainingsdaten berechnet wurden. Folgende Werte wurden auf eine Nachkommastelle gerundet. Aus dem Boxplot geht hervor, dass der Median der Badtemperatur $963,1^{\circ}\text{C}$ beträgt. 50 % der Badtemperaturwerte liegen im Bereich von $956,6^{\circ}\text{C}$ bis $970,9^{\circ}\text{C}$. Der Interquartilabstand (IQA) beträgt $14,3^{\circ}\text{C}$. Der untere Whisker ist bei $935,2^{\circ}\text{C}$ und der obere Whisker bei $992,1^{\circ}\text{C}$ eingezeichnet (die Proportion des IQA beträgt 1,5). Der Mittelwert der Badtemperatur liegt bei $963,7^{\circ}\text{C}$. Die Werte streuen mit einer Standardabweichung von $9,9^{\circ}\text{C}$ um den Mittelwert. Zum Vergleich sei auf die Literatur [TCY13, S. 58] verwiesen, in der angegeben wird, dass die Badtemperatur bei den besten Aluminiumhütten der Welt mit einer Standardabweichung von 5°C bis 6°C streut. Auffällig ist der minimale Wert von $872,5^{\circ}\text{C}$, der für den Ofenbetrieb nicht plausibel erscheint und vermutlich von einer fehlerhaften Messung verursacht wurde.

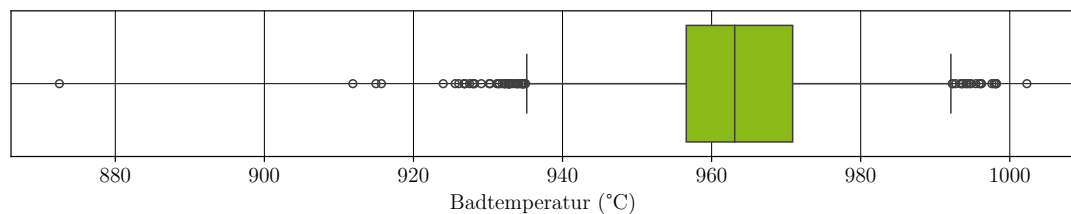


Abbildung 5.6: Ein Boxplot, der anhand aller Badtemperaturdaten berechnet wurde.

Die Abbildung 5.7 zeigt für jeden Ofen einen Boxplot. Anhand des IQA können Öfen identifiziert werden, deren Badtemperaturwerte stärker streuen. Mit einem Wert von $19,4^{\circ}\text{C}$ weist der Ofen 1091 den größten IQA auf, während für den Ofen 1114 mit $9,2^{\circ}\text{C}$ der kleinste IQA berechnet wurde. In Bezug auf den Median weist der Ofen 1108 mit $959,7^{\circ}\text{C}$ den niedrigsten Wert und der Ofen 1095 mit $967,2^{\circ}\text{C}$ den höchsten Wert auf.

Auffällig sind die Boxplots mit Ausreißern. Ausreißer können nach [KND15, S. 69] in plausible und nicht plausible Ausreißer unterteilt werden. Als plausible Ausreißer werden Werte bezeichnet, die im Hinblick auf den zu untersuchenden Prozess gültig sind, sich jedoch von den restlichen Werten deutlich unterscheiden. Bezogen auf die Badtemperatur stellt beispielsweise der maximale Wert von $1002,3^{\circ}\text{C}$, der am Ofen 1042 gemessen wurde, einen Ausreißer dar. Im Hinblick auf den Ofenprozess scheint dieser Ausreißer plausibel zu sein, da in der Aluminiumelektrolyse derartige hohe Badtemperaturen auftreten können. Demgegenüber kann die

Abschnitt 5.3. Zeitreihenanalyse der Badtemperaturdaten

minimale Badtemperatur von 872,5 °C des Ofens 1115 bei einem regulären Ofenbetrieb als ein nicht plausibler Ausreißer eingestuft werden, der auf einen Messfehler zurückzuführen ist.

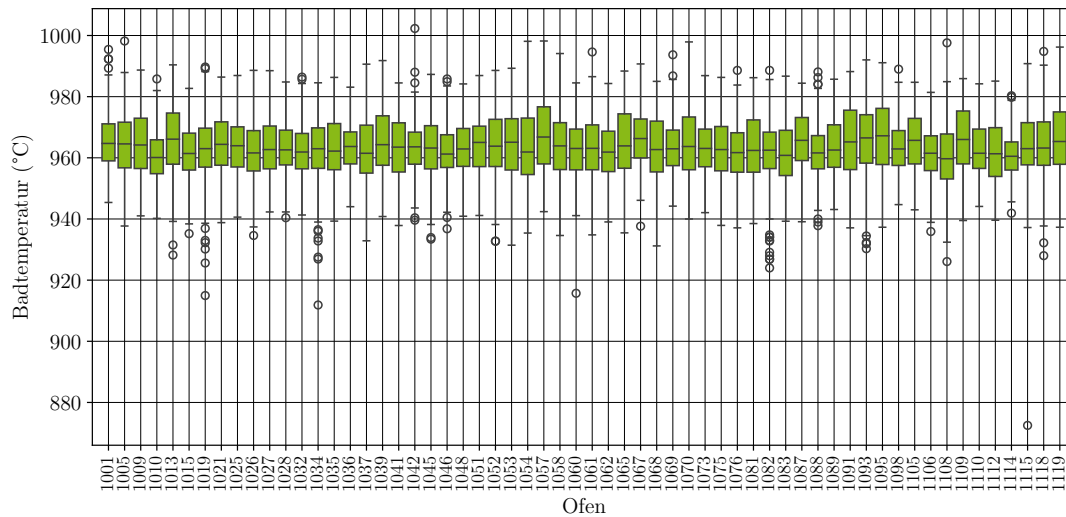


Abbildung 5.7: Ein Boxplot für jeden der 60 untersuchten Öfen, berechnet aus den jeweiligen Badtemperaturen. Eine größere Version dieser Abbildung befindet sich in Anhang A.2.

Um einen Eindruck vom zeitlichen Verlauf der Badtemperaturen zu erhalten, sind in Abbildung 5.8 die Verläufe von vier ausgewählten Öfen dargestellt. Die vertikalen Markierungen kennzeichnen Datenlücken in den Badtemperaturverläufen. Die Badtemperaturen in Abbildung 5.8 befinden sich auf einem konstanten Niveau, da die Badtemperatur eines Ofens geregelt wird (vgl. Kapitel 3.2). Auffällig ist das zyklische Muster der einzelnen Verläufe, das sich in einer langsamen Auf- und Abbewegung bemerkbar macht und mit einer Periode von ungefähr 30 Tagen auftritt. Zudem sind Temperaturspitzen zu sehen: Ein niedriger Temperaturwert ist Anfang November 2022 für den Ofen 1115 zu erkennen, der unter 930 °C liegt. Eine hohe Temperaturspitze ist gegen Ende April für den Ofen 1058 ersichtliche.

Abschnitt 5.3. Zeitreihenanalyse der Badtemperaturdaten

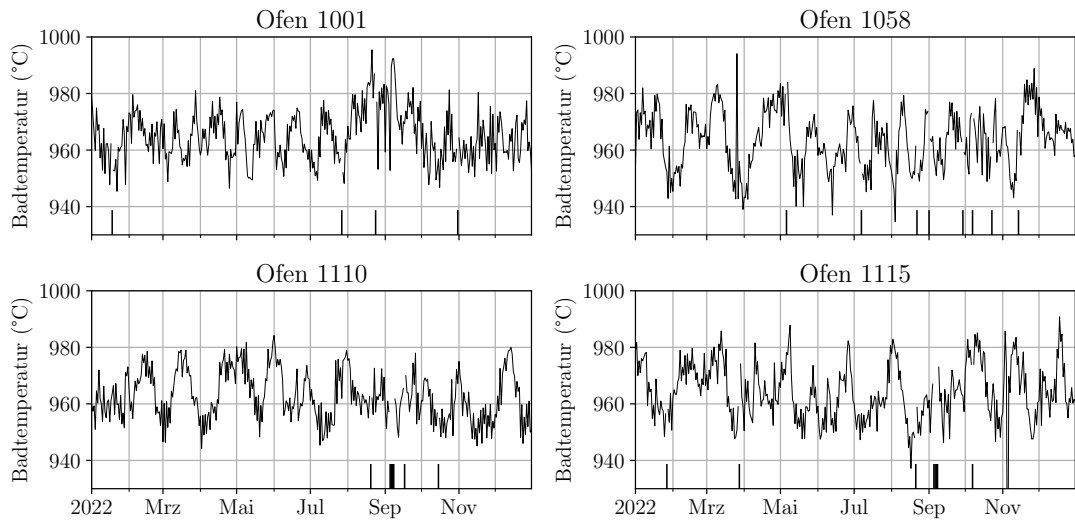


Abbildung 5.8: Badtemperaturverläufe von vier Öfen in einem Zeitraum vom 01.01.2022 bis zum 31.12.2022. Die vertikalen Markierungen zeigen Datenlücken in den Badtemperaturverläufen an.

Abbildung 5.9 zeigt einen Ausschnitt der Badtemperaturverläufe für die gleichen Öfen im Oktober 2022. In dieser sind zusätzlich unterschiedliche starke Schwankungen zu erkennen, die dem zyklischen Muster überlagert sind. Teilweise geht aus diesen Verläufen ein Dreiecksverlauf hervor. Die Temperaturwerte liegen an jedem zweiten Tag höher als am Vortag.

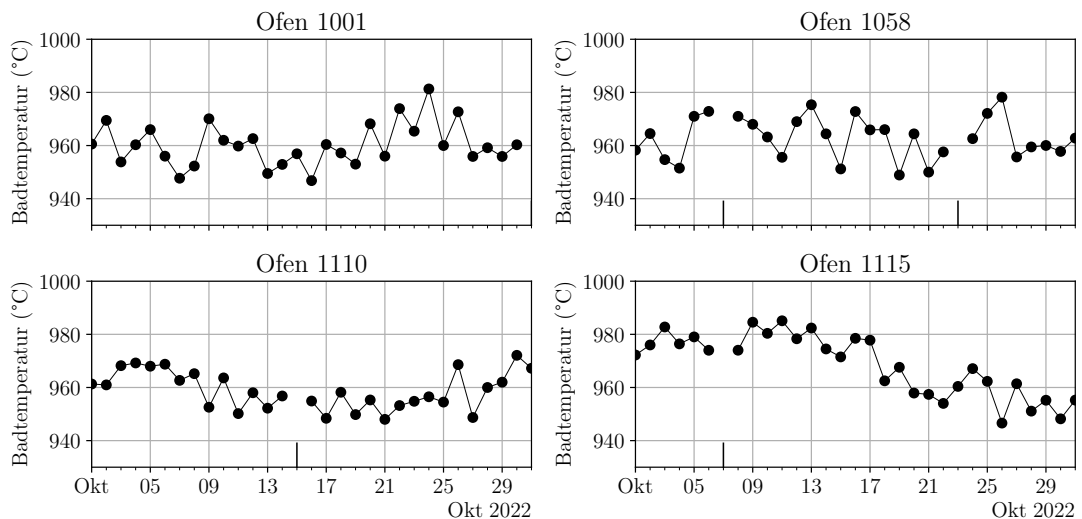


Abbildung 5.9: Badtemperaturverläufe von vier Öfen in einem Zeitraum vom 01.10.2022 bis zum 31.10.2022. Die vertikalen Markierungen zeigen Datenlücken in den Badtemperaturverläufen an.

Abschnitt 5.3. Zeitreihenanalyse der Badtemperaturdaten

Bei der visuellen Betrachtung der Badtemperaturverläufe fallen große Badtemperaturänderungen auf, beispielsweise für den Ofen 1058 vom 26.10.2022 auf den 27.10.2022 (vgl. Abbildung 5.9). Da starke Badtemperaturänderungen in der Aluminiumelektrolyse auffällig sind, wurden diese genauer untersucht.

Um starke Badtemperaturänderungen zu identifizieren, wurde für jeden Ofen die Differenz zum vorherigen Badtemperaturwert berechnet. Anschließend wurden die berechneten Differenzen über einen weiteren Boxplot dargestellt, um Öfen zu identifizieren, die eine besonders starke Badtemperaturänderung aufweisen. Die Frage bestand darin, ob starke Badtemperaturänderungen aufgrund einer fehlerhaften Messung zustande gekommen sind oder ein plausibles Ofenverhalten widerspiegeln. Abbildung 5.10 zeigt für jeden Ofen einen solchen Boxplot der Badtemperaturdifferenzen. Aus diesen geht hervor, dass Badtemperaturänderungen in Höhe von $\pm 20^\circ\text{C}$ von einem auf den nächsten Tag nicht unüblich sind. Einige Ausreißer stechen besonders hervor, wie beispielsweise die Ausreißer der Öfen 1019, 1034, 1054 und 1058. Die beiden extremen Ausreißer des Ofens 1115 sind auf den bekannten Messfehler von $872,5^\circ\text{C}$ zurückzuführen.

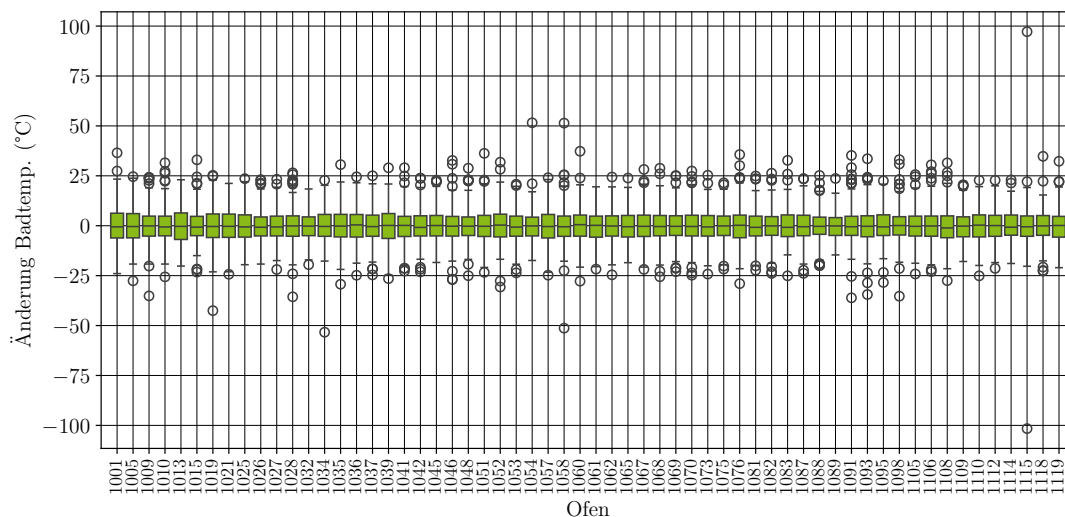


Abbildung 5.10: Ein Boxplot für jeden der 60 untersuchten Öfen, berechnet aus den jeweiligen Badtemperaturänderungen. Eine größere Version dieser Abbildung befindet sich in Anhang A.3.

In Abbildung 5.11 sind die Badtemperaturverläufe der Öfen 1019, 1034, 1054 und 1058 zu sehen, die besonders starke Badtemperaturänderungen aufweisen. Für die Öfen 1019 und 1034 ist eine starke negative Badtemperaturänderung vom 07.07.2022 auf den 08.07.2022 erkennbar. Für den Ofen 1019 beträgt die Änderung $-42,5^\circ\text{C}$ und für den Ofen 1034 $-53,3^\circ\text{C}$. Eine weitere Badtemperaturänderung von $51,6^\circ\text{C}$ ist für den Ofen 1054 zu sehen. Der Badtemperaturverlauf für den Ofen 1058 zeigt zunächst eine Badtemperaturänderung von $51,4^\circ\text{C}$ mit einer anschlie-

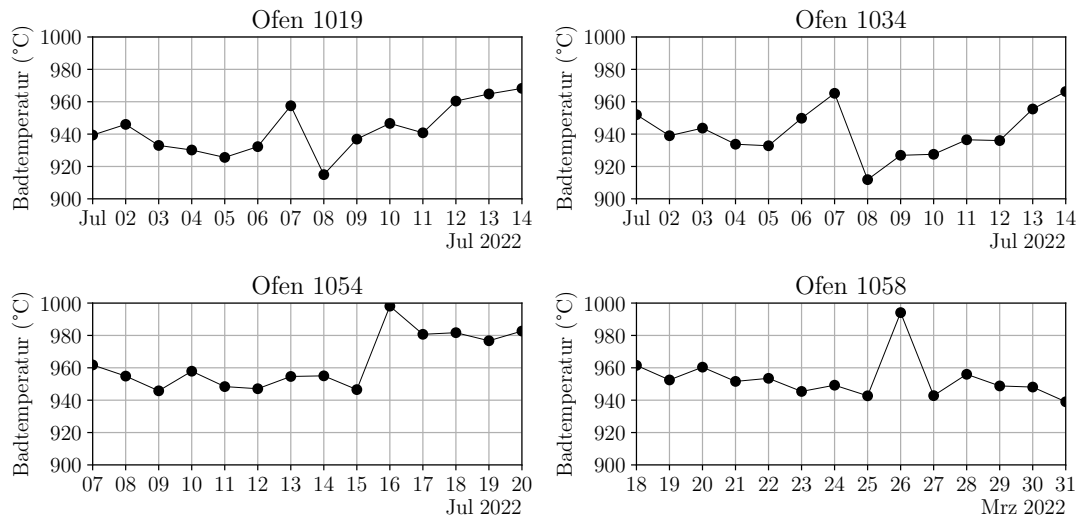


Abbildung 5.11: Badtemperaturverläufe von vier Öfen, die eine starke Badtemperaturänderung aufweisen.

ßenden Änderung von $-51,3^{\circ}\text{C}$. Auffällig an dem Ausreißer des Ofens 1058 ist, dass dieser nicht mit dem weiteren Badtemperaturverlauf übereinstimmt, während die Badtemperaturen der anderen Öfen nach dem jeweiligen Ausreißer plausibel weiter verlaufen.

5.3.3 Badtemperaturänderungen

Zusammen mit der TAE wurden die Ausreißer von den Öfen 1019, 1034 und 1054 genauer untersucht. Aufgrund der in Kapitel 3.2 vorgestellten Einflüsse, die auf die Badtemperatur eines Ofens wirken, erwies sich eine Plausibilitätsbetrachtung von Ausreißern lediglich auf Basis der Badtemperaturverläufe als wenig aussagekräftig. Daher wurden weitere Prozessvariablen herangezogen, um die starken Badtemperaturänderungen der einzelnen Öfen zu erklären. In Abbildung 5.12 ist der Badtemperaturverlauf des Ofens 1019 zu sehen. Der Badtemperaturverlauf wurde mithilfe der tatsächlichen Messzeitpunkte aufgetragen. Für einen besseren Überblick sind die Temperaturmesswerte und die Badtemperaturänderungen zum Vortag im unteren Teil der Abbildung dargestellt. Die angegebenen Temperaturwerte sind auf eine Nachkommastelle gerundet. Des Weiteren sind in dieser die Zeitpunkte von AlF_3 -Zugaben, entdeckten und beseitigten AlF_3 -Durchläufern, Ansätzen, Anodenkontrollen, Metallsaugen, Flusssaugen, Flusszugaben, Sodazugaben, Anodeneffekten, Anodenwechseln und Stromabfällen, beispielsweise aufgrund einer Hallenschaltung, berücksichtigt. Somit kann ein möglicher Zusammenhang zwischen den Einflüssen und den Badtemperaturänderungen hergestellt werden.

Bei der Betrachtung von Abbildung 5.12 ist zu beachten, dass für Flusssaugen, Flusszugaben, Sodazugaben und entdeckte Spikes lediglich der Schichtbeginnzeitstempel von der Schicht angegeben ist, in der der jeweilige Einfluss stattgefunden hat bzw. entdeckt wurde. Beispielsweise

Abschnitt 5.3. Zeitreihenanalyse der Badtemperaturdaten

weist ein Datenbankeintrag für eine Sodazugabe den Zeitstempel 03.01.2022 05:00 Uhr (UTC, Beginn der ersten Schicht) auf. Das bedeutet, dass eine Sodazugabe am 03.01.2022 zwischen 5:00 Uhr (UTC) und 13:00 Uhr (UTC) stattgefunden hat. Die genaue Uhrzeit für die Sodazugabe ist jedoch nicht bekannt. Für Metallsaugen, Flusssaugen, Flusszugaben und Sodazugaben sind zusätzliche Gewichtsangaben für die Zugaben bzw. für das jeweilige Saugen angegeben. Im Falle eines Anodeneffekts wird die umgesetzte Energie angegeben.

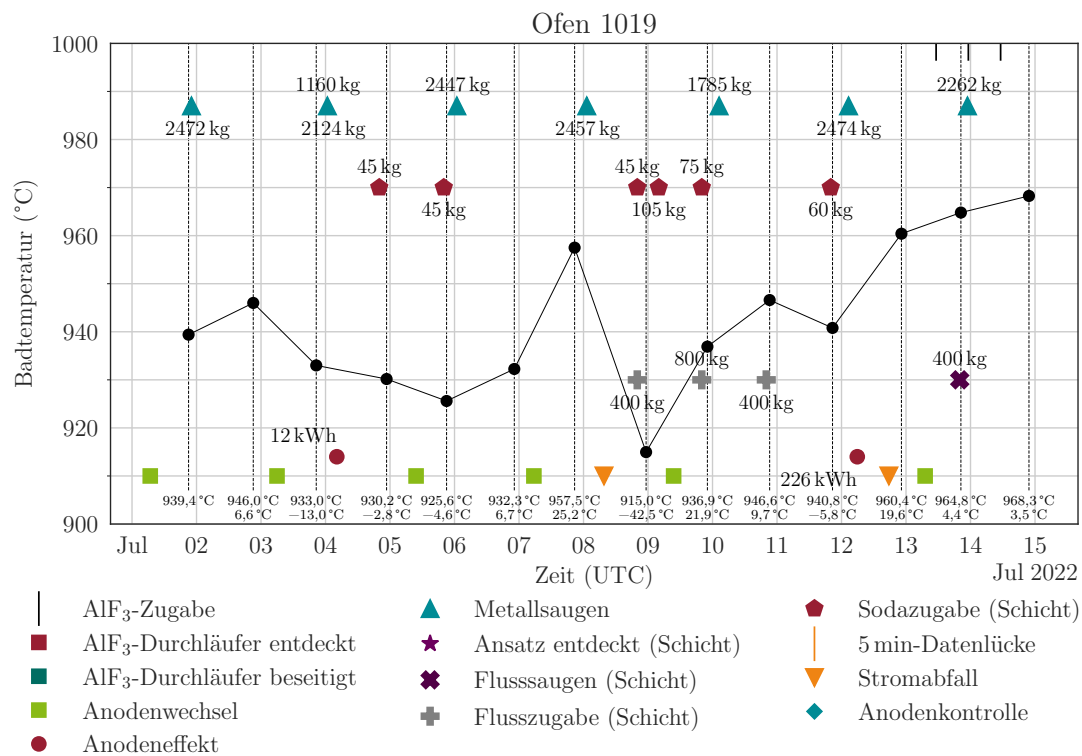


Abbildung 5.12: Zeitlicher Verlauf der Prozessvariablen des Ofens 1019.

Vor dem 07.07.2022 befand sich die Badtemperatur des Ofens 1019 auf einem relativ niedrigen Niveau. Am 04.07.2022 und am 05.07.2022 wurde dem Ofen jeweils 45 kg Soda hinzugegeben. Die Badtemperatur stieg in den darauffolgenden zwei Tagen auf bis zu 957,5 °C an und fiel anschließend auf den niedrigen Temperaturwert von 915,0 °C. Vermutlich waren die Sodazugaben nicht ausreichend, um den starken Abfall der Badtemperatur auf 915,0 °C zu verhindern. Zudem könnte die Flusszugabe von 400 kg einen Einfluss auf die Badtemperaturmessung vom 08.07.2022 gespielt haben, abhängig davon, zu welcher Uhrzeit die Flusszugabe stattgefunden hat. Wie bereits in Kapitel 3.2 beschrieben, erhöht eine Flusszugabe die Kontaktfläche zwischen Fluss und Seitenwand, die zu einer Abkühlung des Ofens führt. Inwiefern der Stromabfall am 08.07.2022 zu der starken Badtemperaturänderung beigetragen hat, lässt sich nicht beantworten. Weitere Sodazugaben fanden am 08.07.2022, 09.07.2022 und 11.07.2022 statt. In diesem Zeitraum ist ein Anstieg der Badtemperatur zu erkennen. Der Anodeneffekt am 12.07.2022

Abschnitt 5.3. Zeitreihenanalyse der Badtemperaturdaten

könnte zusätzlich zu einem weiteren Anstieg der Badtemperatur beigetragen haben. Die Badtemperaturänderung um $-42,5^{\circ}\text{C}$ auf $915,0^{\circ}\text{C}$ scheint insgesamt plausibel zu sein.

In Abbildung 5.13 sind die Badtemperatur und Prozessvariablen für den Ofen 1034 zeitlich aufgetragen. Dabei ist zu sehen, dass vom 07.07.2022 auf den 08.07.2022 eine Badtemperaturänderung von $-53,3^{\circ}\text{C}$ aufgetreten ist, die wahrscheinlich auf den AlF_3 -Durchläufer zurückzuführen ist, der am 06.07.2022 entdeckt wurde. AlF_3 -Durchläufer treten auf, wenn der Zylinder der AlF_3 -Dosiereinheit nicht richtig schließt, wodurch kontinuierlich AlF_3 in den Ofen gelangt. Aufgrund dieser unkontrollierten Zugabe an AlF_3 können die Liquidus- und Badtemperatur abfallen. Der AlF_3 -Durchläufer wurde schließlich am 07.07.2022 beseitigt. Auch für diesen Ofen lässt sich die Auswirkung des Stromabfalls am 08.07.2022 auf die Badtemperatur nicht genau beantworten. Jedoch scheint die Badtemperaturänderung von $-53,3^{\circ}\text{C}$ insgesamt plausibel zu sein.

Eine Maßnahme gegen einen überhöhten Gehalt an AlF_3 im Fluss ist die Zugabe von Soda [TRI13, S. 163], die auch mehrfach in den darauffolgenden Tagen am Ofen 1034 durchgeführt wurde. Der Anodeneffekt am 11.07.2022 mit einem Energieeintrag von 244 kWh könnte zusätzlich zu einem Anstieg der Badtemperatur beigetragen haben.

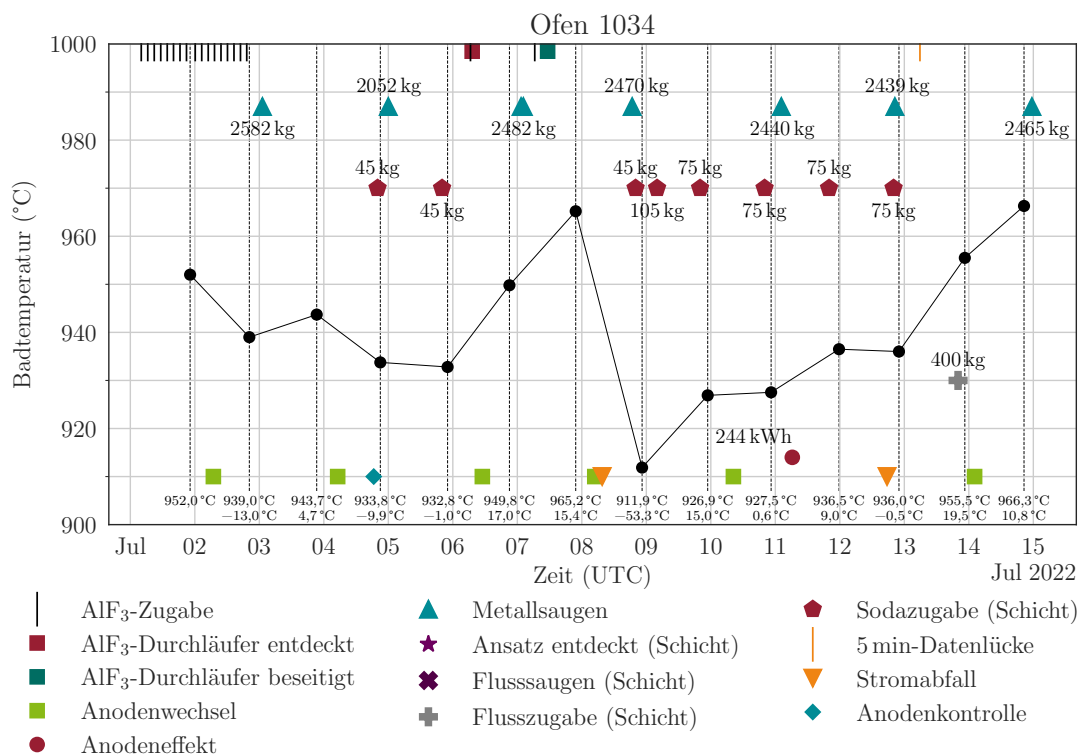


Abbildung 5.13: Zeitlicher Verlauf der Prozessvariablen des Ofens 1034.

Abschnitt 5.3. Zeitreihenanalyse der Badtemperaturdaten

In Abbildung 5.14 ist ein starker Temperaturanstieg von 946,5 °C auf 998,1 °C für den Ofen 1054 zu sehen. Dieser Temperaturanstieg um 51,6 °C kann mit der Zugabe von 60 kg Soda und dem Saugen von 400 kg Fluss jeweils am 15.07.2022 sowie mit dem Anodeneffekt mit einem Energieeintrag von 37 kWh am 16.07.2022 begründet werden. Daher scheint auch diese Badtemperaturänderung des Ofens 1054 plausibel zu sein.

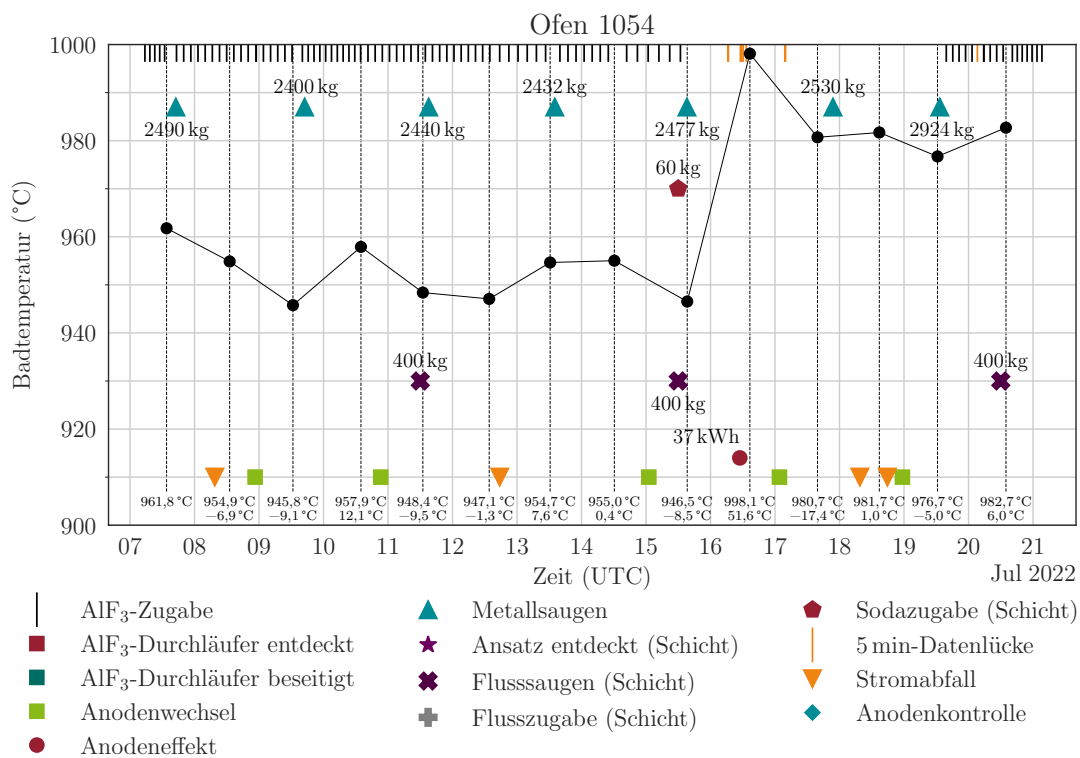


Abbildung 5.14: Zeitlicher Verlauf der Prozessvariablen des Ofens 1054.

Ein Gegenbeispiel stellt der Ofen 1058 dar, dessen Badtemperaturverlauf mit weiteren Prozessvariablen in Abbildung 5.15 zeitlich aufgetragen sind. In dieser ist zu sehen, dass ein kurzzeitiger Temperatursprung auf 994,1 °C am 26.03.2022 stattgefunden hat, der sich nicht über die hier dargestellten Prozessvariablen erklären lässt. Wie bereits festgestellt, widerspricht dieser Temperatursprung dem restlichen Verlauf der Badtemperatur. Daher wird davon ausgegangen, dass es sich bei der Temperaturmessung am 26.03.2022 um eine Fehlmessung handelt. Diese Aussage wird bestätigt, weil in der Datenbank der TAE für den gleichen Tag eine Badtemperaturmessung mit 954,2 °C eingetragen ist, die ca. 3 Stunden vorher stattgefunden hat. Die Messung von 954,2 °C stimmt im Gegensatz zu 994,1 °C auch mit dem restlichen Temperaturverlauf überein. An dieser Stelle ist noch einmal zu erwähnen, dass die Messung von 954,2 °C aufgrund der Vorverarbeitung der Badtemperaturdaten in Kapitel 5.2 nicht in Abbildung 5.15 auftritt, da immer nur die letzte in einer Schicht durchgeführte Messung berücksichtigt wurde. In diesem

Abschnitt 5.3. Zeitreihenanalyse der Badtemperaturdaten

Fall ist die Messung von 994,1 °C die letzte Badtemperaturmessung gewesen, die in der zweiten Schicht durchgeführt wurde und sich daher im vorverarbeiteten Datensatz befindet.

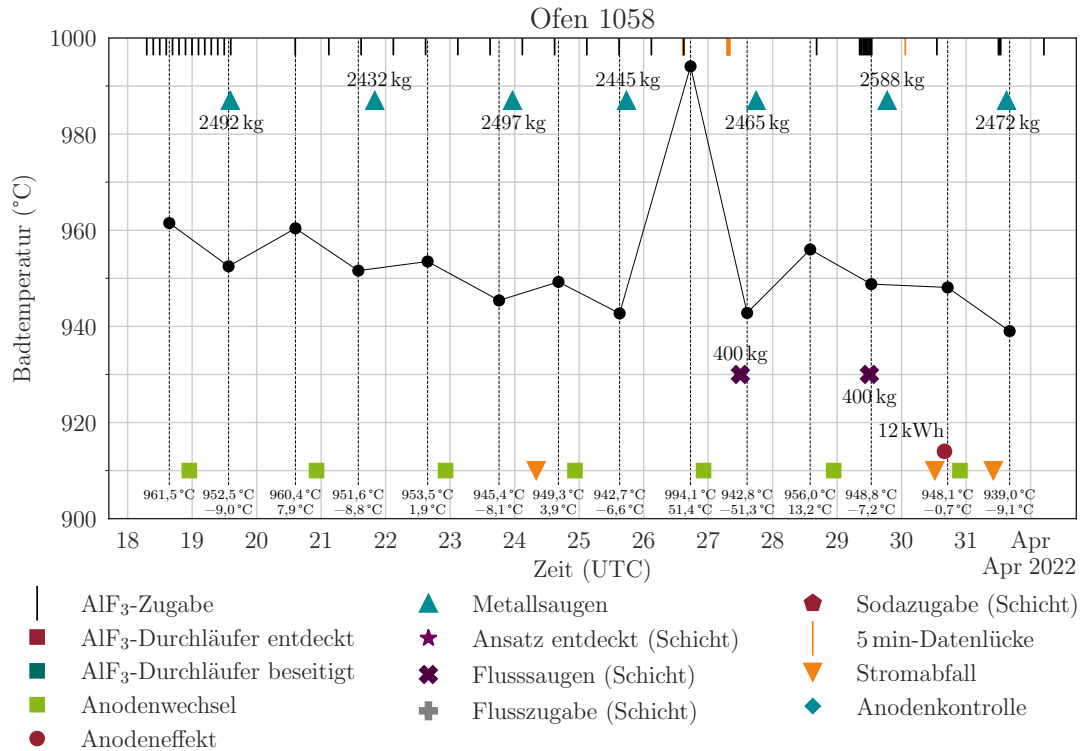


Abbildung 5.15: Zeitlicher Verlauf der Prozessvariablen des Ofens 1058.

Insgesamt zeigen die durchgeführten Plausibilitätsbetrachtungen auf, dass starke Badtemperaturänderungen von ± 50 °C in der Aluminiumelektrolyse auftreten. Für die Erklärung von Badtemperaturänderungen ist es erforderlich, weitere Prozessvariablen in der Analyse zu berücksichtigen. Die unterschiedlichen Einflüsse auf die Badtemperatur wurden bereits in Kapitel 3.2 beleuchtet. Die aufgezeigten Badtemperaturänderungen der Öfen 1019, 1034 und 1054 wurden in dieser Analyse als plausibel eingestuft. Eine Ausnahme stellt die Badtemperaturänderung des Ofens 1058 dar, die auf einen Messfehler zurückgeführt werden kann. Auch die eigentlichen Badtemperaturwerte scheinen bis auf den niedrigen Wert von 872,5 °C, der im Boxplot in Abbildung 5.6 zu erkennen ist, plausibel zu sein.

Eine Filterung von Ausreißern anhand von Temperaturschwellwerten erscheint nicht sinnvoll, da mit einer solchen Filterung auch Badtemperaturwerte herausgefiltert werden könnten, die im Hinblick auf den Zustand eines Ofens plausibel sind. Aufgrund des hohen Aufwands wurde in dieser Arbeit auf eine ausführliche Plausibilitätsanalyse von Ausreißern verzichtet. In Kapitel 5.4.3 wird jedoch eine Möglichkeit vorgestellt, um die Ursachenanalyse mithilfe eines RF-Regressionsmodells verbessern zu können.

Insgesamt wurden für die weitere Analyse Badtemperaturwerte unter 900 °C im Trainings- und Testdatensatz mit n.a. überschrieben. Davon ist beispielsweise der gemessene Badtemperaturwert von 872,5 °C des Ofens 1115 im Trainingsdatensatz betroffen. Die maximale Badtemperatur, die in den Trainingsdaten vorkommt, liegt bei 1002,3 °C und ist eine plausible Badtemperatur. Dennoch ist auch das Auftreten von hohen Badtemperaturmessungen zu berücksichtigen, die nicht plausibel für den Ofenbetrieb sind. In diesem Fall kann es sinnvoll sein, auch einen oberen Schwellwert zu definieren, um hohe Badtemperaturmesswerte zu identifizieren. Ein großzügig gewählter Schwellwert von 1050 °C könnte hierfür eine geeignete Basis sein. Eine Überprüfung der Testdaten ergibt, dass keine Badtemperaturmesswerte im Testdatensatz existieren, die den Schwellwert von 1050 °C übersteigen.

5.3.4 Datenlücken

Aus Tabelle 5.3 geht hervor, dass Datenlücken in den Badtemperaturdaten vorhanden sind, die mit n.a. gekennzeichnet sind. In Abbildung 5.16 ist ein Balkendiagramm zu sehen, das die Anzahl an Datenlücken in den Badtemperaturdaten für die 60 betrachteten Öfen aus Halle 1 im Jahr 2022 (Trainingsdaten) aufzeigt. Der Ofen 1037 hat mit zwei Datenlücken die niedrigste Anzahl, während der Ofen 1118 mit 14 Datenlücken die höchste Anzahl aufweist. Insgesamt sind 404 Datenlücken im gesamten Trainingsdatensatz vorhanden. Es fällt auf, dass die Öfen aus der ersten Schicht im Median eine höhere Anzahl an Datenlücken aufweisen als die Öfen der übrigen zwei Schichten.

Die höhere Anzahl an Datenlücken für die Öfen aus der ersten Schicht ist in Abbildung 5.17 erkennbar. Hier fallen vor allem die zusammenhängenden Datenlücken vom 05.09.2022 bis einschließlich 08.09.2022 für die Öfen zwischen 1081 und 1120 auf. Der Grund für die zusammenhängenden Datenlücken ist darin zu finden, dass die Badtemperaturmessungen an den betreffenden Öfen vom 05.09.2022 bis 07.09.2022 in der dritten Schicht durchgeführt wurden, anstatt wie vorgesehen von der ersten Schicht. Badtemperaturmessungen, die nicht von der zuständigen Schicht durchgeführt worden sind, wurden in der Datenvorverarbeitung dieser Arbeit verworfen. Dadurch sind Datenlücken im Datensatz entstanden (vgl. Kapitel 5.2.1). Für den 08.09.2022 liegen gar keine Badtemperaturmessungen für die Öfen zwischen 1081 und 1120 vor. Ab dem 09.09.2022 wurden wieder die Badtemperaturmessungen an diesen Öfen innerhalb der ersten Schicht durchgeführt.

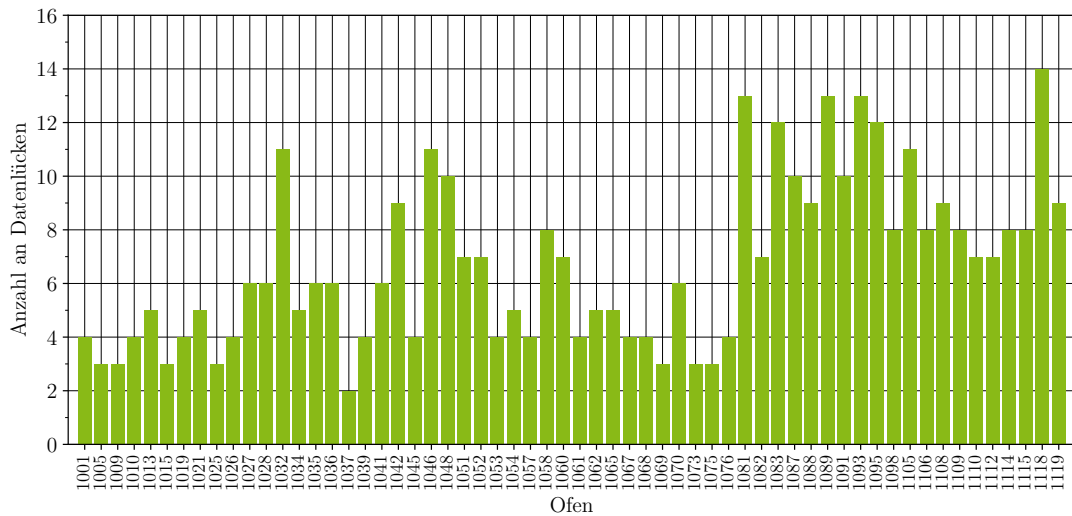


Abbildung 5.16: Anzahl an Datenlücken in den Badtemperaturdaten für jeden der 60 untersuchten Öfen aus Halle 1 im Jahr 2022. Eine größere Version dieser Abbildung befindet sich in Anhang A.4.

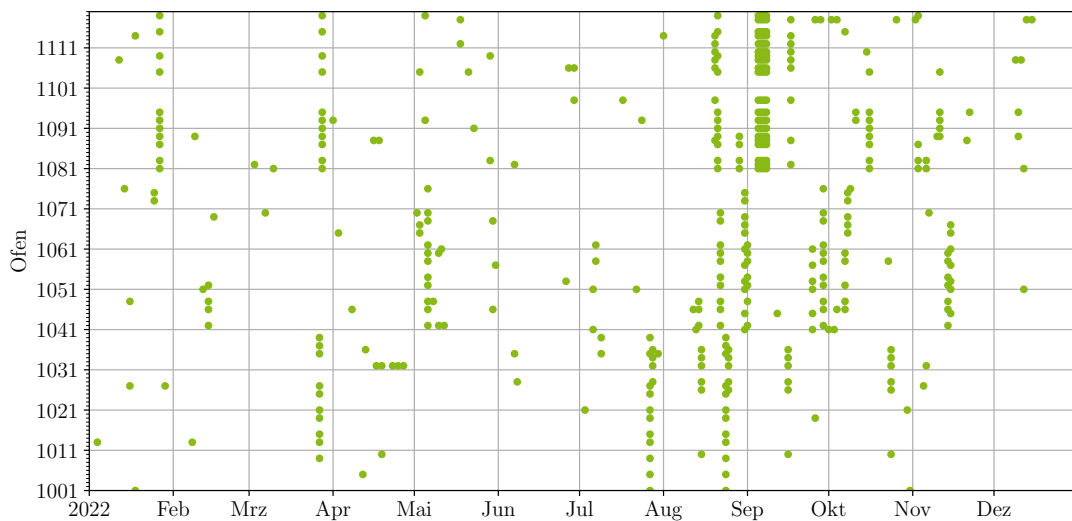


Abbildung 5.17: Zeitlich aufgetragene Datenlücken in den Badtemperaturdaten, die für 60 Öfen aus Halle 1 im Jahr 2022 ermittelt wurden. Ein grüner Punkt stellt eine Datenlücke dar.

Der Umgang mit Datenlücken stellt eine wesentliche Herausforderung dar, da nicht jede Analyse- und Vorhersagemethode mit Datenlücken umgehen kann [RM17, S. 107; HA21, S. 427]. Um diese Schwierigkeit zu vermeiden, können die Datenlücken in den Badtemperaturverläufen mit einer geeigneten Methode aufgefüllt werden. Aufgrund der zeitlichen Anordnung der Badtemperaturwerte bieten sich Methoden an, die die zeitliche Komponente der Daten

berücksichtigen. In [MBB17] werden verschiedene Methoden erörtert, die darauf abzielen, Datenlücken in univariaten Zeitreihen zu schließen.

Nach [KND15, S. 74] sind zunächst einfache Methoden für das Auffüllen von Datenlücken zu bevorzugen. Daher wurde in dieser Arbeit die Methode Last Observation Carried Forward (LOCF) verwendet, da diese einfach zu implementieren ist. Bei der LOCF-Methode wird jeweils die letzte vorhandene Beobachtung vor der Datenlücke verwendet, um die Lücke zu schließen. Für die Werte in Tabelle 5.3 bedeutet das, dass die Datenlücke (n.a.) vom 12.01.2022 mit dem Wert 958,4 °C vom Vortag aufgefüllt wird, während für die Datenlücke (n.a.) vom 16.01.2022 der Badtemperaturwert 954,0 °C verwendet wird. Die Anwendung von LOCF erfolgte in gleicher Weise auf dem Trainings- und Testdatensatz.

Abbildung 5.18 zeigt vier Badtemperaturverläufe der Öfen 1001, 1058, 1110 und 1115. Dabei stellen die grünen Punkte die Werte dar, die durch LOCF für Datenlücken eingesetzt wurden. Für die Öfen 1110 und 1115 ist zu sehen, dass die Datenlücken vom 05.09., 06.09., 07.09. und 08.09.2022 mit dem Badtemperaturwert vom 04.09.2022 aufgefüllt wurden.

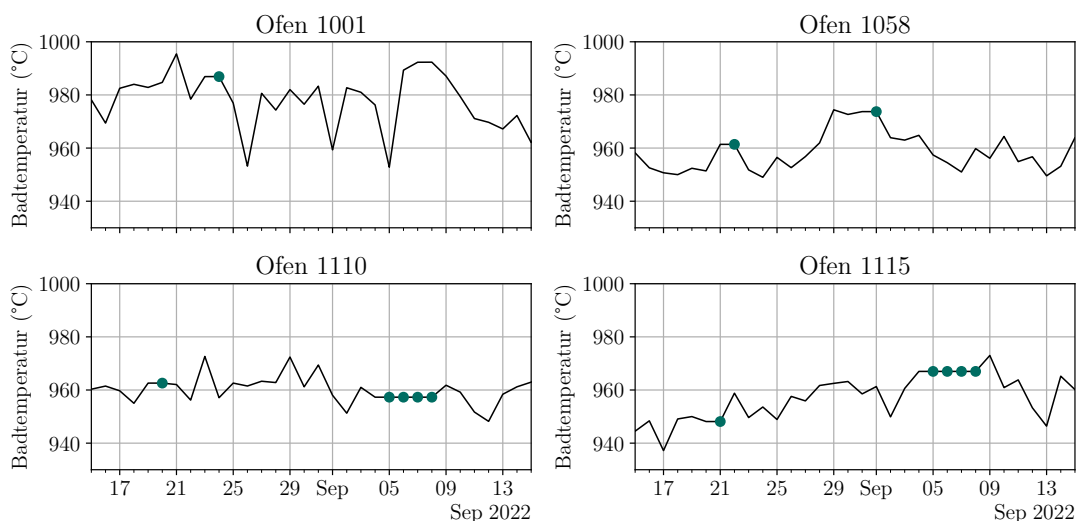


Abbildung 5.18: Badtemperaturverläufe von den Öfen 1001, 1058, 1110 und 1115. Die grünen Punkte stellen die Werte dar, die für das Auffüllen der Datenlücken mithilfe der LOCF-Methode verwendet wurden.

5.3.5 Autokorrelation

Um einen ersten Eindruck über die Korrelationen der Badtemperaturen zu gewinnen, wurden, ähnlich wie in [MAW01, S. 298], über alle 60 Öfen die Koeffizienten der AKF und PAKF berechnet und gemittelt. In Abbildung 5.19 sind die gemittelten Koeffizienten der AKF und PAKF zu sehen, die jeweils bis zur Verzögerungsstelle $k = 35$ berechnet wurden. Die gemittelten Koeffizienten der AKF (vgl. Abbildung 5.19a) zeigen einen annähernd sinusförmigen Verlauf.

Die Koeffizienten der PAKF (vgl. Abbildung 5.19b) sind nach der Verzögerungsstelle $k = 9$ nicht mehr signifikant. Auffallend sind die ersten acht signifikanten Koeffizienten der AKF. Dabei lässt sich beobachten, dass jeder zweite signifikante Koeffizient höher liegt als der vorherige Koeffizient. Dieses Verhalten deutet auf eine Periodizität bzw. Saisonalität hin, die sich alle zwei Tage wiederholt. Das ist auch in Abbildung 5.9 anhand des Dreiecksverlaufs der Badtemperaturen ersichtlich. In vielen Fällen liegt jede zweite gemessene Badtemperatur höher als die am Vortag gemessene Temperatur, was mit dem Anodenwechsel begründet werden kann, der in der Regel alle zwei Tage bei der TAE stattfindet.

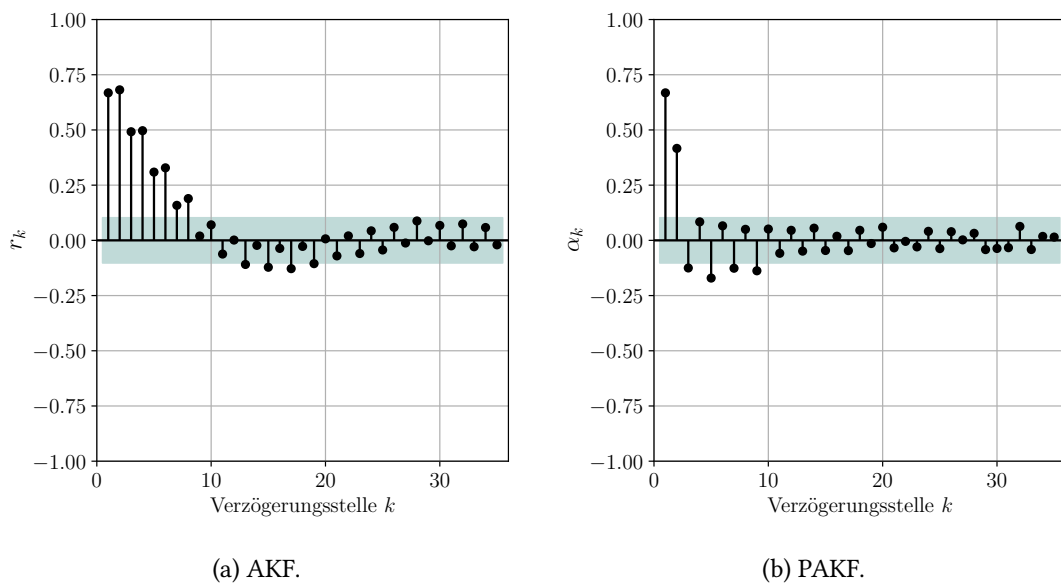


Abbildung 5.19: Für jeden Ofen wurden die AKF und PAKF bis zur Verzögerungsstelle $k = 35$ berechnet. Die Koeffizienten wurden anschließend gemittelt und gegen die Verzögerungsstellen aufgetragen. In (a) sind die gemittelten Koeffizienten der AKF und in (b) die gemittelten Koeffizienten der PAKF jeweils mit einem Konfidenzintervall von 95 % zu sehen.

Darüber hinaus geht aus [MAW01, S. 298] hervor, dass es sich bei den dort untersuchten Badtemperaturverläufen um Zeitreihen handelt, die aus einem autoregressiven Prozess resultieren. Bei der Betrachtung der Korrelogramme in Abbildung 5.19 kann diese Aussage auch auf die vorliegenden Badtemperaturverläufe übertragen werden, da die Koeffizienten der AKF einen sinusförmigen Verlauf aufweisen und die Koeffizienten der PAKF nach der neunten Verzögerungsstelle nicht mehr signifikant sind. Nach [HA21, S. 283; SS17, S. 99] lassen sich diese Eigenschaften als Indizien für einen autoregressiven Prozess werten.

5.3.6 Periodogramm

In Abbildung 5.20 ist für jeden einzelnen Badtemperaturverlauf das Periodogramm sowie das gemittelte Periodogramm zu sehen. Das Periodogramm wurde bereits in Kapitel 4.1.5 beschrieben. Aus Abbildung 5.20 geht hervor, dass eine Häufung der Spektren im unteren Frequenzbereich vorhanden ist. Diese Erkenntnis wurde bereits von McFadden et al. [MAW01] erlangt und deutet darauf hin, dass es sich bei den Badtemperaturverläufen um einen sich „langsam bewegenden“ [MAW01, S. 298] Prozess handelt. Die linke gemittelte Frequenzspitze entspricht einer Periode von ungefähr 30 Tagen. Das bedeutet, dass sich in den Badtemperaturverläufen ein ausgeprägtes Muster befindet, das sich alle 30 Tage wiederholt. Diese Erkenntnis wurde bereits in Abbildung 5.8 ersichtlich. Insgesamt wurden in dieser Arbeit solche „langsamen“ Bewegungen als Zyklus gewertet (vgl. Kapitel 4.1.1).

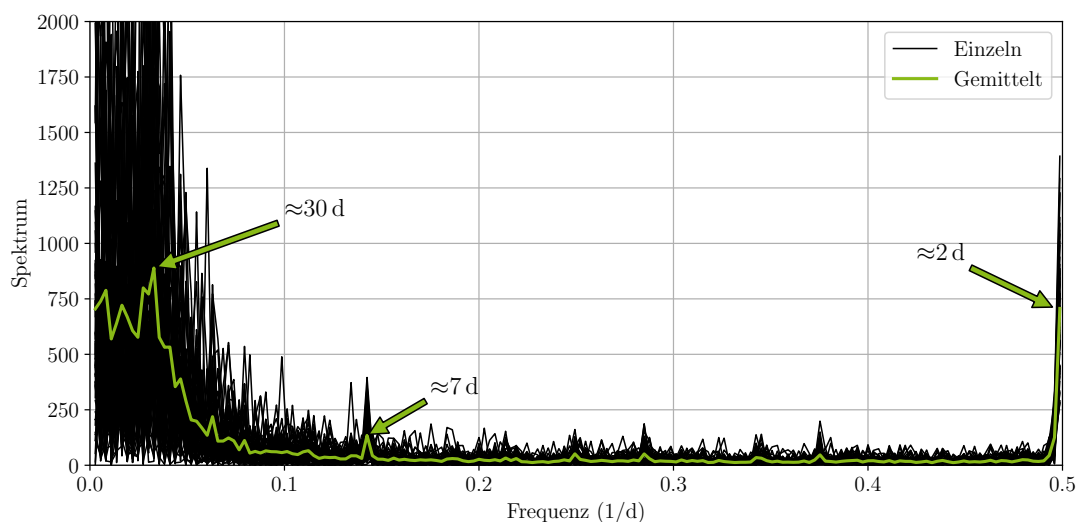


Abbildung 5.20: Für jeden einzelnen Badtemperaturverlauf wurde ein Periodogramm ohne anschließende Glättung berechnet. Anschließend wurden die einzelnen Periodogramme gemittelt, um einen globalen Eindruck über sich wiederholende Muster in den Badtemperaturverläufen zu erhalten.

Eine weitere Frequenzspitze deutet auf eine Bewegung hin, die sich ungefähr alle sieben Tage wiederholt. Diese ist jedoch nicht sonderlich stark ausgeprägt. Ursachen konnten für diese sich wiederholende Bewegung nicht identifiziert werden. Des Weiteren ist eine Frequenzspitze bei zwei Tagen ersichtlich. Diese wurde bereits bei der Analyse der Korrelogramme in Kapitel 5.3.5 erkannt und mit dem Anodenwechsel der TAE begründet.

Für die weitere Vorgehensweise wurden die Badtemperaturverläufe auf Stationarität überprüft, wobei die zweitägige Periode im Fokus stand. Die siebentägige Periode wurde nicht weiter berücksichtigt, da diese nicht sonderlich stark im gemittelten Periodogramm ausgeprägt ist.

5.3.7 Stationarität

Die Koeffizienten der AKF in Abbildung 5.19 und das Periodogramm in Abbildung 5.20 weisen auf eine Periodizität in den Badtemperaturverläufen hin, die mit dem Anodenwechsel begründet wird, der üblicherweise alle zwei Tage bei der TAE stattfindet. Periodische bzw. saisonale Zeitreihen sind nicht stationär (vgl. Kapitel 4.1.3). Es wurde daher überprüft, in welchem Umfang die zweitägige Saisonalität in den Badtemperaturverläufen ausgeprägt ist, um eine saisonale Differenzierung zu rechtfertigen. Dazu wurde die Stärke der Saisonalität mithilfe der Gleichung 4.5 für jeden Badtemperaturverlauf berechnet.

Sofern der errechnete Wert unter dem Schwellwert von 0,64 liegt, ist eine saisonale Differenzierung für den entsprechenden Badtemperaturverlauf nicht notwendig [HA21, S. 273]. Insgesamt wurde für 58 von 60 Öfen jeweils ein Wert berechnet, der unter dem Schwellwert von 0,64 liegt. Nur für die Badtemperaturverläufe der Öfen 1005 und 1013 liegt der jeweilige Wert knapp über dem Schwellwert. Somit zeigt sich, dass die zweitägige Saisonalität für 58 Badtemperaturverläufe nur eine untergeordnete Rolle spielt, sodass auf eine saisonale Differenzierung verzichtet wurde. Die ermittelten Werte sind als Boxplot in Abbildung 5.21 aufgeführt. Aus diesem lässt sich ableiten, dass der Median bei 0,49 liegt. Das untere Quartil liegt bei 0,43, während das obere Quartil bei 0,54 liegt. Aufgrund dieser Ergebnisse wurde für den anschließenden KPSS-Test keine saisonale Differenzierung an den Badtemperaturverläufen vorgenommen.

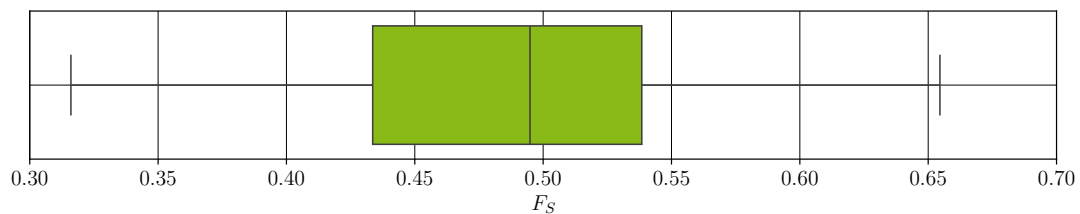


Abbildung 5.21: Stärke der zweitägigen Saisonalität, die für jeden Badtemperaturverlauf in den Trainingsdaten berechnet wurde. Die Verteilung der berechneten Werte ist mit dem Boxplot dargestellt.

Mithilfe des KPSS-Tests, der im Kapitel 4.1.3 vorgestellt wurde, wurden die Badtemperaturen für jeden Ofen auf Stationarität überprüft. Das Signifikanzniveau wurde mit 5 % ($\alpha = 0,05$) festgelegt. Für 14 von 60 Badtemperaturverläufen wird die Nullhypothese des KPSS-Tests verworfen. Diese Badtemperaturverläufe sind nach dem KPSS-Test nicht stationär. In Abbildung 5.22 sind die Badtemperaturverläufe von vier Öfen zu sehen, für die die Nullhypothese des KPSS-Tests verworfen wird. Da für die restlichen 46 Badtemperaturverläufe die Nullhypothese des KPSS-Tests nicht abgelehnt wird, wurden diese als stationär eingestuft. Dieses Ergebnis ist vergleichbar mit den Ergebnissen von McFadden et al. [MAW01], die ihre untersuchten

Badtemperaturverläufe anhand eines durchgeführten „run test“ ebenfalls als stationär eingestuft haben.

Anhand der in diesem Kapitel dargelegten Ergebnisse wurde angenommen, dass die in dieser Arbeit untersuchten Badtemperaturverläufe insgesamt stationär sind. Anhand dieses Ergebnisses wurde eine geeignete Methode gewählt, um die Performanz von Vorhersagemodellen abzuschätzen. Wie bereits in Kapitel 4.3 erläutert, wird nach [CTM20, S. 1] die CV-BI für stationäre Zeitreihen empfohlen.

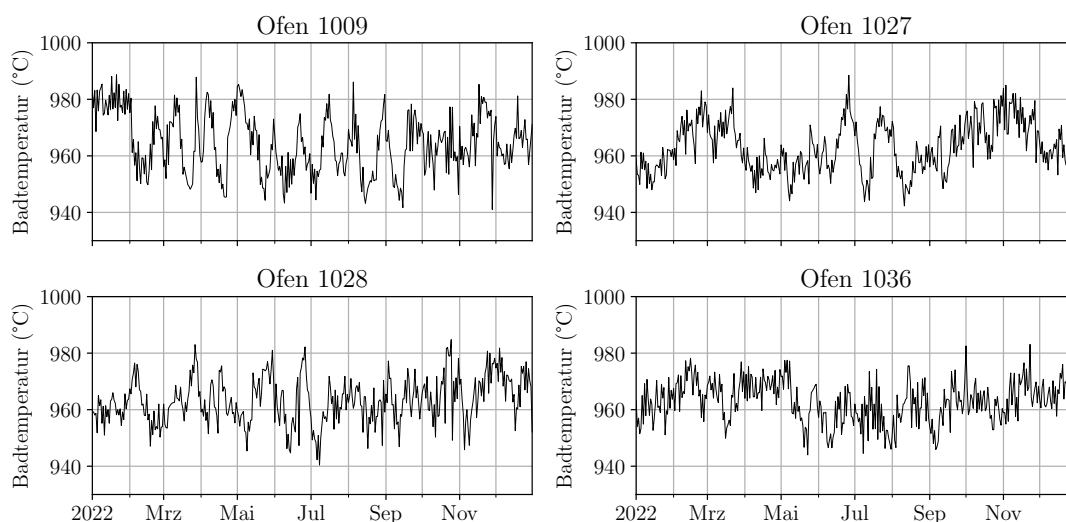


Abbildung 5.22: Badtemperaturverläufe (nicht saisonal differenziert) von den Öfen 1009, 1027, 1028 und 1036, für die die Nullhypothese des KPSS-Tests abgelehnt wird.

5.4 Lösungsentwicklung

Im Rahmen der weiteren Vorgehensweise nach Abbildung 5.1 werden in diesem Kapitel Lösungen aufgezeigt, die eine Badtemperaturvorhersage ermöglichen sollen. Es wird zwischen einer Tagesprognose (Forecasting) und einer Gegenwartsvorhersage (Nowcasting) unterschieden. Bei der Tagesprognose wird eine Badtemperatur für den nächsten Tag vorhergesagt. Dazu wurde ein globales AR-Modell trainiert und validiert. Um die Modellperformanz einschätzen zu können, erfolgte ein Vergleich des globalen AR-Modells mit vier ausgewählten lokalen Zeitreihenmodellen auf einem Testdatensatz.

Als Gegenwartsvorhersage (Nowcasting) kann die Vorhersage bezeichnet werden, die für den aktuellen Zeitpunkt, für die nahe Zukunft und die nahe Vergangenheit getätigt wird [BGR10, S. 5]. Für die Gegenwartsvorhersage der Badtemperatur wurde ein globales RF-Regressionsmodell mit weiteren Prozessvariablen aus der Aluminiumelektrolyse sowie un-

terschiedlichen Hyperparameterkombinationen trainiert und validiert. Zuvor wurde eine Verarbeitung der Prozessdaten durchgeführt, bei der Zeitreihenmerkmale in einem festgelegten zeitlichen Intervall berechnet und mit der gemessenen Badtemperatur verknüpft wurden.

Die Analyse des RF-Regressionsmodells mithilfe einer PFI diente der Identifikation von Prozessvariablen, die einen wesentlichen Einfluss auf die Gegenwartsvorhersage ausüben. Außerdem wird eine Möglichkeit aufgezeigt, die eine Interpretation der Gegenwartsvorhersagen des RF-Modells ermöglichte. Im Rahmen eines abschließenden Tests wurden alle verwendeten Modelle anhand eines reduzierten Testdatensatzes verglichen.

5.4.1 Forecasting (AR-Modell)

In diesem Kapitel wird die Anwendung des in Kapitel 4.2 vorgestellten globalen AR-Modells dargelegt. Dieses wurde mit lokalen Vorhersagemodellen verglichen, um die Eignung des globalen AR-Modells zu überprüfen. Der Fokus lag auf der Badtemperaturvorhersage für den nächsten Tag (Tagesprognose). Vorhersagen für einen größeren Horizont waren nicht Bestandteil dieser Arbeit. Das globale AR-Modell und die lokalen Modelle wurden anhand der Trainingsdaten trainiert. Die folgenden lokalen Modelle wurden mithilfe des R-Pakets `fable` [OWHW23] im Schritt der Evaluation nach Abbildung 5.1 verwendet:

- MEAN
- ETS (auto)
- NAÏVE
- ARIMA (auto)

Das MEAN-Modell berechnet den Mittelwert der Trainingsdaten, der anschließend für die Vorhersage verwendet wird. Das NAÏVE-Modell verwendet jeweils die letzte Beobachtung für die Vorhersage. Exponential Smoothing (ETS) ist ein statistisches Framework [HA21, S. 250, 252], das mithilfe eines Informationskriteriums aus unterschiedlichen Exponential-Smoothing-Modellen ein für die Zeitreihe passendes Modell bestimmt. Die Suche nach einem passenden ETS-Modell wurde in dieser Arbeit mit dem R-Paket `fable` [OWHW23] automatisiert durchgeführt. Darüber hinaus wurde für jeden Badtemperaturverlauf ein ARIMA-Modell in einer automatisierten Vorgehensweise gewählt, wofür der Hyndman-Khandakar-Algorithmus [HK08; HA21, S. 285–287] mithilfe des R-Pakets `fable` zum Einsatz kam. Für die Modelle ETS und ARIMA wurde als Argument eine Periode von zwei Tagen angegeben, um eine potenzielle zweitägige Saisonalität zu berücksichtigen. Weitere Informationen zu den hier genannten lokalen Modellen können der Literatur [HA21] entnommen werden. Als Bewertungsmaß der einzelnen Modelle wurde der Mean Absolute Error (MAE) verwendet, der mithilfe der folgende Gleichung 5.1 berechnet wird [KND15, S. 444; HK06, S. 682]:

$$\text{MAE} = \frac{\sum_{t=1}^n \text{abs}(y_t - \hat{y}_t)}{n} \quad (5.1)$$

y_t ist eine Beobachtung und \hat{y}_t eine Vorhersage jeweils zum Zeitpunkt t . Die Anzahl der Vorhersagen ist mit n angegeben. Im Vergleich zum Root Mean Squared Error (RMSE) ist der MAE weniger empfindlich gegenüber Ausreißern [HK06, S. 682]. Zudem lässt sich der MAE leichter interpretieren. Der Vergleich zwischen den lokalen Modellen und dem globalen AR-Modell wurde mit dem Testdatensatz aus Kapitel 5.3.1 durchgeführt.

Um das globale AR-Modell zu trainieren, wurde ein Lag Embedding mit jedem Badtemperaturverlauf durchgeführt. Nähere Informationen zum Lag Embedding sind in Kapitel 4.2.2 zu finden. Die Ordnung des globalen AR-Modells wurde anhand der CV-BI bestimmt, da diese von Cerqueira et al. [CTM20, S. 2019] für stationäre Zeitreihen vorgeschlagen wird. Wie bereits in Kapitel 5.3.7 festgestellt wurde, ist die Mehrheit der untersuchten Badtemperaturverläufe stationär. Die maximale zu überprüfende Ordnung (Anzahl an Verzögerungsstellen) des AR-Modells wurde auf 35 festgelegt, um die ausgeprägte Frequenzspitze von etwa 30 Tagen zu berücksichtigen (vgl. Kapitel 5.3.6).

Die Anzahl an Folds, die mit der CV-BI erstellt werden, wurde mit $k = 10$ festgelegt. Dieser Wert ist im Kontext der CV ein üblicher Literaturwert, der in der Arbeit von Kohavi [Koh95] ermittelt wurde. In Abbildung 5.23 sind die über die zehn Folds gemittelten MAE-Werte zu sehen, die jeweils gegen die Verzögerungsstellen des globalen AR-Modells aufgetragen sind. Neben dem Mittelwert ist die Standardabweichung für jede Verzögerungsstelle eingezeichnet.

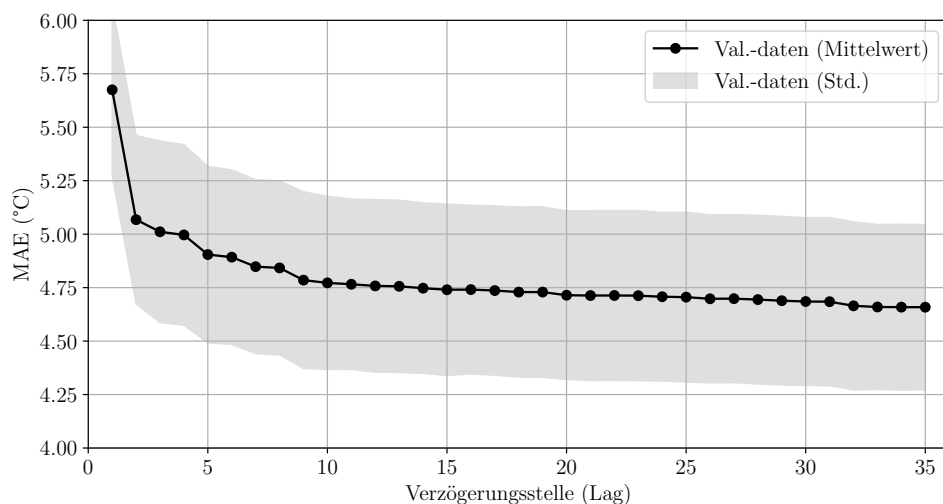


Abbildung 5.23: Gemittelter MAE, der jeweils gegen die berücksichtigten Verzögerungsstellen eines globalen AR-Modells aufgetragen ist.

Aus Abbildung 5.23 geht hervor, dass die größte Verbesserung des MAE mit Hinzufügen der zweiten Verzögerungsstelle erreicht wird. Nach der zehnten Verzögerungsstelle treten keine wesentlichen Verbesserungen mehr auf. Daher wurde für die weitere Vorgehensweise ein

globales AR(10)-Modell mithilfe der gesamten Trainingsdaten trainiert. Anschließend wurde dieses anhand des Testdatensatzes mit den lokalen Modellen verglichen. Der Testdatensatz umfasst die Badtemperaturdaten vom 01.01.2023 bis einschließlich dem 23.04.2023 (vgl. Kapitel 5.3.1). Tabelle 5.4 zeigt den MAE, den das globale AR(10)-Modell und die lokalen Modelle auf dem Testdatensatz erreichen.

Modell	Art der Vorhersage	MAE (°C)
AR(10)	Global	4,59
ETS (auto)	Lokal	4,60
ARIMA (auto)	Lokal	4,67
NAÏVE	Lokal	5,81
MEAN	Lokal	7,95

Tabelle 5.4: MAE, den die Modelle auf den Testdaten erzielten. Angaben sind auf zwei Nachkommastellen gerundet.

Die mit der LOCF-Methode aufgefüllten Badtemperaturwerte (vgl. Kapitel 5.3.4) wurden bei der Berechnung des MAE nicht berücksichtigt. Aus Tabelle 5.4 geht hervor, dass das globale AR(10)-Modell den kleinsten MAE erreicht. Die lokalen Modelle ETS und ARIMA liegen dicht dahinter. Das MEAN-Modell erreicht den höchsten MAE, da dieses lediglich den Mittelwert der Badtemperaturverläufe aus den Trainingsdaten als Vorhersage verwendet. Insgesamt zeigen die Ergebnisse, dass das globale AR(10)-Modell in der Lage ist, die lokalen Modelle hinsichtlich der Badtemperaturvorhersage zu substituieren. Daraus resultiert ein praktischer Vorteil, da ein einziges globales AR(10)-Modell in der Lage ist, vergleichbare Badtemperaturvorhersagen für alle 60 Öfen zu erstellen. Die lokalen Modelle ETS und ARIMA erreichen eine ähnliche Vorhersageperformanz wie das globale AR(10)-Modell, jedoch werden diese pro Ofen trainiert, wodurch 60 ETS und 60 ARIMA Modelle entstehen, die anschließend in den Produktivbetrieb überführt werden müssen. Ein weiterer Vorteil eines bereits trainierten globalen AR-Modells besteht darin, dass dieses auch unmittelbar für neu in Betrieb genommene Öfen eingesetzt werden kann. Ein weiterer Trainingsvorgang ist in diesem Fall nicht mehr notwendig. Dagegen müssen für eine Vorhersage mit einem lokalen Modell zunächst ausreichend Daten eines neuen Ofens verfügbar sein, um ein lokales Modell zu trainieren. Es zeigt sich, dass der Einsatz des globalen AR(10)-Modells gegenüber den lokalen Modellen praktische Vorteile mit sich bringt, ohne eine Verschlechterung hinsichtlich der Vorhersageperformanz eingehen zu müssen.

5.4.2 Nowcasting (Random Forest)

Das globale AR(10)-Modell führt eine Badtemperaturvorhersage für den nächsten Tag (Tagessprognose) anhand vergangener Badtemperaturmessungen durch. Es kann als Basismodell angesehen werden und als Vergleichsgrundlage für weitere Vorhersagemodelle dienen. Wie bereits im Kapitel 3.2 erwähnt, können verschiedene Faktoren die Badtemperatur beeinflussen.

Anhand der Abbildungen 5.12, 5.13, 5.14 und 5.15 wurde der Einfluss mehrerer Prozessvariablen auf die Badtemperatur untersucht. Die Analyse der Auswirkungen einzelner Prozessvariablen auf die Badtemperatur ist jedoch zeitaufwendig. Es stellte sich daher die Frage, ob eine solche Analyse mit einem Vorhersagemodell unterstützt werden kann.

Für die nachfolgende Analyse muss erwähnt werden, dass sich in dem Trainings- und Testzeitraum (vgl. Abbildung 5.5) Öfen befinden, die mit einer neuen Prozesssteuerung ausgestattet wurden. Für diese Öfen wurden bestimmte Prozessvariablen nicht mehr in der ursprünglichen Datenbank aufgezeichnet. Aus diesem Grund wurden Öfen mit neuer Prozesssteuerung vom zweiten Lösungsansatz ausgeschlossen. Die Anzahl verkleinerte sich auf 51 Öfen, die in der anschließenden Untersuchung berücksichtigt wurden.

In unserer Arbeit [GKD+18] wurde ein RF als Regressionsmodell verwendet, um die Badtemperatur anhand von mehreren Prozessvariablen vorherzusagen. Anschließend wurde das RF-Modell mithilfe der PFI und von Partial Dependence Plots analysiert, um den Zusammenhang zwischen den betrachteten Prozessvariablen und der Badtemperaturvorhersage zu bestimmen. Es wurde ausgesagt, dass die über das RF-Modell gefundenen Zusammenhänge zwischen den Prozessvariablen und der Badtemperaturvorhersage nicht unbedingt auf den tatsächlichen Prozess der Aluminiumelektrolyse zutreffen. Die in [GKD+18] erstellten Vorhersagen können auch als Gegenwartsvorhersagen [BGR10, S. 5; SBMW22, S. 375] bezeichnet werden. Dabei werden die bis zu einem bestimmten Zeitpunkt verfügbaren Informationen genutzt, um eine Vorhersage über die Zielvariable (Badtemperatur) durchzuführen [SBMW22, S. 376]. Abbildung 5.24 zeigt den Ablauf der Gegenwartsvorhersage für die Badtemperatur.

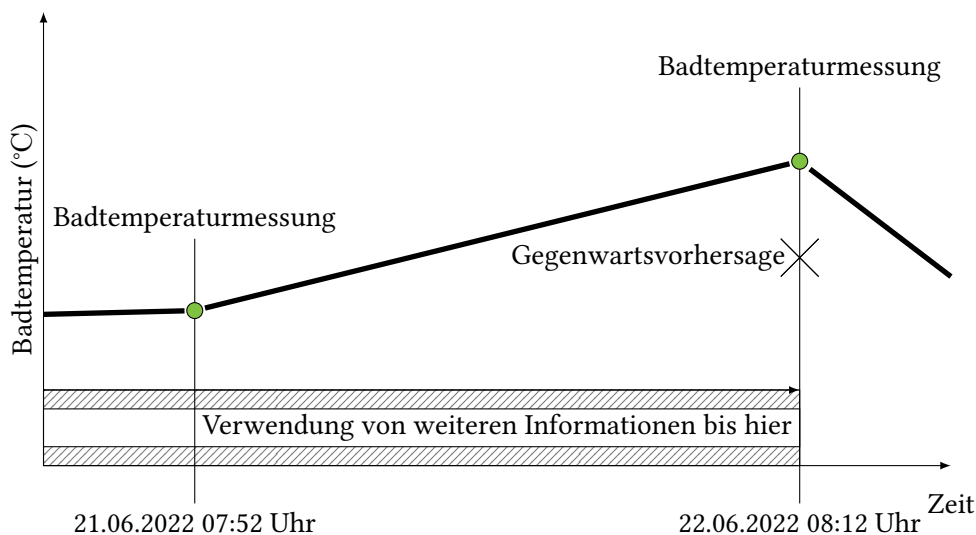


Abbildung 5.24: Ablauf der Gegenwartsvorhersage der Badtemperatur. Abgeändert nach [SBMW22, S. 376].

Die Gegenwartsvorhersage findet hierbei zum Zeitpunkt einer Badtemperaturmessung statt, um die Gegenwartsvorhersage mit der tatsächlichen Badtemperatur vergleichen zu können. Der tatsächliche Verlauf der Badtemperatur ist zwischen den Badtemperaturmessungen unbekannt. Wie bereits in unserer Arbeit [GKD+18, S. 4] erwähnt wird, werden in der Aluminiumindustrie die Prozessvariablen eines Ofens mit unterschiedlichen zeitlichen Auflösungen aufgezeichnet. Anhand der in Kapitel 3.2 und Kapitel 5.3.3 aufgezeigten Einflüsse auf die Badtemperatur wurde eine Vorauswahl an Prozessvariablen getroffen, die anschließend für das RF-Regressionsmodell eingesetzt wurden. Die in Tabelle 5.5 dargestellten Prozessvariablen bildeten die Ausgangsbasis für die in dieser Arbeit durchgeführten Gegenwartsvorhersage.

Prozessvariable	Einheit/Datentyp	Auflösung
Hallenstrom	A	5 min
Ofenspannung	V	5 min
Fütterungsrate	%	5 min
Zugabe von AlF_3	bool	5 min
Anodenwechsel	bool	5 min
Metallsaugen	bool	5 min
Anodeneffekt	bool	5 min
AlF_3 -Durchläufer entdeckt	bool	Zeitpunkt der Entdeckung
Flusssaugen	kg	Schichtzeitstempel
Zugabe von Soda (Na_2CO_3)	kg	Schichtzeitstempel
Ansatz entdeckt	Anzahl	Schichtzeitstempel

Tabelle 5.5: Für die Gegenwartsvorhersage verwendete Prozessvariablen.

Die in Tabelle 5.5 aufgezeigten Prozessvariablen stammen aus verschiedenen Datenquellen, die für diese Arbeit zusammengeführt wurden, um eine Gegenwartsvorhersage der Badtemperatur zu ermöglichen. Eine Schwierigkeit bei der Zusammenführung der Prozessvariablen waren die unterschiedlichen zeitlichen Auflösungen. Während für den Hallenstrom und die Ofenspannung alle fünf Minuten ein Wert vorlag, wurde für die Entdeckung eines Ansatzes oder die Zugabe von Soda (Na_2CO_3) nur der Beginn der Schicht als Zeitstempel angegeben, in der der Ansatz entdeckt wurde oder eine Zugabe von Soda erfolgt ist. Daher wurde bei der Vorverarbeitung der Prozessvariablen darauf geachtet, dass alle Prozessvariablen zeitlich korrekt miteinander verknüpft wurden. Als Referenz wurden die Zeitstempel der 5 min-Daten genutzt, an die alle weiteren Daten angefügt wurden. Diese Vorgehensweise gewährleistet, dass nur zum Zeitpunkt der Gegenwartsvorhersage bereits vorhandene Daten einbezogen wurden.

Damit das RF-Regressionsmodell aus unterschiedlich aufgelösten Daten lernen kann, wurden die Prozessvariablen geeignet aufbereitet. In [GKD+18, S. 4; SBMW22, S. 375–376] werden dazu Zeitreihenmerkmale innerhalb von Zeitintervallen über die betrachteten Prozessvariablen gebildet. Dabei wird der jeweilige Verlauf der Prozessvariablen innerhalb der zeitlichen Intervalle mit einem numerischen Wert quantifiziert. Als Zeitreihenmerkmale werden in [GKD+18,

S. 4; SBMW22, S. 376] unter anderem der Mittelwert, der Median und die Standardabweichung verwendet.

In dieser Arbeit wurden ebenfalls Zeitreihenmerkmale anhand der in Tabelle 5.5 dargestellten Prozessvariablen berechnet. Für den Hallenstrom, die Ofenspannung sowie die Fütterungsrate wurde jeweils der Mittelwert berechnet, während für die restlichen Prozessvariablen die Anzahl ihrer Vorkommnisse (Summe) berechnet wurde. Das Zeitintervall für die Berechnung der Zeitreihenmerkmale wurde mit 24 Stunden festgelegt. Die Wahl des gewählten Zeitintervalls und der Zeitreihenmerkmale basiert auf den Ergebnissen unserer Arbeit [GKD+18] und gewährleistet, dass vergangene Werte der Prozessvariablen für die Gegenwartsvorhersage berücksichtigt werden. Mittels dieser Vorgehensweise kann die Performanz der Gegenwartsvorhersage gesteigert werden [GKD+18, S. 5]. Abbildung 5.25 zeigt die Berechnung von Zeitreihenmerkmalen über die 5 min-Daten in dem 24 Stunden Intervall. Anschließend wurden die berechneten Zeitreihenmerkmale mit der jeweiligen Badtemperaturmessung verknüpft.

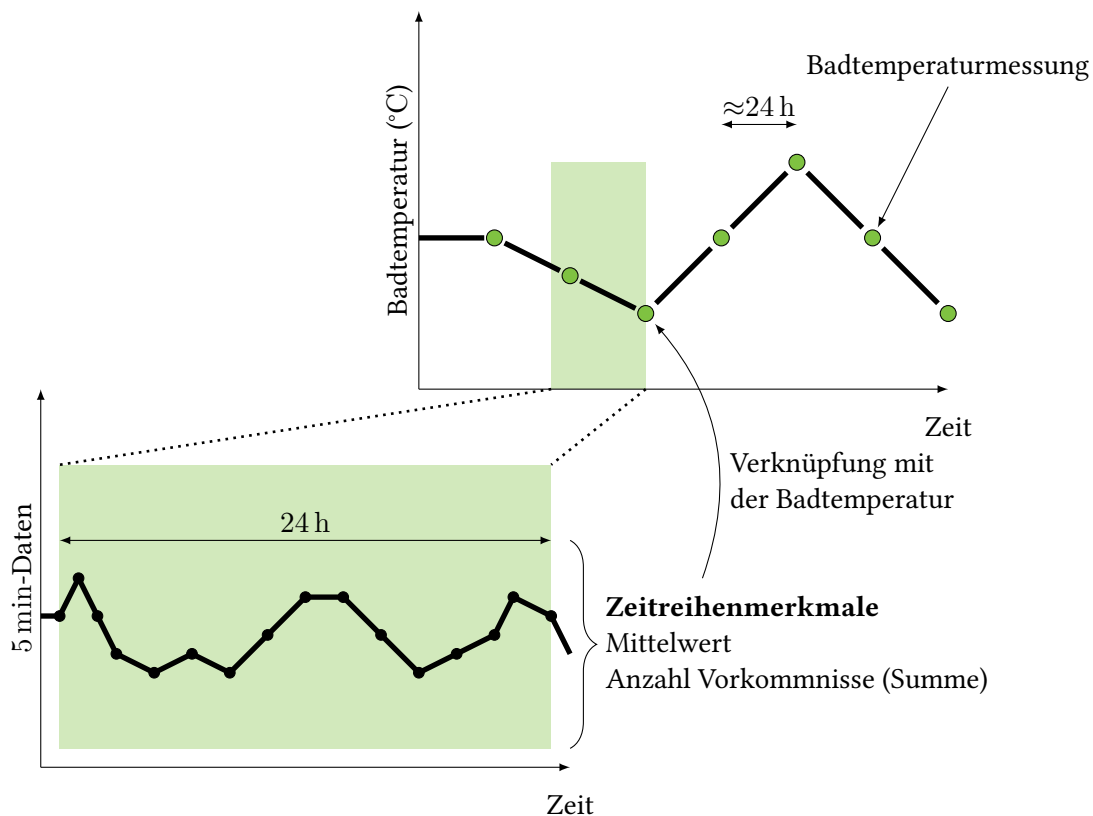


Abbildung 5.25: Erstellung von Zeitreihenmerkmalen über die 5 min-Daten in einem 24 Stunden Intervall. Abgeändert nach [KGD+21, S. 771].

Für die Validierung des RF-Regressionsmodells wurde die CV-BI mit zehn Folds durchgeführt. Die in Abbildung 5.5 dargestellte Aufteilung des Trainings- und Testzeitraums blieb weiterhin

erhalten. Die Validierung des RF-Regressionsmodells wurde ausschließlich auf den Trainingsdaten durchgeführt. Die Bewertung des Modells erfolgte über den MAE nach Gleichung 5.1. Während der Validierung wurden mehrere Versuche durchgeführt und die Auswirkungen auf die Modellperformanz ausgewertet. Im ersten Versuch wurden die gemessene Badtemperatur als Zielvariable und die zuvor beschriebenen Zeitreihenmerkmale verwendet. Im zweiten Versuch wurde zusätzlich der Wert der letzten Badtemperaturmessung als Merkmal berücksichtigt. Beide Versuche wurden wiederholt, wobei diesmal die Änderung der Badtemperatur als Zielvariable definiert wurde (Versuch 3 und 4). Bei allen Versuchen wurde darauf geachtet, dass jeweils die gleichen Validierungsdaten verwendet wurden, um einen fairen Vergleich zwischen den Versuchen zu gewährleisten. Während der Validierung wurden unterschiedliche Hyperparameterkombinationen validiert. In Tabelle 5.6 ist der kleinste auf zwei Nachkommastellen gerundete MAE für jeden Versuch zu finden, der jeweils mit der folgenden Hyperparameterkombination ermittelt wurde: Anzahl an Bäumen: 100, Maximale Anzahl an zufälligen Daten für das Training eines Baumes: 1000, Minimale Anzahl an Daten pro Blattknoten: 1, Kein Subspace Sampling. Es sei darauf hingewiesen, dass für den dritten Versuch ein geringfügig kleinerer MAE erreicht wurde, wenn ein Subspace Sampling mit $\sqrt{\text{Anzahl der Merkmale}}$ durchgeführt wird. Gerundet entspricht dieser MAE jedoch dem angegebenen Wert in Tabelle 5.6.

Versuch	MAE (°C)
1	6,39 ± 0,37
2	4,59 ± 0,34
3	4,84 ± 0,40
4	4,55 ± 0,36

Tabelle 5.6: MAE mit Standardabweichung, die in vier Versuchen innerhalb der CV-BI für das RF-Regressionsmodell ermittelt wurden. Angaben sind auf zwei Nachkommastellen gerundet.

Aus Tabelle 5.6 geht hervor, dass im vierten Versuch der kleinste MAE erreicht wird. In diesem Versuch wurde die Badtemperaturänderung als Zielvariable definiert und als zusätzliches Merkmal der Wert der Badtemperaturmessung vom Vortag berücksichtigt. Es ist zu vermuten, dass ähnliche Ausprägungen der einzelnen Merkmale zu einer ähnlichen Badtemperaturänderung führen. So war das RF-Regressionsmodell besser in der Lage, die Daten hinsichtlich der Zielvariable in homogene Gruppen einzuteilen. Wird beispielsweise davon ausgegangen, dass jeder Anodenwechsel die gleiche Badtemperaturänderung hervorruft, dann ist dieser Zusammenhang vermutlich besser durch das RF-Regressionsmodell zu erlernen als der Zusammenhang zwischen Anodenwechsel und der Badtemperatur.

Diese Vermutung wird von den Ergebnissen der weiteren Versuche bestärkt. Im ersten Versuch wurde lediglich die Badtemperatur vorhergesagt, was zum größten MAE innerhalb der CV-BI führt. Der MAE verbessert sich bereits im zweiten Versuch wesentlich, bei dem der

Wert der letzten Badtemperaturmessung als zusätzliches Merkmal berücksichtigt wurde. Daher hatte das RF-Regressionsmodell die Möglichkeit, die vorherige Badtemperatur in der Gegenwartsvorhersage zu berücksichtigen. Der dritte Versuch zeigt gegenüber dem ersten Versuch eine wesentliche Verbesserung hinsichtlich der Vorhersageperformanz und bestätigt die zuvor aufgestellte Vermutung. In diesem Versuch wurde lediglich die Badtemperaturänderung vorhergesagt. Wie bereits erwähnt, wird der kleinste MAE im vierten Versuch erreicht, bei dem die Badtemperaturänderung als Zielvariable definiert und der Wert der letzten Badtemperaturmessung zusätzlich berücksichtigt wurden. Aus den Versuchen lässt sich insgesamt ableiten, dass eine Vorhersage der Badtemperaturänderung gegenüber der Badtemperatur zu bevorzugen ist, da eine bessere Modellperformanz erreicht wird.

Um einen Einblick in das Zusammenspiel zwischen den verwendeten Merkmalen und der Zielvariable zu erhalten, wurde die PFI berechnet. Die PFI gibt an, wie wesentlich ein Merkmal für die Vorhersage der Zielvariable ist. Dazu wird zunächst ein entsprechendes Vorhersagemodell trainiert und die Performanz anhand von Validierungsdaten ermittelt. Diese ermittelte Performanz stellt die Basis für die Berechnung der PFI dar. Anschließend werden die Ausprägungen eines Merkmals aus den Validierungsdaten zufällig permutiert. Mit dem permutierten Merkmal und allen weiteren Merkmalen wird dann erneut eine Performanz ermittelt und aufgezeichnet. Dieses Vorgehen wird solange wiederholt, bis alle Ausprägungen eines jeden Merkmals zufällig permutiert wurden. Der Unterschied zwischen der Basisperformanz und der Performanz, die mit einem zufällig permutierten Merkmal ermittelt wird, gibt eine Einschätzung darüber, wie wesentlich das entsprechende Merkmal für die Vorhersage ist [Bre01a, S. 23].

Bei der Interpretation der PFI sollten die Korrelationen zwischen den einzelnen Merkmalen berücksichtigt werden. In Abbildung A.5 sind die Korrelationen der berechneten Zeitreihenmerkmale und der beiden Zielvariablen, Badtemperatur und Badtemperaturänderung, angegeben. Die angegebenen Korrelationskoeffizienten wurden nach Pearson berechnet. In Abbildung A.6 sind zudem die Korrelationskoeffizienten zu sehen, die nach Spearman berechnet wurden. Merkmale, die miteinander stark korrelieren, können sich gegenseitig substituieren. Bestehende Korrelationen zwischen Merkmalen sollten daher bei der Analyse der PFI beachtet werden, indem die korrelierenden Merkmale bei der Berechnung der PFI gemeinsam permutiert werden [PTCH18]. Aus den Korrelationen geht insgesamt hervor, dass zwischen den Merkmalen keine stärkeren Korrelationen vorliegen, die eine Zusammenführung von Merkmalen für die Berechnung der PFI rechtfertigten. Die stärkste Korrelation besteht in Abbildung A.5 zwischen der Ofenspannung und dem Hallenstrom, während in Abbildung A.6 die stärkste Korrelation zwischen der Fütterungsrate und dem Anodenwechsel vorliegt. Wie bereits erläutert, sind auch die Korrelationen der Zielvariablen in beiden Abbildungen A.5 und A.6 zu sehen. Aus diesen geht hervor, dass die Badtemperatur (Zielvariable) mit der vorherigen Badtemperatur (Messwert vom Vortag) korreliert. Eine solche Korrelation wurde bereits aus der berechneten Autokorrelation in Kapitel 5.3.5 ersichtlich. Mit der Badtemperaturänderung (Zielvariable)

Abschnitt 5.4. Lösungsentwicklung

korrelieren der Anodenwechsel und die vorherige Badtemperatur am stärksten. In Abbildung 5.26 ist die berechnete PFI als Balkendiagramm mit zugehöriger Standardabweichung für jedes Merkmal dargestellt. Die PFI wurde während der CV-BI auf den jeweiligen Validierungsdaten für jeden der vier Versuche anhand des RF-Regressionsmodells ermittelt. Für die Messung der jeweiligen Performanz wurde der MAE berechnet. Die Daten geben die Änderung des MAE für die einzelnen Merkmale an.

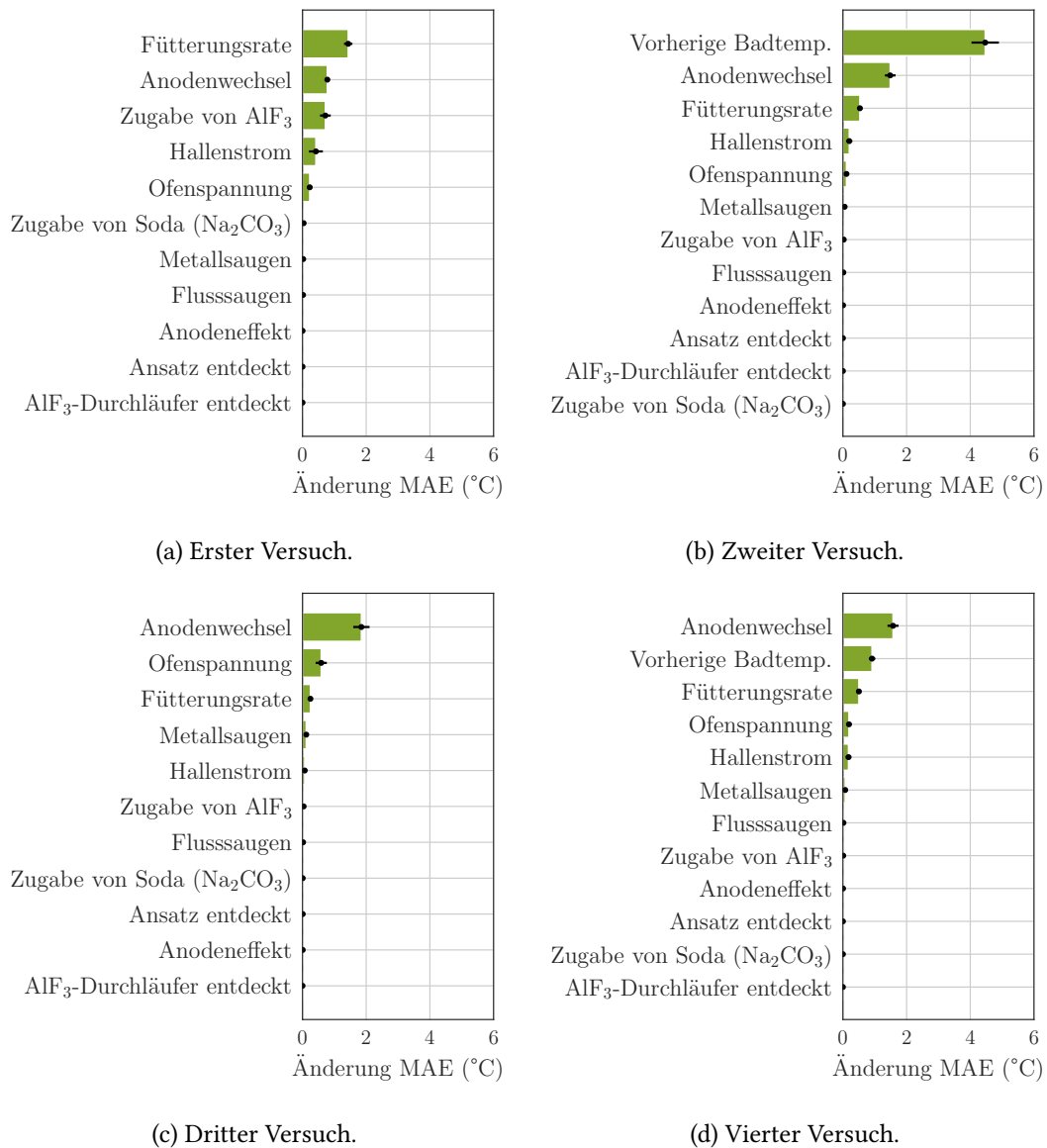


Abbildung 5.26: Berechnete PFI für jeden Versuch. (a) Die Zielvariable ist die Badtemperatur. (b) Die Zielvariable ist die Badtemperatur. Der Wert der letzten Badtemperaturmessung wird als Merkmal berücksichtigt. (c) Die Zielvariable ist die Badtemperaturänderung. (d) Die Zielvariable ist die Badtemperaturänderung. Der Wert der letzten Badtemperaturmessung wird als Merkmal berücksichtigt.

Im ersten Versuch (vgl. Abbildung 5.26a) kommt die größte Änderung des MAE durch das Merkmal Fütterungsrate zustande. In diesem Fall steigt der MAE um etwa $1,4^{\circ}\text{C}$ gegenüber der Basisperformanz. Im zweiten Versuch (vgl. Abbildung 5.26b) verursacht der Wert der letzten Badtemperaturmessung die größte Änderung des MAE, während im dritten und vierten Versuch (vgl. Abbildung 5.26c und 5.26d) jeweils der Anodenwechsel sich als wesentliches Merkmal herausstellt. Wie bereits in [GKD+18, S. 9] festgestellt, gilt es bei der Interpretation der PFI zu beachten, dass diese nicht zwangsläufig den kausalen Zusammenhang zwischen den Merkmalen und der Zielvariable darstellen. Dennoch gibt die PFI einen ersten Überblick über Merkmale, die wesentlich für die Vorhersage der Badtemperatur sind.

Das RF-Regressionsmodell aus dem vierten Versuch wurde näher untersucht, da dieses den geringsten MAE erzielt hat. Die für den vierten Versuch berechnete PFI (vgl. Abbildung 5.26d) zeigt, dass einige Merkmale kaum zu einer Änderung des MAE beitragen. Aus diesem Grund wurden Merkmale mit einem geringen Beitrag entfernt und erneut ein RF-Regressionsmodell mit einer reduzierten Anzahl an Merkmalen mithilfe der CV-BI trainiert und validiert. Folgende Merkmale wurden weiter berücksichtigt: Anodenwechsel, Wert der Badtemperaturmessung vom Vortag (vorherige Badtemp.), Fütterungsrate, Ofenspannung, Hallenstrom und Metallsaugen. Die restlichen Merkmale wurden aufgrund ihrer geringen Werte in der PFI verworfen. Als Zielvariable wurde die Änderung der Badtemperatur definiert und unterschiedliche Hyperparameterkombinationen validiert. Der kleinste MAE von $4,56^{\circ}\text{C} \pm 0,36^{\circ}\text{C}$ wurde mit den Hyperparametern aus dem ersten Durchlauf ohne Subspace Sampling erzielt. Der MAE ist nur geringfügig schlechter als der MAE aus dem vierten Versuch im ersten Durchlauf (vgl. Tabelle 5.6).

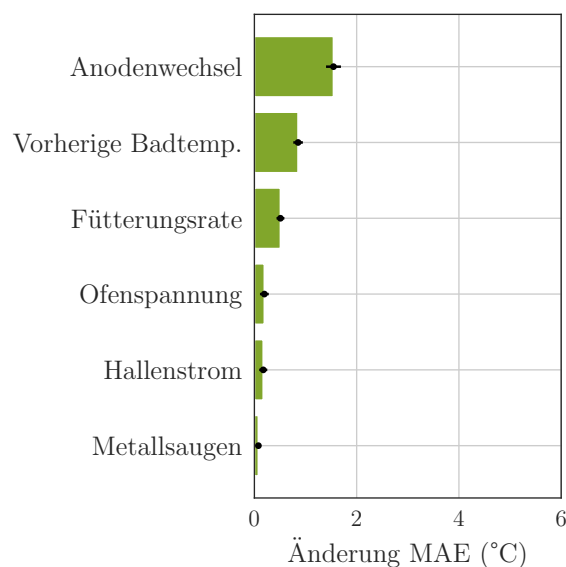


Abbildung 5.27: PFI des RF-Regressionsmodells mit reduzierten Merkmalen.

Diese Erkenntnis legt nahe, dass die ausgeschlossenen Merkmale nicht wesentlich für die Vorhersage der Badtemperatur sind. Abbildung 5.27 veranschaulicht die PFI des RF-Regressionsmodells mit den reduzierten Merkmalen. Ein Vergleich mit Abbildung 5.26d zeigt, dass die Reihenfolge der dargestellten Merkmale erhalten bleibt.

Im nächsten Schritt wurde das RF-Regressionsmodell auf den kompletten Trainingsdaten trainiert, um das endgültige Modell zu erhalten. Für den Trainingsvorgang wurde die Hyperparameterkombination aus dem ersten bzw. zweiten Durchlauf verwendet. Die Anzahl an Entscheidungsbäumen wurde jedoch auf 1000 erhöht. Nach [HG20, S. 299] verbessert sich die Modellperformanz, je mehr Entscheidungsbäume trainiert werden. Eine zusätzlich in dieser Arbeit durchgeführte CV-BI unterstützt diese Aussage. Allerdings verbesserten 2000 Entscheidungsbäume die Modellperformanz nicht weiter. Aus diesem Grund wurden 1000 Entscheidungsbäume für den abschließenden Test verwendet.

Das Modell wurde anschließend auf den Testdaten getestet, um eine abschließende Performanz zu ermitteln. Bei diesem Test erreichte das RF-Regressionsmodell einen MAE von 4,53 °C auf den Testdaten. Auch in diesem abschließenden Test wurden die Badtemperaturwerte, die durch die LOCF-Methode ersetzt wurden (vgl. Kapitel 5.3.4), nicht berücksichtigt. In Abbildung 5.28 ist weiterhin der MAE für jeden Ofen dargestellt, den das RF-Regressionsmodell auf den entsprechenden Testdaten erzielt hat.

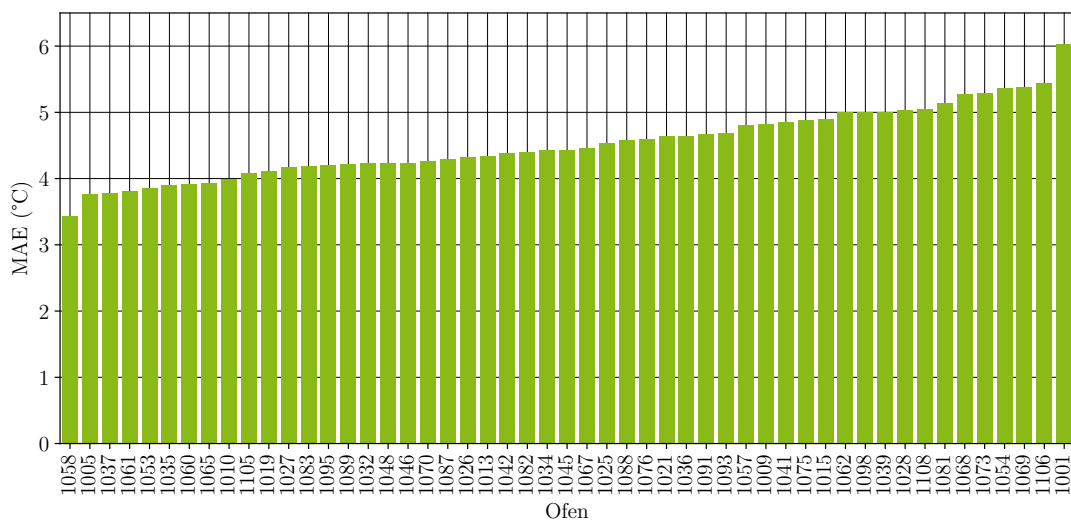


Abbildung 5.28: MAE für jeden Ofen, den das RF-Regressionsmodell auf den Testdaten erzielt.

Der Abbildung 5.28 ist zu entnehmen, dass das RF-Regressionsmodell für den Ofen 1001 den höchsten MAE und für den Ofen 1058 den geringsten MAE erzielte. Für die beiden Öfen 1001 und 1058 lassen sich exemplarisch die Gegenwartsvorhersagen des RF-Modells, die Standardab-

weichung für jede Vorhersage (σ RF) und die gemessene Badtemperatur (Ziel) vom 09.04.2023 bis zum 23.04.2023 in Abbildung 5.29 entnehmen. In dieser sind die Gegenwartsvorhersagen der Badtemperaturänderung und der Badtemperatur zu sehen. Die Vorhersage der Badtemperatur ergibt sich hierbei aus der Vorhersage der Badtemperaturänderung. Die Standardabweichung wurde über die Gegenwartsvorhersagen der einzelnen Entscheidungsbäume berechnet. Die Standardabweichung dient als Maß für die Unsicherheit der jeweiligen Vorhersage. Eine kleine Standardabweichung bedeutet ein größeres Vertrauen, das in die Vorhersage gesetzt werden kann. Vorhersagen, die eine relativ hohe Standardabweichung aufweisen, sollten mit Vorsicht betrachtet werden [HG20, S. 302–303]. Zum Beispiel besitzt die Gegenwartsvorhersage des Ofens 1001 für den 10.04.2024 eine relativ hohe Standardabweichung.

Für den Ofen 1001 ist in Abbildung 5.29 zu erkennen, dass für den 15.04.2023 kein Messwert bzw. Zielwert vorhanden ist. Lediglich die Gegenwartsvorhersage des RF-Regressionsmodells ist zu sehen, die für diesen Tag eine negative Badtemperaturänderung prognostiziert. Die Vorhersage wurde zu Beginn der dritten Schicht getätigt, da diese Schicht für die Badtemperaturmessung des Ofens 1001 zuständig ist (vgl. Abbildung A.1).

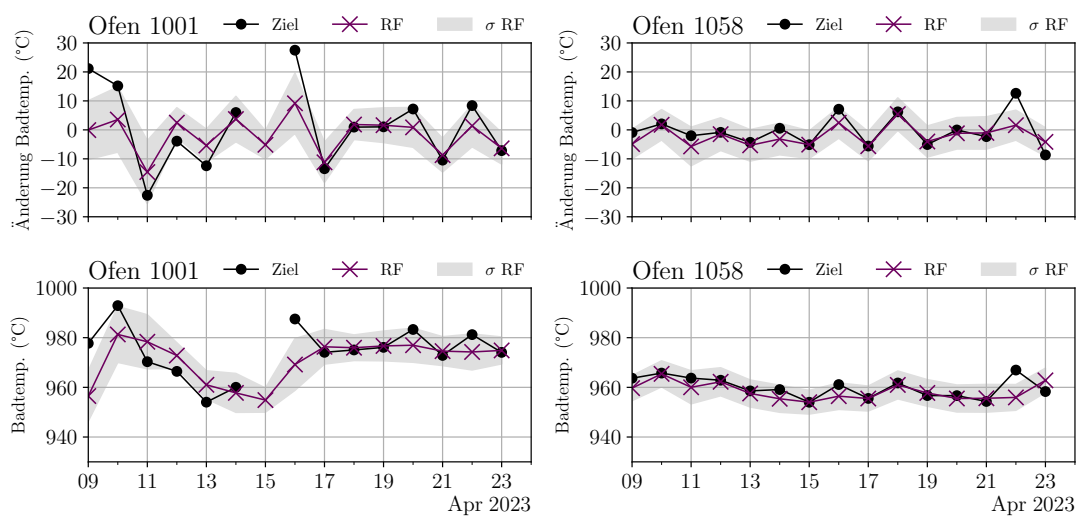


Abbildung 5.29: Zeitlich aufgetragene Gegenwartsvorhersagen des RF-Regressionsmodells für die Öfen 1001 und 1058. Zum Vergleich sind weiterhin die tatsächlich gemessenen Badtemperaturen aufgetragen. Die Standardabweichung (σ RF) gibt die Streuung der Gegenwartsvorhersagen der einzelnen Entscheidungsbäume des RF-Modells an.

5.4.3 Analyse (Nowcasting)

Die Gegenwartsvorhersagen der einzelnen Entscheidungsbäume, die in dem RF-Regressionsmodell enthalten sind, werden zu einer Gesamtvorhersage aggregiert. Mit dem Aufbau eines Entscheidungsbaums lässt sich der Beitrag der Merkmale zu jeder Gegen-

wartsvorhersage ermitteln. Hierzu wird der Pfad vom Wurzelknoten bis zum Blattknoten eines jeden Entscheidungsbaums für eine bestimmte Vorhersage analysiert [Saa14]. Eine solche Analyse hilft bei der Interpretation einer jeden Gegenwartsvorhersage und bietet die Möglichkeit einer Ursachenanalyse von Badtemperaturänderungen, die nach bestem Wissen im Bereich der Aluminiumelektrolyse noch nicht aufgezeigt wurde.

Für die Interpretation der Gegenwartsvorhersagen wurde das Python-Paket `treeinterpreter` [Saa21] in der Version 0.2.3 verwendet. Um den Einfluss der einzelnen Merkmale auf die Gegenwartsvorhersage darstellen zu können, wurde des Weiteren ein Waterfall-Chart eingesetzt, der mithilfe des Python-Pakets `waterfallcharts` [Csi17] in der Version 3.8 mit eigenen Modifikationen realisiert wurde.

Ausgehend von dem Bias, der sich aus den Trainingsdaten ergibt, sind in Abbildung 5.30 die einzelnen Beiträge der Merkmale zu sehen, aus denen sich die Gesamtvorhersage für den 14.04.2024 für den Ofen 1001 zusammensetzt. Die Summe der einzelnen Beiträge ergibt die Gegenwartsvorhersage des RF-Modells. An dieser Stelle wird darauf hingewiesen, dass die in Abbildung 5.30 aufgezeigten Merkmale die berechneten Zeitreihenmerkmale darstellen, die innerhalb des 24 Stunden Intervalls berechnet wurden (vgl. Kapitel 5.4.2). Zum Beispiel stellt das Merkmal Anodenwechsel die Anzahl der Anodenwechsel der letzten 24 Stunden dar. Für die Ofenspannung oder den Hallenstrom ist das der jeweilige Mittelwert in diesem Intervall. Aus Abbildung 5.30 geht hervor, dass der Anodenwechsel den größten Beitrag zur Gegenwartsvorhersage vom 14.04.2023 des Ofens 1001 liefert.

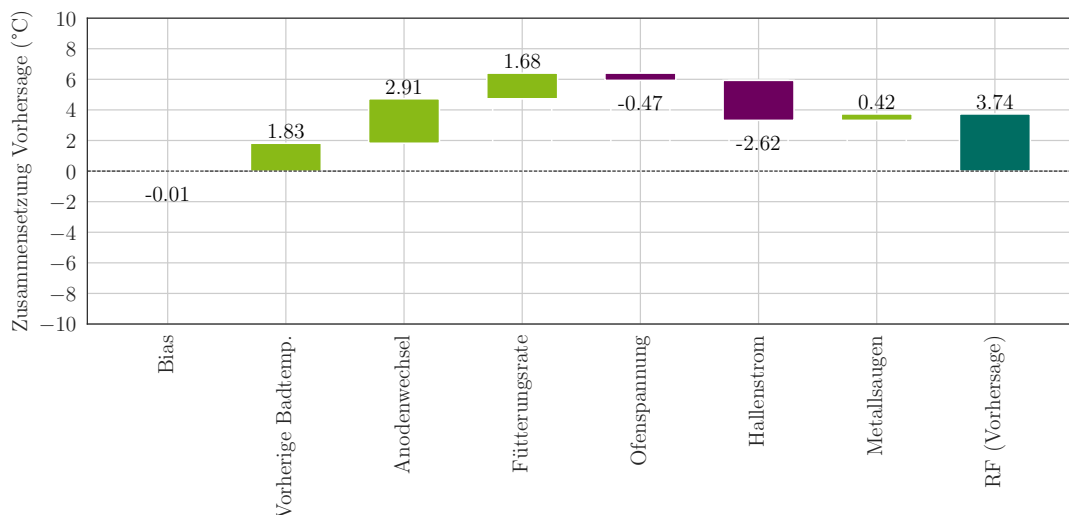


Abbildung 5.30: Ein Waterfall-Chart, der die Zusammensetzung der Gegenwartsvorhersage vom 14.04.2023 des Ofens 1001 darstellt. Angaben sind auf zwei Nachkommastellen gerundet.

Erwähnenswert hierbei ist, dass ausgehend vom Zeitpunkt der Gegenwartsvorhersage in den letzten 24 Stunden kein Anodenwechsel am Ofen 1001 stattgefunden hat. Das führt zu einem positiven Beitrag bei der Gegenwartsvorhersage. Auch der Hallenstrom trägt mit einem negativen Beitrag wesentlich zur Gesamtvorhersage bei. Die Summe aus Bias und den Beiträgen der Merkmale ergibt die Gegenwartsvorhersage, die mit $3,74^{\circ}\text{C}$ auf der rechten Seite der Abbildung 5.30 angegeben ist. Ein weiteres Beispiel ist in Abbildung 5.31 zu sehen. In dieser sind die Beiträge der einzelnen Gegenwartsvorhersage vom 19.04.2023 für den Ofen 1058 zu sehen. Die vorhergesagte Badtemperaturänderung beträgt $-4,01^{\circ}\text{C}$. Wesentlich für diese Gegenwartsvorhersage ist das Merkmal Anodenwechsel. In dem 24-Stunden-Zeitraum haben zwei Anodenwechsel stattgefunden, die insgesamt mit $-4,06^{\circ}\text{C}$ zur Gegenwartsvorhersage beitragen.

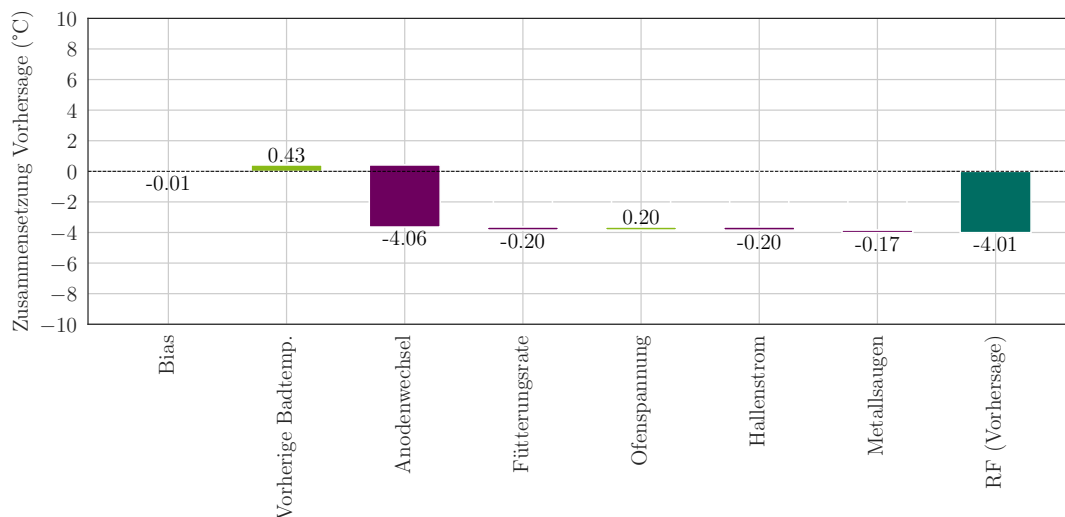


Abbildung 5.31: Ein Waterfall-Chart, der die Zusammensetzung der Gegenwartsvorhersage vom 19.04.2023 des Ofens 1058 aufzeigt. Angaben sind auf zwei Nachkommastellen gerundet.

Eine solche Zusammensetzung kann auch für jede weitere Gegenwartsvorhersage des RF-Regressionsmodells erzeugt werden, um eine Interpretation der Gegenwartsvorhersagen zu ermöglichen. Anhand des Waterfall-Charts können Prozessvariablen identifiziert werden, die die Gegenwartsvorhersage maßgeblich beeinflussen. Allerdings muss bei diesen Analysen beachtet werden, dass die Zusammenhänge nicht unbedingt den realen Prozess der Aluminiumherstellung widerspiegeln, wie bereits in unserer Arbeit [GKD+18, S. 9] festgestellt wurde. Die hier gezeigten Beiträge, aus denen sich die jeweilige Gegenwartsvorhersage zusammensetzt, lassen dennoch Rückschlüsse auf die Modellzusammenhänge zu, die von den Trainingsdaten erlernt wurden. Die Interpretation der Gegenwartsvorhersagen (Ursachenanalyse) muss allerdings im Praxisbetrieb der Aluminiumelektrolyse erprobt werden.

5.4.4 Evaluation

Der MAE von 4,53 °C des RF-Regressionsmodells wurde anhand eines Testdatensatzes ermittelt, der lediglich Öfen ohne neue Prozesssteuerung beinhaltet. Ein direkter Vergleich mit den lokalen Modellen und dem globalen AR(10)-Modell (vgl. Tabelle 5.4) war daher zunächst nicht möglich. Um dennoch einen Vergleich zwischen den einzelnen Modellen durchführen zu können, wurde jeweils die Performanz der lokalen Modelle und des globalen AR(10)-Modells auf dem reduzierten Testdatensatz berechnet. So war ein direkter Vergleich zwischen allen Modellen möglich, da die gleichen Testdaten verwendet wurden. Auch für diesen Test wurden die Badtemperaturen, die durch die LOCF-Methode ermittelt wurden (vgl. Kapitel 5.3.4), vom Test ausgeschlossen. Die Ergebnisse können Tabelle 5.7 entnommen werden.

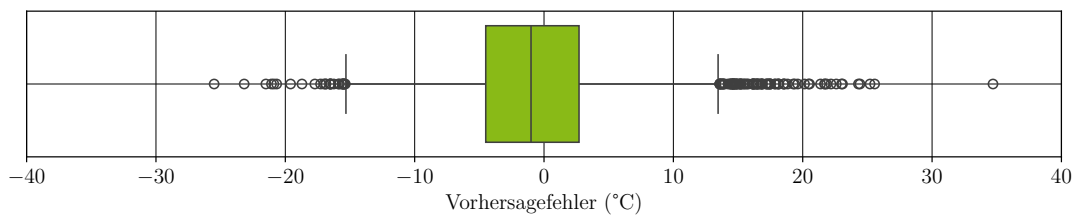
Modell	Art der Vorhersage	MAE (°C)
RF	Nowcasting (Global)	4,53
AR(10)	Forecasting (Global)	4,55
ETS (auto)	Forecasting (Lokal)	4,55
ARIMA (auto)	Forecasting (Lokal)	4,62
NAÏVE	Forecasting (Lokal)	5,76
MEAN	Forecasting (Lokal)	7,88

Tabelle 5.7: MAE, der von dem jeweiligen Modell auf den Testdaten (ohne Öfen mit neuer Prozesssteuerung) erzielt wurde. Die Angaben wurden zunächst aufsteigend sortiert und anschließend auf zwei Nachkommastellen gerundet. Als Forecasting wird die Tagesprognose bezeichnet.

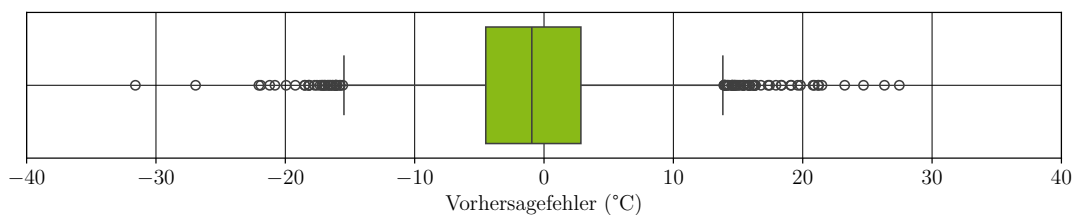
Aus Tabelle 5.7 geht hervor, dass das globale RF-Modell mit einem MAE von 4,53 °C die beste Performanz auf den Testdaten erzielt. Es sei darauf hingewiesen, dass dieses im Vergleich zu den restlichen Modellen eine Gegenwartsvorhersage durchführt. Auf das RF-Modell folgen das globale AR(10)-Modell und das lokale ETS-Modell mit einem MAE von jeweils 4,55 °C. Den höchsten MAE erreicht das MEAN-Modell. Der Performanzvergleich zeigt auch hier, dass das globale AR(10)-Modell eine Alternative zu den lokalen Modellen ist, um die Badtemperatur vorherzusagen. Während bei den lokalen Modellen für jede Zeitreihe ein entsprechendes Modell trainiert wird, wird beim globalen Ansatz ein Modell anhand von allen verfügbaren Zeitreihen trainiert. Daraus ergibt sich ein praktischer Vorteil, da mit einem einzigen globalen Modell Badtemperaturvorhersagen für mehrere Aluminiumelektrolyseöfen erstellt werden können. Je nach Einsatzszenario sind beide globalen Modelle für eine Vorhersage der Badtemperatur zu empfehlen. Für eine Tagesprognose kann das globale lineare AR-Modell als Basismodell eingesetzt werden. Bei der Gegenwartsvorhersage mit dem RF-Modell steht zusätzlich die Möglichkeit einer Interpretation der erzeugten Vorhersagen zur Verfügung.

In Abbildung 5.32 sind die Verteilungen der Vorhersagefehler auf den Testdaten jeweils anhand eines Boxplots für beide globalen Modelle zu sehen. Der Vorhersagefehler ist die Differenz

zwischen der gemessenen Badtemperatur und der Vorhersage. Bei beiden Boxplots ist der Median jeweils negativ. Das bedeutet, dass die Vorhersagen beider globalen Modelle im Median höher liegen als die tatsächlich gemessene Badtemperatur. Zudem sind für beide globalen Modelle Ausreißer hinsichtlich der jeweiligen Vorhersagefehler zu erkennen. Für das globale AR(10)-Modell in Abbildung 5.32a liegt der größte absolute Vorhersagefehler bei $34,71^\circ\text{C}$. Beim RF-Modell in Abbildung 5.32b liegt der größte absolute Fehler bei $31,59^\circ\text{C}$. Die Standardabweichung der Vorhersagefehler liegt für das globale AR(10)-Modell bei $5,82^\circ\text{C}$ und für das RF-Modell bei $5,72^\circ\text{C}$.



(a) Boxplot für die Vorhersagefehler des globalen AR(10)-Modells.



(b) Boxplot für die Vorhersagefehler des globalen RF-Regressionsmodells.

Abbildung 5.32: Verteilungen der Vorhersagefehler der globalen Modelle auf den Testdaten.

Eine weitere Idee bestand noch darin, die Vorhersagen der beiden globalen Modelle miteinander zu kombinieren, um eine potenzielle Verbesserung hinsichtlich der Vorhersageperformanz zu erreichen. Dazu wurde der Mittelwert anhand der Gegenwartsvorhersage des globalen RF-Regressionsmodells und der Tagesprognose des globalen AR(10)-Modells für jeden Tag berechnet. Die gemittelten Vorhersagen wurden anschließend auf den Testdaten evaluiert. Die Modellkombination erreichte einen MAE von $4,31^\circ\text{C}$ auf den Testdaten. Aus diesem Ergebnis ist ersichtlich, dass mittels einfacher Mittelung eine Verbesserung der Vorhersageperformanz erreicht werden konnte. In Abbildung 5.33 ist ein Boxplot für die Vorhersagefehler der Modellkombination zu sehen. Aus dieser geht unter anderem hervor, dass der größte absolute Vorhersagefehler unter 30°C liegt. Die Standardabweichung der Vorhersagefehler der Modellkombination beträgt $5,49^\circ\text{C}$ und ist kleiner als die der einzelnen globalen Modelle.

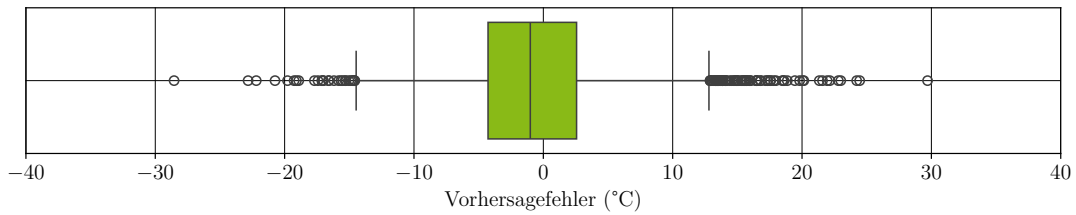


Abbildung 5.33: Boxplot für die Vorhersagefehler der Modellkombination auf den Testdaten.

5.5 Zusammenfassung

In diesem Kapitel wurde zunächst die umfangreiche Aufbereitung der Badtemperaturdaten aufgezeigt. Es wurden Daten von Öfen aus der ersten Halle der TAE in einem Zeitraum vom 01.01.2022 06:00 Uhr bis zum 24.04.2023 06:00 Uhr betrachtet. Da die Messzeitpunkte der Badtemperatur eines Ofens variieren, wurden die Zeitstempel angepasst. Außerplanmäßige Badtemperaturmessungen wurden von der weiteren Analyse ausgeschlossen. Anschließend wurde die Verteilung der Badtemperaturmesswerte anhand von mehreren Boxplots untersucht. Dabei ist ein nicht plausibler Badtemperaturwert von 872,5 °C aufgefallen, der daraufhin aus den Daten entfernt wurde. Des Weiteren wurden starke Badtemperaturänderungen von vier Öfen identifiziert, die mit Hilfe der TAE und der Betrachtung weiterer Prozessvariablen näher untersucht wurden. Es stellte sich heraus, dass Badtemperaturänderungen von ± 50 °C von einem auf den anderen Tag auftreten und auf bestimmte Ursachen zurückgeführt werden können. Lediglich die am 26.03.2022 gemessene Badtemperatur des Ofens 1058 wurde auf einen Messfehler zurückgeführt (vgl. Abbildung 5.15). Da die Unterscheidung von plausiblen und nicht plausiblen Ausreißern in den Badtemperaturdaten zeitlich aufwendig ist, wurde in dieser Arbeit auf eine ausführliche Korrektur von Ausreißern verzichtet.

Datenlücken in den Badtemperaturdaten wurden mithilfe der LOCF-Methode geschlossen. Anschließend wurden die Badtemperaturverläufe hinsichtlich ihrer Autokorrelation untersucht, aus der eine zweitägige Periodizität bzw. Saisonalität ersichtlich wurde. Die zweitägige Saisonalität wurde mit Anodenwechsel erklärt, der bei der TAE in der Regel alle zwei Tage an einem Ofen durchgeführt wird und zu einem Dreiecksverlauf in den Badtemperaturen führt. Zusätzlich wurde für jeden Badtemperaturverlauf ein Periodogramm erstellt, um ausgeprägte Frequenzen in den Badtemperaturverläufen zu identifizieren. Aus den Periodogrammen gingen ausgeprägte Frequenzspitzen bei ungefähr sieben und 30 Tagen hervor. Die Frequenzspitze bei 30 Tagen wurde in Verbindung mit einem zyklischen Muster in den Badtemperaturverläufen gebracht. Für die Frequenzspitze bei sieben Tagen konnte keine Ursache identifiziert werden.

Die Überprüfung der Stationarität der Badtemperaturverläufe wurde mittels Berechnung der Stärke der Saisonalität und mittels Anwendung des KPSS-Tests durchgeführt. Anhand der

Ergebnisse wurde festgestellt, dass die Mehrheit der untersuchten Badtemperaturverläufe stationär ist. Diese Erkenntnis deckt sich mit den Ergebnissen aus [MAW01, S. 298–299].

Für die Badtemperaturvorhersage wurde zunächst ein globales AR-Modell trainiert, das eine Badtemperaturvorhersage (Forecasting) für den nächsten Tag ermöglicht. Dabei wurde ein AR-Modell mit unterschiedlichen Ordnungen validiert. Aus der Validierung ging hervor, dass nach der zehnten Ordnung keine wesentlichen Verbesserungen in der Modellperformanz erreicht wurden. Für die weitere Analyse wurde daher ein globales AR(10)-Modell auf den kompletten Trainingsdaten trainiert und anschließend mit vier lokalen Modellen auf den Testdaten verglichen. Bei diesem Vergleich erreichte das globale AR(10)-Modell mit $4,59^\circ\text{C}$ den niedrigsten MAE. Darauf folgte mit einem MAE von $4,60^\circ\text{C}$ das lokale ETS-Modell.

Zudem wurde die Vorgehensweisen aus unserer Arbeit [GKD+18] aufgegriffen, um eine Gegenwartsvorhersage (Nowcasting) durchzuführen. Dazu wurde ein RF-Regressionsmodell mit weiteren Prozessvariablen trainiert und validiert. Während der Validierung ist aufgefallen, dass die Gegenwartsvorhersage der Badtemperaturänderung zu einem kleineren MAE führte als die Vorhersage der Badtemperatur. Mithilfe der PFI wurden Prozessvariablen identifiziert, die nicht wesentlich für die Gegenwartsvorhersage waren und für das endgültige Training auf dem kompletten Trainingsdatensatz ausgeschlossen wurden. Das endgültige RF-Regressionsmodell erreichte auf den Testdaten einen MAE von $4,53^\circ\text{C}$. Es wurde erwähnt, dass dieser Wert auf einem reduzierten Testdatensatz erzielt wurde, da Öfen mit neuer Prozesssteuerung vom Training und Test ausgeschlossen wurden. Grund ist, dass wegen der neuen Prozesssteuerung bestimmte Prozessvariablen nicht mehr in der ursprünglichen Datenbank aufgezeichnet wurden. Daher wurden für das globale AR(10)-Modell und die lokalen Modelle ebenfalls die Performanz auf dem reduzierten Testdatensatz ermittelt, um einen direkten Vergleich mit dem RF-Regressionsmodell durchführen zu können.

Tabelle 5.7 zeigt, dass das RF-Regressionsmodell mit dem bereits erwähnten MAE von $4,53^\circ\text{C}$ die beste Performanz auf den reduzierten Testdaten erzielt. Darauf folgen das globale AR(10)-Modell und das lokale ETS-Modell mit einem MAE von jeweils $4,55^\circ\text{C}$. Bei diesem Vergleich sollte berücksichtigt werden, dass das RF-Regressionsmodell eine Gegenwartsvorhersage erzeugt, während das globale AR(10)-Modell und die lokalen Modelle jeweils eine Badtemperaturvorhersage für den nächsten Tag (Tagesprognose) ermitteln. Insgesamt stellt das globale AR(10)-Modell eine Alternative zu den lokalen Modellen dar, um eine Badtemperaturvorhersage in der Aluminiumelektrolyse durchzuführen. Die praktischen Vorteile beider globalen Modellen wurden in diesem Kapitel präsentiert. Des Weiteren wurde gezeigt, dass die Modellkombination aus dem globalen AR(10)-Modell und dem globalen RF-Regressionsmodell den MAE auf $4,31^\circ\text{C}$ verbessert.

Mithilfe der Python-Pakete `treeinterpreter` [Saa21] und `waterfallcharts` [Csi17] wurde eine Möglichkeit gezeigt, um die Gegenwartsvorhersagen des RF-Regressionsmodells interpretieren

zu können, indem die Zusammensetzung der entsprechenden Gegenwartsvorhersage aus den einzelnen Merkmalen aufgezeigt wird. Zwar ermöglicht eine Interpretation der Gegenwartsvorhersagen auch eine Ursachenanalyse von Badtemperaturänderungen, diese sollte jedoch durch Praxiserfahrungen aus der Aluminiumelektrolyse bestätigt werden.



KAPITEL 6

Fazit und Ausblick

Im Rahmen dieser Arbeit wurden datengetriebene Methoden zur Vorhersage der Elektrolyttemperatur (Badtemperatur) in der Aluminiumelektrolyse untersucht und gegenübergestellt. Die Badtemperatur stellt in der industriellen Aluminiumherstellung eine wesentliche Prozessvariable dar. Mittels regelmäßiger manueller Messung der Badtemperatur können wichtige Rückschlüsse auf den Zustand eines Aluminiumelektrolyseofens gewonnen werden. Zudem dient die Badtemperatur als Maß für die Erkennung von Ofenabnormitäten [GK93, S. 215] und beeinflusst maßgeblich die Effizienz der Aluminiumproduktion [ZJX+94, S. 28]. Die korrosive Produktionsumgebung erschwert jedoch eine kontinuierliche Messung der Badtemperatur mittels einer in den Öfen angebrachte Sensorik, da der Elektrolyt diese innerhalb kurzer Zeit zerstören würde [GK93, S. 215; Dü16, S. 41].

Der Aufbau dieser Arbeit orientierte sich am Vorgehensmodell Machine Learning for Production (ML4P), das unter der Projektleitung des Fraunhofer-Instituts für Optronik, Systemtechnik und Bildauswertung (IOSB) entwickelt wurde [BP20]. In Kapitel 1 wurden der Aluminiumhersteller und Projektpartner TRIMET Aluminium SE (TRIMET) und das EFRE-Drittmittelprojekt „Thermische Flexibilisierung der Aluminiumelektrolyse (FlexTherm)“ vorgestellt. Aus diesem Drittmittelprojekt ist die initiale Motivation für die vorliegende Arbeit entstanden. In Kapitel 2 wurden das Ziel und mögliche Lösungsansätze präsentiert, um eine Badtemperaturvorhersage in der Aluminiumelektrolyse zu ermöglichen. Dazu wurde ein Machine Learning Pipeline Diagramm (ML-Pipeline-Diagramm) erläutert, um den Ist- vom Ziel-Zustand abzugrenzen. Außerdem wurde ein Literaturüberblick gegeben, der den unterschiedlichen Einsatz von datengetriebenen Methoden im Bereich der Aluminiumelektrolyse umfasst. Es wurden Arbeiten aufgezeigt, in denen bereits eine Vorhersage der Badtemperatur durchgeführt wurde.

In Kapitel 3 wurden die Grundlagen der industriellen Aluminiumherstellung vorgestellt. Die Wechselwirkungen von Prozessvariablen eines Elektrolyseofens wurden erläutert und die Auswirkungen unterschiedlicher Einflüsse auf die Badtemperatur aufgezeigt. Unter anderem wurden die Auswirkungen einer kontrollierten Stromabschaltung (Hallenschaltung) auf die Badtemperatur von vier Öfen bei der TRIMET Aluminium SE Essen (TAE) untersucht und die

Ergebnisse dargestellt. Das Kapitel 4 ging auf wesentliche Grundlagen der Zeitreihenanalyse ein, die für die in dieser Arbeit durchgeführten Analysen der Badtemperatur essentiell waren. Die numerische Beschreibung von Zeitreihen mittels Zeitreihenmerkmalen und deren Anwendungsbereiche wurden präsentiert. Die in dieser Arbeit verwendeten Vorhersagemodelle umfassen das AR-Modell und RF-Modell. Die Wahl dieser Modelle erfolgte aufgrund ihrer wettbewerbsfähigen Vorhersageperformanz sowie ihrer einfachen Handhabung und Praxisimplementierung. Des Weiteren wurden unterschiedliche Methoden aufgezeigt, um im Bereich der Zeitreihenanalyse Vorhersagemodelle zu validieren und zu testen.

Der erste Teil des Kapitels 5 zeigte die Vorverarbeitung der Badtemperaturdaten auf. Dabei wurden Schwierigkeiten in den Daten, wie Datenlücken und Datenausreißer, sowie entsprechende Lösungen aufgezeigt. Eine Analyse hinsichtlich der Stationarität ergab, dass die Mehrheit der Badtemperaturverläufe stationär ist. Dieses Ergebnis deckt sich mit den Aussagen aus [MAW01, S. 299–299]. Außerdem wurden starke Badtemperaturänderungen von Öfen analysiert. Dabei wurden Änderungen in der Badtemperatur von $\pm 50^\circ\text{C}$ festgestellt, für die entsprechende Ursachen diskutiert wurden. Diese Ursachenforschungen für starke Badtemperaturänderungen lieferten weitere Ergebnisse, die für die Entwicklung von Vorhersagemodellen eingesetzt werden konnten.

Für die Vorhersage der Elektrolyttemperatur wurden in dieser Arbeit zwei Lösungsansätze verfolgt, die im zweiten Teil des Kapitels 5 dargestellt wurden. Der erste Ansatz umfasste ein globales AR-Modell, mit dem eine Tagesprognose (Forecasting) der Badtemperatur erstellt wurde. Das globale AR-Modell wurde anschließend mit lokalen Zeitreihenmodellen verglichen. Bei diesem Vergleich erzielte das globale AR(10)-Modell eine wettbewerbsfähige Performanz auf den Testdaten. Mit einem MAE von $4,59^\circ\text{C}$ erreichte das AR(10)-Modell den kleinsten Fehler. Das lokale ETS-Modell lag mit einem MAE von $4,60^\circ\text{C}$ dicht dahinter.

Im zweiten Lösungsansatz wurde eine Gegenwartsvorhersage (Nowcasting) mit einem RF-Regressionsmodell durchgeführt. Das RF-Regressionsmodell wurde auf einem reduzierten Datensatz trainiert und getestet, da bestimmte Öfen im untersuchten Zeitraum mit einer neuen Prozesssteuerung ausgestattet waren. Das hatte den Effekt, dass einige Prozessvariablen nicht mehr in der ursprünglichen Datenbank aufgezeichnet wurden. Aus diesem Grund wurden die Öfen mit neuer Prozesssteuerung aus dem Datensatz entfernt. Somit reduzierte sich der Datensatz von 60 auf 51 Öfen, die für das Training und Testen des RF-Regressionsmodells verwendet wurden. Während der Validierung des RF-Regressionsmodells stellte sich heraus, dass die Vorhersage der Badtemperaturänderung einen kleineren MAE innerhalb der CV-BI erzeugt als wenn die Badtemperatur vorhergesagt wird. Diese Erkenntnis wurde damit begründet, dass ähnliche Merkmalsausprägungen zu einer ähnlichen Badtemperaturänderung führen. So sind die einzelnen Entscheidungsbäume besser in der Lage, den Datensatz hinsichtlich der Badtemperaturänderung in homogene Gruppen einzuteilen. Als Beispiel kann

das Merkmal Anodenwechsel genannt werden, das sich auch als prädiktives Merkmal in der CV-BI herausstellte (vgl. Kapitel 5.4.2). Wird davon ausgegangen, dass ein Anodenwechsel zur einer gleichen Badtemperaturänderung führt, dann kann das RF-Regressionsmodell diesen Zusammenhang wahrscheinlich besser lernen als wenn die Badtemperatur direkt vorhergesagt wird. Für das weitere Vorgehen wurde daher entschieden, die Änderung der Badtemperatur von einem auf den nächsten Tag vorherzusagen. Letztendlich erlangte das RF-Regressionsmodell auf dem reduzierten Testdatensatz einen MAE von 4,53 °C. Die für diesen Test verwendeten Hyperparameter wurden innerhalb einer CV-BI mit zehn Folds ermittelt.

In einem abschließenden Test wurden alle in dieser Arbeit verwendeten Vorhersagemodelle auf dem reduzierten Testdatensatz mit 51 Öfen gegenübergestellt. Das RF-Regressionsmodell erreichte in diesem Vergleich den kleinsten MAE. Mit einem MAE von jeweils 4,55 °C lagen das globale AR-Modell und das lokale ETS-Modell dicht dahinter. Den höchsten MAE erzielte das MEAN-Modell (vgl. Tabelle 5.7). Insgesamt zeigte dieser Vergleich, dass das globale AR(10)-Modell eine Alternative zu den verwendeten lokalen Zeitreihenmodellen ist. Das globale AR-Modell besitzt den praktischen Vorteil, dass es die in dieser Arbeit verwendeten lokalen Zeitreihenmodelle substituieren kann. Während bei der lokalen Vorhersage für jeden Badtemperaturverlauf ein lokales Modell trainiert wird, werden beim globalen AR-Modell zunächst alle verfügbaren Badtemperaturverläufe mit einem Lag Embedding (vgl. Kapitel 4.2.2) in eine Matrix eingebettet, mit der das globale AR-Modell anschließend trainiert wird. Dadurch können Tagesprognosen für mehrere Öfen über ein einziges globales AR-Modell erzeugt werden. Die Mittelung der Vorhersagen der beiden globalen Modelle verbesserte den MAE auf den Testdaten auf 4,31 °C.

In dieser Arbeit wurden Gegenwartsvorhersagen zum Zeitpunkt einer Badtemperaturmessung erstellt, um die Performanz des RF-Modells zu ermitteln. Das RF-Regressionsmodell ermöglicht jedoch auch die Erstellung einer Gegenwartsvorhersage zu einem beliebigen Zeitpunkt. Dies eröffnet einen weiteren Weg, um kontinuierlich auf die Badtemperatur eines Ofens zu schließen. Die zeitliche Auflösung der Vorhersagenerstellung richtet sich dabei nach der kleinsten zeitlichen Auflösung der verwendeten Prozessvariablen. Diese lag in dieser Arbeit bei fünf Minuten, da bei der TAE unter anderem der Hallenstrom und die Ofenspannung mit einer zeitlichen Auflösung von fünf Minuten aufgezeichnet wurden. Das RF-Regressionsmodell wäre damit in der Lage, eine Badtemperaturvorhersage alle fünf Minuten zu erstellen. Dennoch ist zu beachten, dass die Gegenwartsvorhersagen für einen beliebigen Zeitpunkt in fortführenden Arbeiten untersucht werden müssen. Dazu kann das RF-Modell parallel zum Betrieb bei der TAE eingesetzt und die Plausibilität der Gegenwartsvorhersagen mit den Mitarbeitenden besprochen werden.

In Kapitel 5.4.3 wurde eine Möglichkeit vorgestellt, um die Gegenwartsvorhersagen des RF-Modells interpretieren zu können. Es wurde gezeigt, dass diese Interpretation auch für eine

Ursachenanalyse von Badtemperaturänderungen eingesetzt werden kann. Ein solche Ursachenanalyse mithilfe des RF-Regressionsmodells wurde im Bereich der Aluminiumelektrolyse nach bestem Wissen noch nicht aufgezeigt. Eine Schwierigkeit besteht jedoch darin, dass die in dem Kapitel 5.4.3 präsentierten Zusammenhänge nicht unbedingt auf den realen Prozess zutreffen müssen. Weiterhin gilt es zu beachten, dass stellenweise die Gegenwartsvorhersage und die tatsächliche Badtemperatur weit auseinanderliegen. Das kann dazu führen, dass die Interpretation der Gegenwartsvorhersage und die Ursachenforschung zu diesen Zeitpunkten nicht besonders aussagekräftig sind. Daher sind auch in diesem Bereich weitere Arbeiten erforderlich, um den praktischen Nutzen einer Ursachenanalyse von Badtemperaturänderungen im Praxisbetrieb zu evaluieren. Die Zerlegung der Gegenwartsvorhersagen und die anschließende Darstellung über einen Waterfall-Chart können dennoch als Hilfsmittel eingesetzt werden, um einen ersten Hinweis auf Ursachen für bestimmte Badtemperaturänderungen zu erhalten.

Folgend wird ein interaktives Dashboard vorgestellt, das in [Abd23] entwickelt wurde und einen praxistauglichen Einsatz der Badtemperaturvorhersagen in der Aluminiumelektrolyse ermöglichen soll. In Abbildung 6.1 ist die Übersichtsseite des Dashboards für das RF-Regressionsmodell zu sehen.

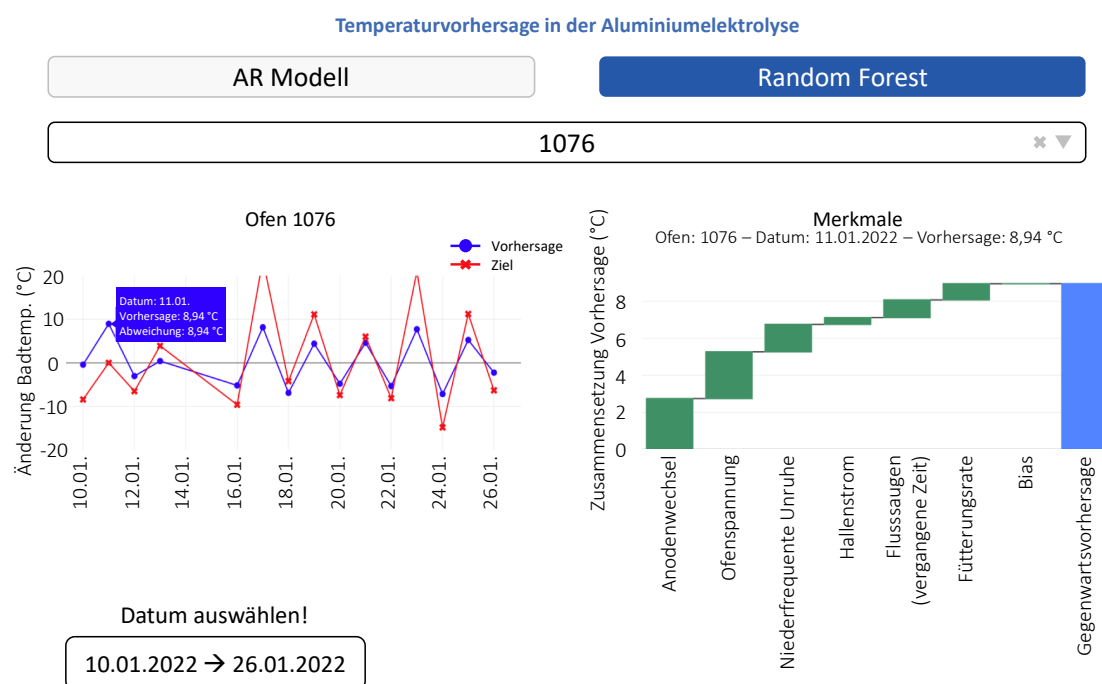


Abbildung 6.1: Dashboard für die Analyse der Gegenwartsvorhersagen des RF-Regressionsmodells. Abgeändert nach [Abd23, S. 51].

Auf der linken Seite der Abbildung 6.1 sind die Gegenwartsvorhersagen in blau und die aus den gemessenen Badtemperaturen berechneten Änderungen in rot für den Ofen 1076 zeit-

lich aufgetragen. Weitere Öfen lassen sich über das Dropdown-Menü auswählen. Für jeden Zeitpunkt einer Badtemperaturmessung wird zudem die Abweichung zwischen der Vorhersage und der gemessenen Badtemperatur angezeigt. Weiter unten lässt sich der gewünschte Zeitraum auswählen, der angezeigt werden soll. Auf der rechten Seite des Dashboards ist ein Waterfall-Chart dargestellt, der die Zerlegung der Gegenwartsvorhersage für den 11.01.2022 aufzeigt und somit eine interaktive Ursachenanalyse für Badtemperaturänderungen ermöglicht. Der Waterfall-Chart lässt sich für jede erstellte Gegenwartsvorhersage anzeigen. Im oberen Bereich des Dashboards kann zudem die Übersichtsseite für die Tagesprognosen (Forecasting) des globalen AR-Modells ausgewählt werden. Insgesamt soll das Dashboard einen benutzerfreundlichen Umgang mit den erstellten Badtemperaturvorhersagen im Produktivbetrieb der Aluminiumelektrolyse ermöglichen. Für die weitere Vorgehensweise gilt es, innerhalb der dritten Phase des ML4P-Vorgehensmodells in einen Austausch mit den Prozessexperten/-innen der TAE zu treten, um eine sinnvolle Integration des Dashboards in den Produktionsbetrieb zu erörtern.

In dieser Arbeit wurden die ersten drei der sechs Phasen des ML4P-Vorgehensmodells [BP20] behandelt, um eine Badtemperaturvorhersage in der Aluminiumelektrolyse zu ermöglichen. In weiterführenden Arbeiten sollten die letzten drei Phasen des Vorgehensmodells abgearbeitet werden. Das beinhaltet unter anderem die Absprache mit den Mitarbeitenden der TAE, um die erstellten Vorhersagemodelle in den Produktivbetrieb zu überführen. Eine Schwierigkeit, die es dabei zu bewältigen gilt, ist die Änderung der Messfrequenz der Badtemperatur. Die Messungen finden bei der TAE für die Öfen aus Halle 1 mittlerweile nur noch alle zwei Tage statt. Eine mögliche Lösung hierfür stellen die Badtemperaturvorhersagen des globalen AR(10)-Modells dar. Diese können für die Tage eingesetzt werden, an denen keine Badtemperaturmessung stattfindet. Das AR(10)-Modell erstellt somit eine Tagesprognose basierend auf gemessenen und vorhergesagten Badtemperaturen. Die Vorhersagen können anschließend für die Ofenregelung eingesetzt werden. Da das RF-Regressionsmodell den letzten gemessenen Badtemperaturwert als Merkmal nutzt, könnte dieses auch auf die Tagesprognose des globalen AR(10)-Modells zurückgreifen. Diese Vorgehensweise gilt es aber mit weiteren Arbeiten zu validieren. Eine Verbesserung der Vorhersageperformanz des RF-Regressionsmodells könnte unter Berücksichtigung weiterer Merkmale erreicht werden. In dieser Arbeit wurde beispielsweise der potenzielle Einfluss der an den Öfen installierten Wärmetauscher (vgl. Kapitel 1) auf die Badtemperatur nicht berücksichtigt. Aus diesem Grund sollte in einer weiteren Analyse der Einfluss der Wärmetauscher untersucht und geeignete Merkmale für das Training des RF-Regressionsmodells ausgewählt werden.

Die Liquidustemperatur spielt in der Aluminiumelektrolyse ebenfalls eine wichtige Rolle. Sie stellt die Mindesttemperatur dar, bei der der Prozess betrieben werden muss, ohne dass der Elektrolyt zu erstarren beginnt [Hau16c, S. 804]. Daher wird die Liquidustemperatur in regel-

mäßigen Abständen bei der TAE ermittelt. Es ist möglich, dass die in dieser Arbeit entwickelten Lösungen auch für eine Vorhersage der Liquidustemperatur eingesetzt werden könnten.

In der vorliegenden Arbeit wurden zwei Lösungen präsentiert, die eine Badtemperaturvorhersage in der Aluminiumelektrolyse ermöglichen. Der Ziel-Zustand, der in Kapitel 2 beschrieben und vom ML-Pipeline-Diagramm in Abbildung 2.1 dargestellt wird, wurde in dieser Arbeit bis auf die Integration der Vorhersagemodelle in die bestehende Ofenregelung erreicht. Vor allem in Zeiten der Energiewende sind neuartige Konzepte hinsichtlich der Analyse von Ofenprozessdaten notwendig, da die Aluminiumelektrolyse nicht mehr mit einer konstanten sondern variablen Energiezufuhr betrieben wird [Dü16, S. 163–164]. Vor diesem Hintergrund wurden die in dieser Arbeit aufgezeigten Lösungen entwickelt.

Für eine Tagesprognose der Badtemperatur in der Aluminiumelektrolyse wird der Einsatz eines globalen linearen AR-Modells empfohlen, da es in der Lage ist, die in dieser Arbeit verwendeten klassischen lokalen Vorhersagemodelle zu substituieren. Ein globales lineares AR-Modell sollte daher in zukünftigen Arbeiten für eine Tagesprognose zusätzlich berücksichtigt werden, da es gegenüber den verwendeten lokalen Modellen praktische Vorteile mit sich bringt. Liegt der Fokus auf einer Gegenwartsvorhersage, so bietet sich ein RF-Regressionsmodell an. Hierbei sollte die Badtemperaturänderung anstatt der Badtemperatur vorhergesagt werden, da so eine bessere Vorhersageperformanz erreicht werden konnte. Darüber hinaus können die Gegenwartsvorhersagen des RF-Regressionsmodells durch die in dieser Arbeit vorgestellten Analysemethoden interpretiert werden.

Die Arbeitsschritte, die notwendig sind, um das abschließende Ziel einer Aluminiumelektrolyse mit flexiblem Energieeintrag unter dem Einsatz datengetriebener Methoden zu erreichen, wurden in diesem Kapitel aufgezeigt. Nun gilt es die in dieser Arbeit entwickelten Methoden in den Produktivbetrieb zu überführen.

Anhang

A Abbildungen

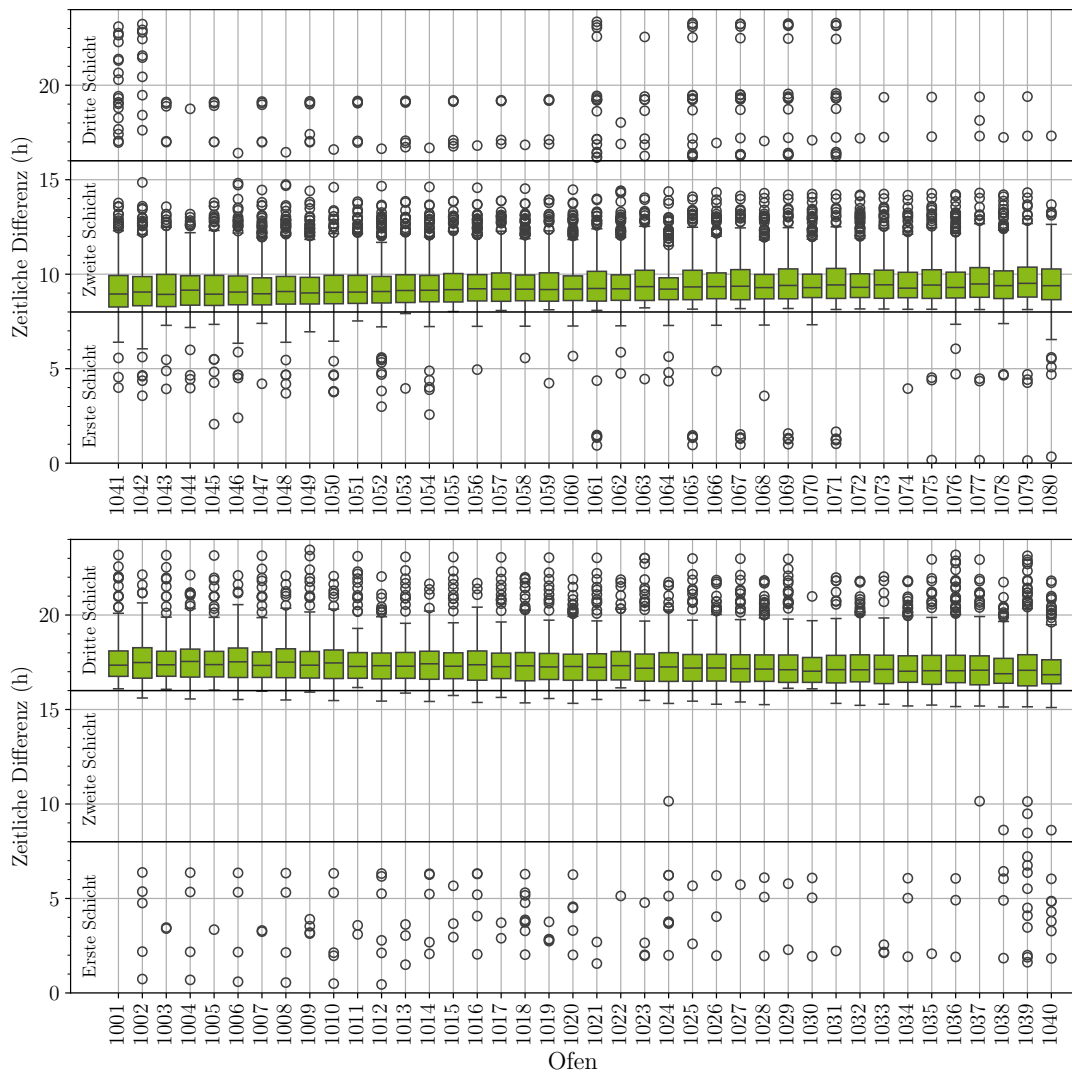
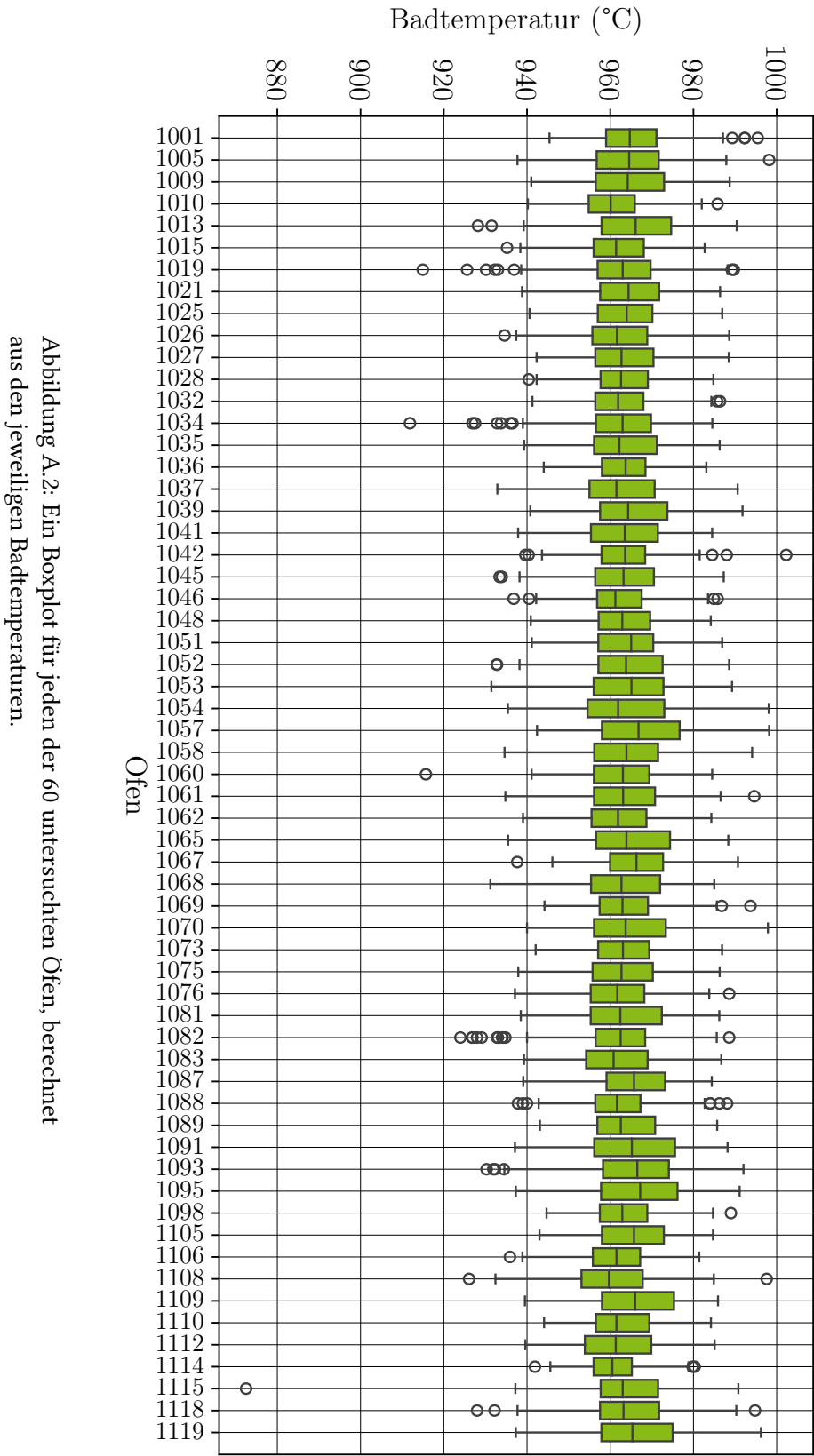


Abbildung A.1: Die Boxplots zeigen die Verteilung der zeitlichen Differenzen für die Öfen der zweiten und dritten Schicht. Die Differenzen wurden jeweils zwischen dem Messzeitpunkt (UTC) der Badtemperatur und Beginn (UTC) der ersten Schicht berechnet.



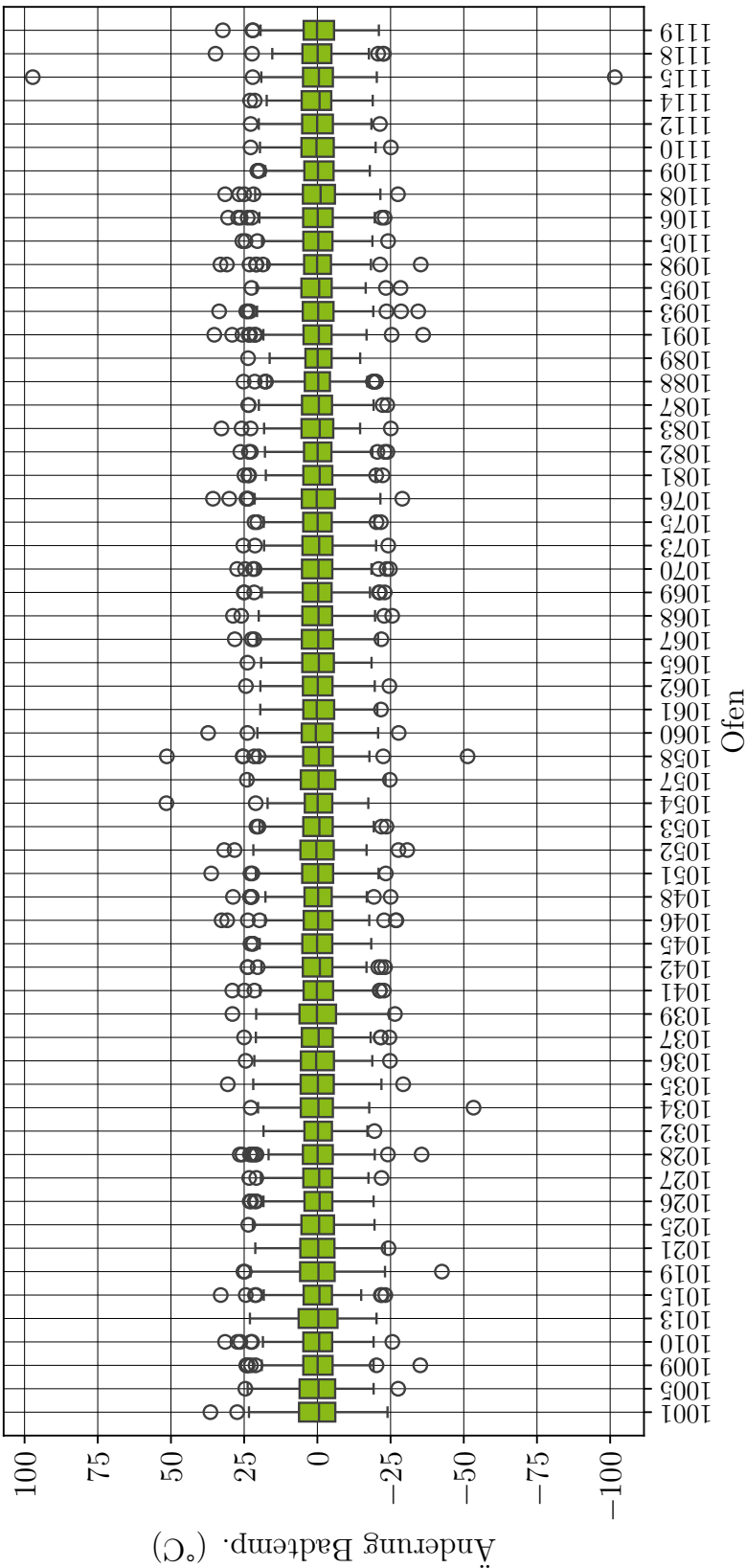


Abbildung A.3: Ein Boxplot für jeden der 60 untersuchten Öfen, berechnet aus den jeweiligen Badtemperaturänderungen.

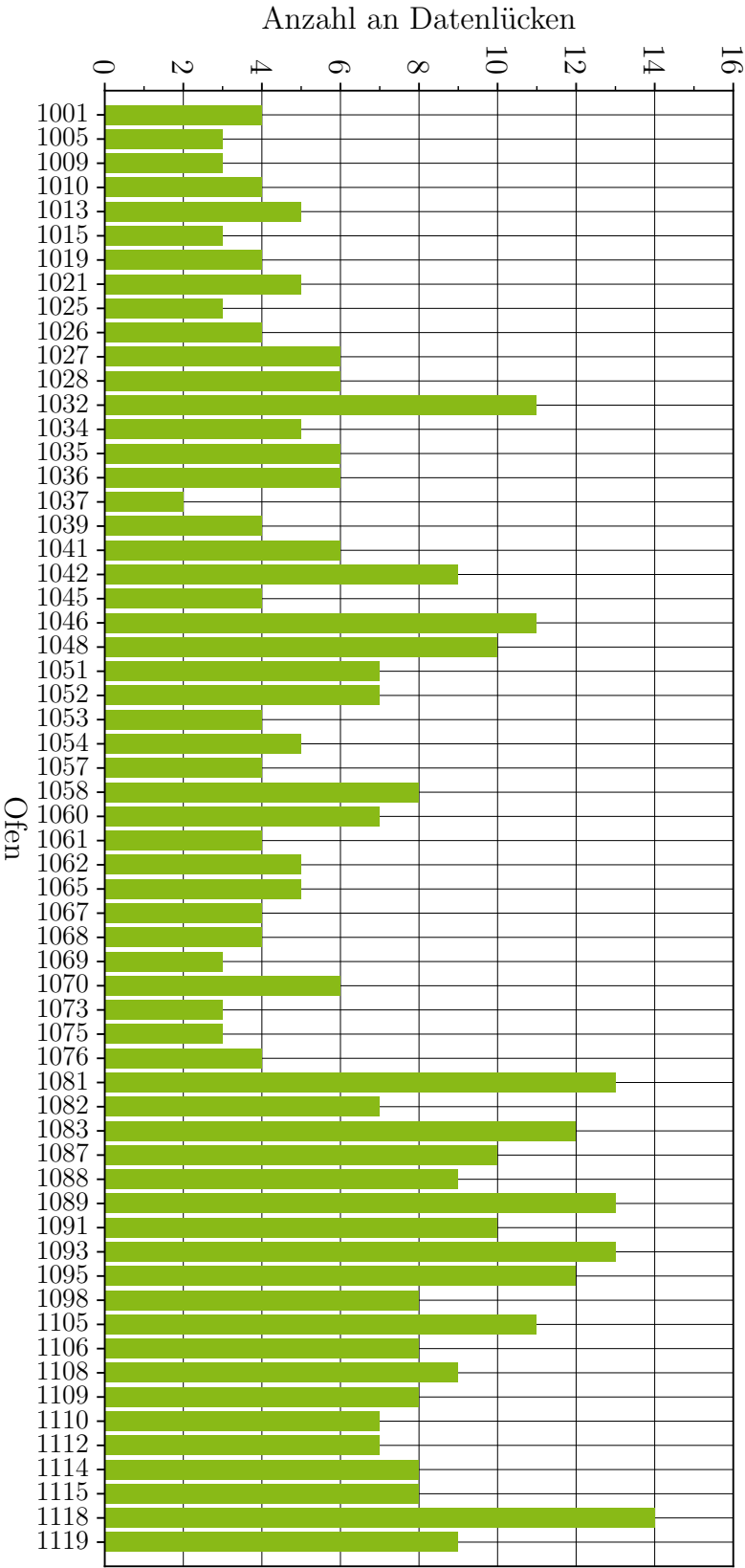


Abbildung A.4: Anzahl an Datenlücken in den Badtemperaturdaten für jeden der 60 untersuchten Öfen aus Halle 1 im Jahr 2022.

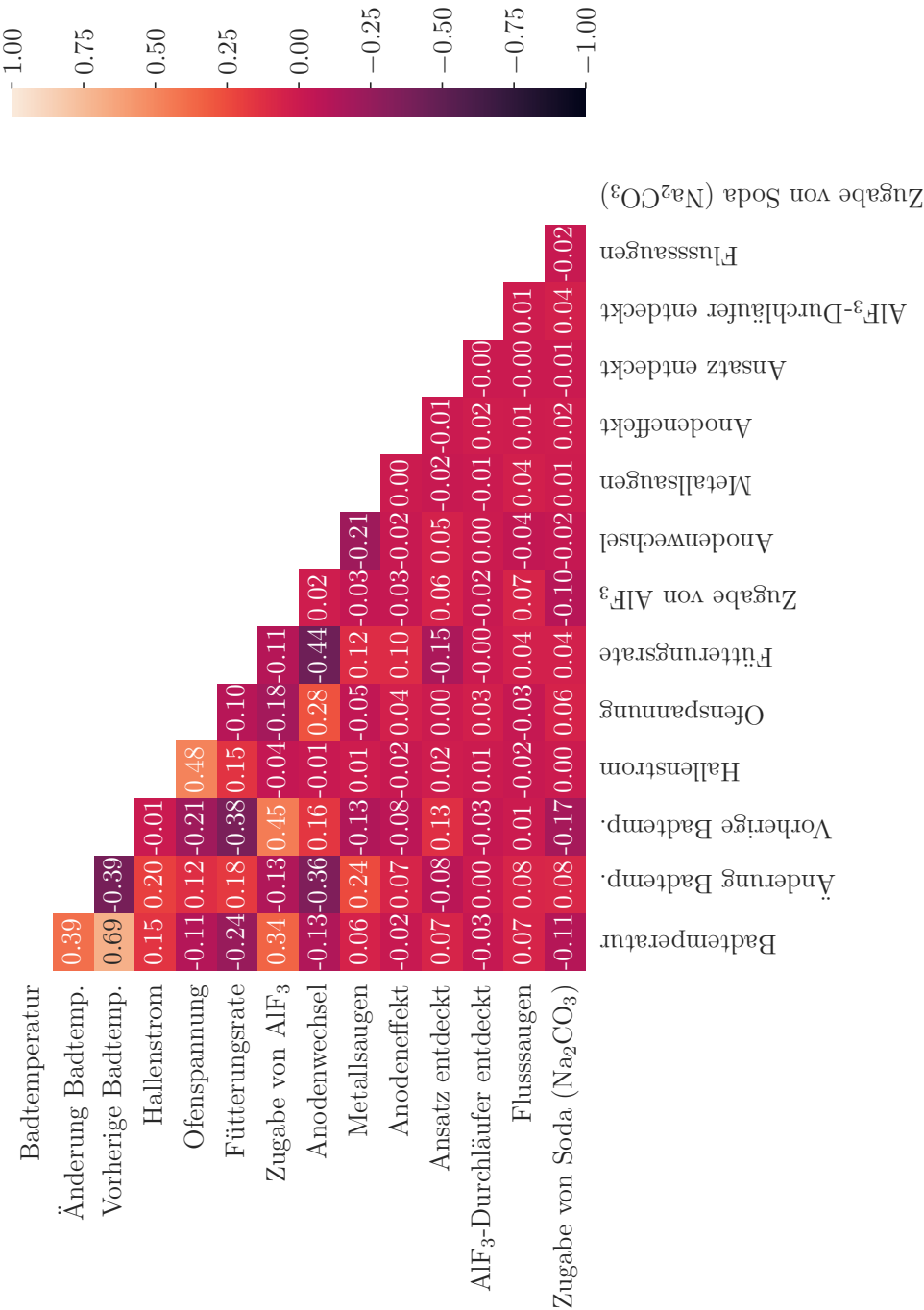


Abbildung A.5: Korrelationen nach Pearson zwischen den berechneten Zeitreihenmerkmalen in den Trainingsdaten. Weiterhin sind die Korrelationen der Badtemperatur und der Badtemperaturänderung zu sehen.

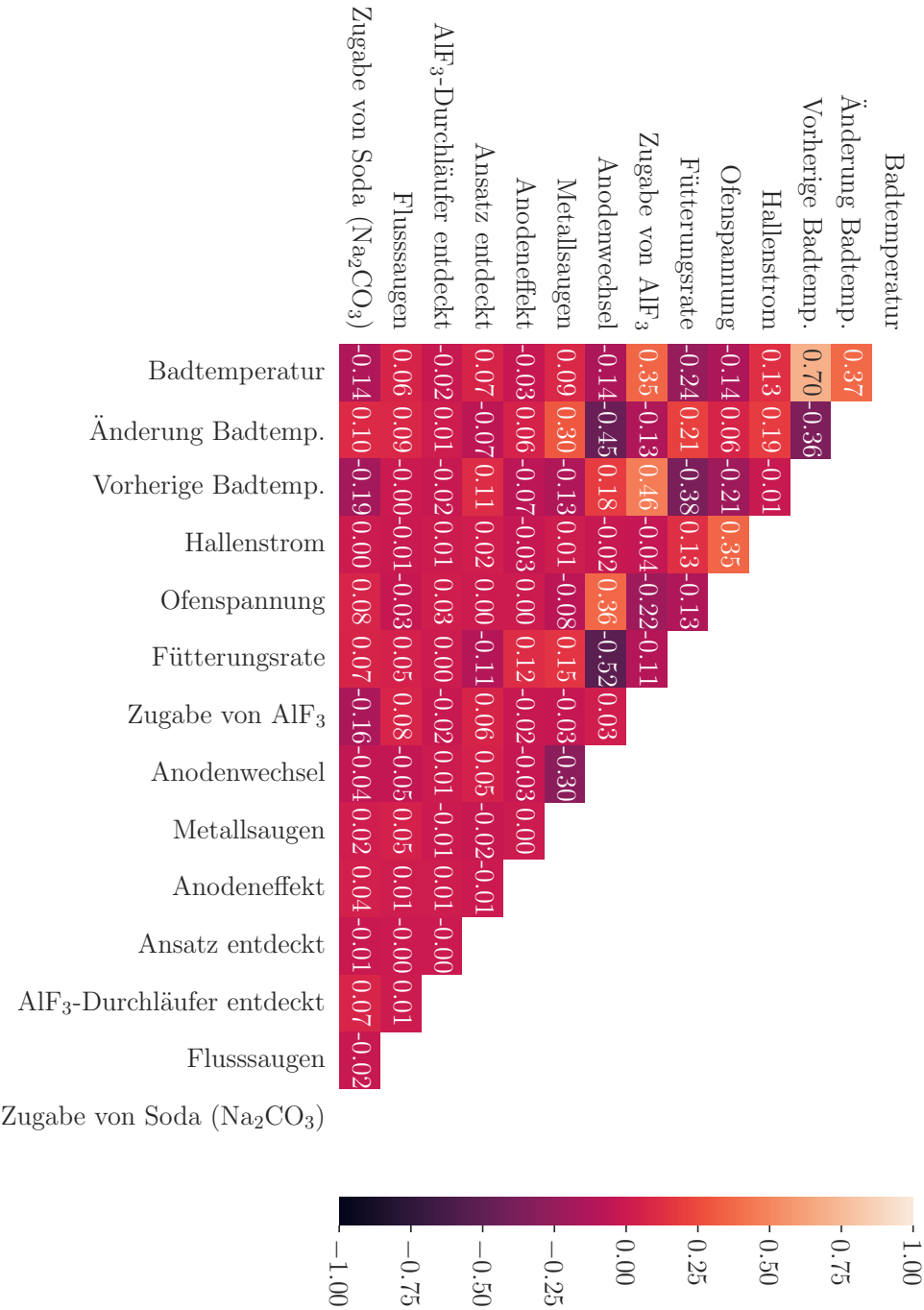


Abbildung A.6: Korrelationen nach Spearman zwischen den berechneten Zeitreihenmerkmalen in den Trainingsdaten. Weiterhin sind die Korrelationen der Badtemperatur und der Badtemperaturänderung zu sehen.

Literatur

- [Abd23] Abdullateef Abdullateef. „Entwicklung eines Dashboards für den praxistauglichen Einsatz von Temperaturvorhersagen in der Aluminiumindustrie“. Bachelor-Thesis. Bergische Universität Wuppertal, Apr. 2023.
- [AGW97] Y. Amit, D. Geman und K. Wilder. „Joint induction of shape features and tree classifiers“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.11 (1997), S. 1300–1305. issn: 0162-8828. doi: 10.1109/34.632990.
- [BGR10] Marta Banbura, Domenico Giannone und Lucrezia Reichlin. „Nowcasting“. In: *SSRN Electronic Journal* (Nov. 2010). ECB Working Paper No. 1275. issn: 1725-2806. doi: 10.2139/ssrn.1717887.
- [BBS20] Kasun Bandara, Christoph Bergmeir und Slawek Smyl. „Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach“. In: *Expert Systems with Applications* 140 (Feb. 2020), S. 112896. issn: 0957-4174. doi: 10.1016/j.eswa.2019.112896.
- [BHK18] C. Bergmeir, R. J. Hyndman und B. Koo. „A Note on the Validity of Cross-Validation for Evaluating Autoregressive Time Series Prediction“. In: *Computational Statistics & Data Analysis* 120 (Apr. 2018), S. 70–83. doi: 10.1016/j.csda.2017.11.003.
- [BP20] Jürgen Beyerer und Julius Pfrommer. *ML4P-Vorgehensmodell: Machine learning for production*. Techn. Ber. Abgerufen am 23.12.2022 um 11:12 Uhr. Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung, 2020. URL: https://www.iosb.fraunhofer.de/content/dam/iosb/iosbtest/documents/projekte/ml4p/ML4P_whitepaper.pdf.
- [BLS+24] Vadim Borisov, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk und Gjergji Kasneci. „Deep Neural Networks and Tabular Data: A Survey“. In: *IEEE Transactions on Neural Networks and Learning Systems* 35.6 (Juni 2024), S. 7499–7519. issn: 2162-2388. doi: 10.1109/tnnls.2022.3229161.
- [Bre96] Leo Breiman. „Bagging predictors“. In: *Machine Learning* 24.2 (Aug. 1996), S. 123–140. issn: 1573-0565. doi: 10.1007/bf00058655.
- [Bre01a] Leo Breiman. „Random Forests“. In: *Machine Learning* 45.1 (2001), S. 5–32. issn: 0885-6125. doi: 10.1023/a:1010933404324.
- [Bre01b] Leo Breiman. „Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)“. In: *Statistical Science* 16.3 (Aug. 2001). doi: 10.1214/ss/1009213726.
- [Bun] Bundesministerium für Wirtschaft und Klimaschutz. *Ein Stromnetz für die Energiewende*. Abgerufen am 05.07.2024 um 19:53 Uhr. Berlin, Deutschland. URL: <https://www.bmwk.de/Redaktion/DE/Dossier/netze-und-netzausbau.html>.
- [Dig] C-V-Therm Combined Sensor For Primary Aluminium. Abgerufen am 19.06.2024 um 16:39 Uhr. Heraeus Electro-Nite. Feb. 2024. URL: <https://www.heraeus-electro-nite.com/dam/jcr:d03a171d-032a-47a8-96a0-fe7a9aff4bab/c-v-therm.pdf>.

- [CTM20] Vitor Cerqueira, Luis Torgo und Igor Mozetič. „Evaluating time series forecasting models: an empirical study on performance estimation methods“. In: *Machine Learning* 109.11 (Okt. 2020), S. 1997–2028. DOI: 10.1007/s10994-020-05910-7.
- [CSO16] Patrizia R. S. Chermont, Fábio M. Soares und Roberto C. L. de Oliveira. „Simulations on the Bath Chemistry Variables Using Neural Networks“. In: *Light Metals 2016*. Springer International Publishing, 2016, S. 607–612. DOI: 10.1007/978-3-319-48251-4_102.
- [CT65] James W. Cooley und John W. Tukey. „An Algorithm for the Machine Calculation of Complex Fourier Series“. In: *Mathematics of Computation* 19.90 (1965), S. 297–301. ISSN: 1088-6842. DOI: 10.1090/s0025-5718-1965-0178586-1.
- [CCV21] Pedro Costa, Vitor Cerqueira und João Vinagre. „AutoFITS: Automatic Feature Engineering for Irregular Time Series“. Dez. 2021. arXiv: 2112.14806 [cs.LG].
- [CM09] Paul S. P. Cowpertwait und Andrew V. Metcalfe. *Introductory Time Series with R*. Springer New York, 2009. DOI: 10.1007/978-0-387-88698-5.
- [Csi17] Christopher Paul Csiszar. *waterfallcharts*. PyPi Paket Version 3.8. Nov. 2017. URL: <https://pypi.org/project/waterfallcharts>.
- [DDPR16] N. Depree, R. Düssel, P. Patel und T. Reek. „The ‘Virtual Battery’ — Operating an Aluminium Smelter with Flexible Energy Input“. In: *Light Metals 2016*. Springer International Publishing, 2016, S. 571–576. DOI: 10.1007/978-3-319-48251-4_96.
- [DMB19] Roman Düssel, Albert Mulder und Louis Bugnion. „Transformation of a Potline from Conventional to a Full Flexible Production Unit“. In: *Light Metals 2019*. Hrsg. von Corleen Chesonis. Cham: Springer International Publishing, 2019, S. 533–541. ISBN: 978-3-030-05864-7. DOI: 10.1007/978-3-030-05864-7_68.
- [Dü16] Roman Düssel. „Entwicklung eines Regelungskonzepts für Aluminium-Elektrolysezellen unter Berücksichtigung einer variablen Stromstärke und eines regelbaren Wärmeverlusts“. Diss. Deutschland: Bergische Universität Wuppertal, 2016.
- [FHK+16] Ludwig Fahrmeir, Christian Heumann, Rita Künstler, Iris Pigeot und Gerhard Tutz. *Statistik*. Springer Berlin Heidelberg, 2016. DOI: 10.1007/978-3-662-50372-0.
- [Fiba] *FiberLab® Online bath analysis in cryolite*. Abgerufen am 19.06.2024 um 16:49 Uhr. Heraeus Electro-Nite. Jan. 2023. URL: <https://www.heraeus-electro-nite.com/dam/jcr:6880eb6f-2b6a-4c98-ad4e-d2f10edd905d/fiberlab.pdf>.
- [Fibb] *FiberLab®*. Abgerufen am 19.06.2024 um 17:05 Uhr. Heraeus Electro-Nite. URL: <https://www.heraeus-electro-nite.com/en/non-ferrous/aluminium/>.
- [Fro18] Jörg Frochte. *Maschinelles Lernen: Grundlagen und Algorithmen in Python*. München: Carl Hanser Verlag, 2018. ISBN: 978-3-446-45291-6.
- [FK00a] F. Frost und V. Karri. „Identifying Significant Parameters for Hall-Heroult Process Using General Regression Neural Networks“. In: *Intelligent Problem Solving. Methodologies and Approaches*. Springer Berlin Heidelberg, 2000, S. 73–78. ISBN: 9783540450498. DOI: 10.1007/3-540-45049-1_9.
- [FK00b] F. Frost und V. Karri. „Productivity Improvements through Prediction of Electrolyte Temperature in Aluminium Reduction Cell Using BP Neural Network“. In: *PRICAI 2000 Topics in Artificial Intelligence*. Springer Berlin Heidelberg, 2000, S. 490–499. ISBN: 9783540445333. DOI: 10.1007/3-540-44533-1_50.

- [Ful18] Ben D. Fulcher. „Feature-Based Time-Series Analysis“. In: *Feature engineering for machine learning and data analytics*. Boca Raton, FL: CRC Press/Taylor & Francis Group, 2018, S. 87–116. ISBN: 9781315181080. URL: <https://www.taylorfrancis.com/chapters/edit/10.1201/9781315181080-4/feature-based-time-series-analysis-ben-fulcher?context=ubx&refId=9f3349b9-95c3-4b3a-9ec7-12953a9af75f>.
- [GGC+] Federico Garza, Kin Gutierrez, Cristian Challu, Jose Morales, Ricardo Olivares und Max Mergenthaler. *tsfeatures*. Python package. URL: <https://github.com/FedericoGarza/tsfeatures>.
- [GKD+18] Niclas Grabowski, Ron Kremser, Roman Düssel, Albert Mulder und Dietmar Tutsch. „Using Random Forest Regression for Predicting and Analysing Reduction Cell Behaviour“. In: *12th Australasian Aluminium Smelting Technology Conference*. Queenstown, New Zealand, 2018.
- [GKD+23] Niclas Grabowski, Ron Kremser, Roman Düssel, Albert Mulder und Dietmar Tutsch. „Information Extraction from Industrial Sensor Data Using Time Series Meta-Features“. In: *Applied Sciences* 13.12 (Juni 2023), S. 7065. DOI: 10.3390/app13127065.
- [GK93] Kai Grjotheim und Halvor Kvande. *Introduction to Aluminium Electrolysis: Understanding the Hall-Héroult Process*. 2. Aufl. Düsseldorf, Deutschland: Alu Media GmbH, 1993. ISBN: 978-3-942486-06-4.
- [GB19] Amit Gupta und Biswajit Basu. „Sustainable Primary Aluminium Production: Technology Status and Future Opportunities“. In: *Transactions of the Indian Institute of Metals* 72.8 (Mai 2019), S. 2135–2150. DOI: 10.1007/s12666-019-01699-9.
- [GSWSK12] Vanderlei Gusberti, Dagoberte S. Severo, Barry J. Welch und Maria Skyllas-Kazacos. „Modeling the Mass and Energy Balance of Different Aluminium Smelting Cell Technologies“. In: *Light Metals 2012*. Springer International Publishing, 2012, S. 929–934. DOI: 10.1007/978-3-319-48179-1_161.
- [HK17] Andreas Handl und Torben Kuhlenkasper. *Multivariate Analysemethoden*. 3. Aufl. Springer Berlin Heidelberg, 2017. ISBN: 9783662547540. DOI: 10.1007/978-3-662-54754-0.
- [HTF09] Trevor Hastie, Robert Tibshirani und Jerome Friedman. *The Elements of Statistical Learning*. 2. Aufl. Springer New York, 2009. ISBN: 978-0-387-84858-7. DOI: 10.1007/978-0-387-84858-7.
- [HBGH03] Elin Haugland, Håvard Børset, Håvard Gikling und Helge Høie. „Effects of Ambient Temperature and Ventilation on Shell Temperature, Heat Balance and Side Ledge of an Alumina Reduction Cell“. In: *Light Metals 2003*. Hrsg. von Paul N. Crepeau. TMS (The Minerals, Metals & Materials Society), 2003, S. 269–276.
- [Hau16a] Warren E. Haupin. „Principles of Aluminum Electrolysis“. In: *Essential Readings in Light Metals*. Springer International Publishing, 2016, S. 3–11. DOI: 10.1007/978-3-319-48156-2_1.
- [Hau16b] Warren Haupin. „Interpreting the Components of Cell Voltage“. In: *Essential Readings in Light Metals*. Springer International Publishing, 2016, S. 153–159. DOI: 10.1007/978-3-319-48156-2_21.
- [Hau16c] Warren Haupin. „The Liquidus Enigma“. In: *Essential Readings in Light Metals*. Springer International Publishing, 2016, S. 804–807. ISBN: 9783319481562. DOI: 10.1007/978-3-319-48156-2_119.
- [HV12] M. Horvath und E. Vircikova. „Data Mining For Quality Control of Primary Aluminium Production Process“. In: *Management and Production Engineering Review*. Bd. 3. 4. Production Engineering Committee of the Polish Academy of Sciences, Polish Association for Production Management, 2012, S. 47–53.

- [HG20] Jeremy Howard und Sylvain Gugger. *Deep Learning for Coders with fastai and PyTorch. AI Applications Without a PhD*. 1. Aufl. O'Reilly, Juli 2020. ISBN: 9781492045526.
- [HA21] Rob J. Hyndman und George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Mai 2021. 442 S. ISBN: 0987507133.
- [HK08] Rob J. Hyndman und Yeasmin Khandakar. „Automatic Time Series Forecasting: The forecast Package for R“. In: *Journal of Statistical Software* 27.3 (2008). ISSN: 1548-7660. DOI: 10.18637/jss.v027.i03.
- [HK06] Rob J. Hyndman und Anne B. Koehler. „Another look at measures of forecast accuracy“. In: *International Journal of Forecasting* 22.4 (2006), S. 679–688. ISSN: 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2006.03.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0169207006000239>.
- [HWL15] Rob J. Hyndman, Earo Wang und Nikolay Laptev. „Large-Scale Unusual Time Series Detection“. In: *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, Nov. 2015. DOI: 10.1109/icdmw.2015.104.
- [Hyn18] Rob Hyndman. *Mcomp: Data from the M-Competitions*. R package version 2.8. 2018. URL: <https://CRAN.R-project.org/package=Mcomp>.
- [JGJ21] Nils Janssen, Hendrik Gesell und Uwe Janoske. „Adaption of the aluminium electrolysis to volatile power supply: Development of a predictive model to investigate the thermal behavior of a cell“. In: *NAFEMS* (2021), S. 1–18.
- [KHL20] Yanfei Kang, Rob J. Hyndman und Feng Li. „GRATIS: GeneRAting Time Series with diverse and controllable characteristics“. In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 13.4 (Mai 2020), S. 354–376. DOI: 10.1002/sam.11461. URL: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sam.11461>.
- [KHSM17] Yanfei Kang, Rob J. Hyndman und Kate Smith-Miles. „Visualising forecasting algorithm performance using time series instance spaces“. In: *International Journal of Forecasting* 33.2 (2017), S. 345–358. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2016.09.004. URL: <https://www.sciencedirect.com/science/article/pii/S0169207016301030>.
- [KND15] John D. Kelleher, Brian Mac Namee und Aoife D'Arcy. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. Cambridge, Massachusetts: MIT Press, 2015. ISBN: 978-0-262-02944-5.
- [Koh95] Ron Kohavi. „A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection“. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. Bd. 2. IJCAI'95. 1995, 1137–1143. ISBN: 1558603638.
- [KGD+21] R. Kremser, N. Grabowski, R. Düssel, A. Mulder und D. Tutsch. „Comparison of different spike detection methods in Hall-Héroult cells“. In: *Automation 2021*. VDI Verlag, 2021, S. 763–XVIII. DOI: 10.51202/9783181023921-763.
- [KGD+20] Ron Kremser, Niclas Grabowski, Roman Düssel, Albert Mulder und Dietmar Tutsch. „Anode Effect Prediction in Hall-Héroult Cells Using Time Series Characteristics“. In: *Applied Sciences* 10.24 (Dez. 2020). ISSN: 2076-3417. DOI: 10.3390/app10249050.
- [Kva15] Halvor Kvande. *Heat Balance Fundamentals*. Präsentation. TMS 2015 Industrial Aluminum Electrolysis Course (TRIMET Aluminium SE, Essen, Germany). Okt. 2015.
- [KMSS13] Halvor Kvande, Bjorn P. Moxnes, Jam Skaar und Per A. Solli. „Pseudo Resistance Curves for Aluminium Cell Control - Alumina Dissolution and Cell Dynamics“. In: *Essential Readings in Light Metals*. John Wiley & Sons, Inc., Apr. 2013, S. 760–766. DOI: 10.1002/9781118647851.ch113.

- [KPSS92] Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt und Yongcheol Shin. „Testing the null hypothesis of stationarity against the alternative of a unit root“. In: *Journal of Econometrics* 54.1–3 (Okt. 1992), S. 159–178. issn: 0304-4076. doi: 10.1016/0304-4076(92)90104-y.
- [LSS+17] Flávia A. N. de Lima, Alan M. F. de Souza, Fábio M. Soares, Diego Lisboa Cardoso und Roberto C. L. de Oliveira. „Clustering Aluminum Smelting Potlines Using Fuzzy C-Means and K-Means Algorithms“. In: *Light Metals 2017*. Springer International Publishing, 2017, S. 589–597. isbn: 9783319515410. doi: 10.1007/978-3-319-51541-0_73.
- [Lou14] Gilles Louppe. „Understanding Random Forests: From Theory to Practice“. Juli 2014. arXiv: 1407.7502 [stat.ML].
- [LSK+19] Carl H. Lubba, Sarab S. Sethi, Philip Knaute, Simon R. Schultz, Ben D. Fulcher und Nick S. Jones. „catch22: CAnonical Time-series CHaracteristics“. In: *Data Mining and Knowledge Discovery* 33.6 (Aug. 2019), S. 1821–1852. doi: 10.1007/s10618-019-00647-x.
- [Mad92] D. J. Madsen. „Temperature Measurement and Control in Reduction Cells“. In: *Light Metals: Proceedings of Sessions, TMS Annual Meeting*. The Minerals, Metals & Materials Society, 1992, S. 453–456.
- [Maj11] Nazatul Aini Abd Majid. „Cascade Fault Detection and Diagnosis for the Aluminium Smelting Process using Multivariate Statistical Techniques“. Diss. New Zealand: University of Auckland, 2011.
- [MTC+11] Nazatul Aini Abd Majid, Mark P. Taylor, John J. J. Chen, Marco A. Stam, Albert Mulder und Brent R. Young. „Aluminium process fault detection by Multiway Principal Component Analysis“. In: *Control Engineering Practice* 19.4 (2011), S. 367–379. issn: 0967-0661. doi: 10.1016/j.conengprac.2010.12.005. url: <https://www.sciencedirect.com/science/article/pii/S0967066110002650>.
- [MYTC12] Nazatul Aini Abd Majid, Brent R. Young, Mark P. Taylor und John J. J. Chen. „K-means clustering pre-analysis for fault diagnosis in an aluminium smelting process“. In: *2012 4th Conference on Data Mining and Optimization (DMO)*. IEEE, Sep. 2012. doi: 10.1109/dmo.2012.6329796.
- [MH00] Spyros Makridakis und Michèle Hibon. „The M3-Competition: results, conclusions and implications“. In: *International Journal of Forecasting* 16.4 (2000). The M3- Competition, S. 451–476. issn: 0169-2070. doi: 10.1016/S0169-2070(00)00057-1. url: <http://www.sciencedirect.com/science/article/pii/S0169207000000571>.
- [Mar17] A. Martel. „Anode Spike Detection Using Advanced Analytics and Data Analysis“. In: *Travaux 46, Proceedings of 35th International ICSOBA Conference, Hamburg, Germany, 2 – 5 October, 2017*. Okt. 2017, S. 815–822.
- [Mar18] Arthur Martel. „Spike Detection Using Advanced Analytics and Data Analysis“. In: *Light Metals 2018*. 2018, S. 485–490. doi: 10.1007/978-3-319-72284-9_64.
- [MAW01] F. J. S. McFadden, P. C. Austin und B. J. Welch. „Time-Series Analysis and Modelling of Electrolyte temperature in aluminium reduction cells“. In: *Automation in Mining, Mineral and Metal Processing*. 2001, S. 297–302.
- [MMH21] Pablo Montero-Manso und Rob J. Hyndman. „Principles and algorithms for forecasting groups of time series: Locality and globality“. In: *International Journal of Forecasting* 37.4 (Okt. 2021), S. 1632–1653. issn: 0169-2070. doi: 10.1016/j.ijforecast.2021.03.004.
- [MBB17] Steffen Moritz und Thomas Bartz-Beielstein. „imputeTS: Time Series Missing Value Imputation in R“. In: *The R Journal* 9.1 (2017), S. 207–218. doi: 10.32614/rj-2017-009.

- [NAM01] Alex Nanopoulos, Rob Alcock und Yannis Manolopoulos. „Feature-Based Classification of Time-Series Data“. In: *Information Processing and Technology*. USA: Nova Science Publishers, Inc., 2001, S. 49–61. ISBN: 1590331168.
- [Nel20] Oliver Nelles. *Nonlinear System Identification*. 2. Aufl. Springer International Publishing, 2020. ISBN: 978-3-030-47439-3. DOI: 10.1007/978-3-030-47439-3.
- [OWHW23] Mitchell O’Hara-Wild, Rob Hyndman und Earo Wang. *fable: Forecasting Models for Tidy Time Series*. R Paket Version 0.3.3. 2023. URL: <https://CRAN.R-project.org/package=fable>.
- [PTCH18] Terence Parr, Kerem Turgutlu, Christopher Csiszar und Jeremy Howard. *Beware Default Random Forest Importances*. Abgerufen am 05.08.2024 um 17:28 Uhr. explained.ai. März 2018. URL: <https://explained.ai/rf-importance/>.
- [PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot und E. Duchesnay. „Scikit-learn: Machine Learning in Python“. In: *Journal of Machine Learning Research* 12 (2011), S. 2825–2830.
- [PL04] Ricardo B. C. Prudêncio und Teresa B. Ludermit. „Meta-learning approaches to selecting time series models“. In: *Neurocomputing* 61 (2004). Hybrid Neurocomputing: Selected Papers from the 2nd International Conference on Hybrid Intelligent Systems, S. 121–137. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2004.03.008. URL: <https://www.sciencedirect.com/science/article/pii/S0925231204002310>.
- [Qui86] J. R. Quinlan. „Induction of decision trees“. In: *Machine Learning* 1.1 (März 1986), S. 81–106. ISSN: 1573-0565. DOI: 10.1007/bf00116251.
- [Ras18] Sebastian Raschka. „Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning“. Nov. 2018. arXiv: 1811.12808 [cs.LG].
- [RM17] Sebastian Raschka und Vahid Mirjalili. *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow*. 2. Aufl. Packt Publishing, 2017. ISBN: 978-1-78712-593-3.
- [Ree15] Till Carsten Reek. „Power Modulation of Aluminium Reduction Cells – Operational Constraints and Process Limits“. Diss. University of New South Wales, 2015.
- [RP20] Michel Reverdy und Vinko Potocnik. „History of Inventions and Innovations for Aluminum Production“. In: *TMS 2020 149th Annual Meeting & Exhibition Supplemental Proceedings*. Springer International Publishing, 2020, S. 1895–1910. DOI: 10.1007/978-3-030-36296-6_175.
- [Rhe22] Rheinische Post. *Trimet halbiert Produktion in Essen - Kurzarbeit zunächst nicht geplant*. Abgerufen am 27.10.2023 um 17:11 Uhr. März 2022. URL: https://rp-online.de/wirtschaft/unternehmen/trimet-halbiert-produktion-in-essen-kurzarbeit-zunaechst-nicht-geplant_aid-67304147.
- [RIW+16] Torsten Rieck, Martin Iffert, Peter White, Renuka Rodrigo und Rik Kelchtermans. „Increased Current Efficiency and Reduced Energy Consumption at the TRIMET Smelter Essen Using 9 Box Matrix Control“. In: *Essential Readings in Light Metals*. Springer International Publishing, 2016, S. 817–824. DOI: 10.1007/978-3-319-48156-2_122.
- [Ris12] Andreas Risse. *Fertigungsverfahren der Mechatronik, Feinwerk- und Präzisionsgerätetechnik*. 1. Aufl. Wiesbaden: Springer Vieweg. in Springer Fachmedien Wiesbaden GmbH, 2012. 1553 S. ISBN: 9783834883124.

- [RWP00] Bernd Rolofs und Neal Wai-Poi. „The Effect of Anode Spike Formation on Operational Performance“. In: *Light Metals 2000*. Hrsg. von R. D. Peterson. The Minerals, Metals & Materials Society, 2000, S. 189–193.
- [Saa14] Ando Saabas. *Interpreting random forests*. Abgerufen am 24.04.2024 um 16:17 Uhr. Okt. 2014. URL: <https://blog.datadive.net/interpreting-random-forests/>.
- [Saa21] Ando Saabas. *treeinterpreter*. PyPi Paket Version 0.2.3. Jan. 2021. URL: <https://pypi.org/project/treeinterpreter>.
- [SFGJ20] David Salinas, Valentin Flunkert, Jan Gasthaus und Tim Januschowski. „DeepAR: Probabilistic forecasting with autoregressive recurrent networks“. In: *International Journal of Forecasting* 36.3 (Juli 2020), S. 1181–1191. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2019.07.001.
- [SBMW22] Alexander Sauer, Hans Ulrich Buhl, Alexander Mitsos und Matthias Weigold, Hrsg. *Energieflexibilität in der deutschen Industrie*. Bd. 2. Markt- und Stromsystem, Managementsysteme und Technologien energieflexibler Fabriken. Stuttgart: Fraunhofer Verlag, 2022. 535 S. ISBN: 9783839617786.
- [SS17] Robert H. Shumway und David S. Stoffer. *Time Series Analysis and Its Applications. With R Examples*. 4. Aufl. SpringerLink. Cham: Springer, 2017. ISBN: 9783319524528.
- [SO10] F. M. Soares und R. C. L. Oliveira. „Modelling of Temperature in the Aluminium Smelting Process using Neural Networks“. In: *The 2010 International Joint Conference on Neural Networks (IJCNN)*. Bd. 2010. IEEE, 2010. DOI: 10.1109/IJCNN.2010.5596645.
- [SSC+19] Alan Marcel Fernandes de Souza, Fábio Mendes Soares, Marcos Antonio Gomes de Castro, Nilton Freixo Nagem, Afonso Henrique de Jesus Bitencourt, Carolina de Mattos Affonso und Roberto Célio Limão de Oliveira. „Soft Sensors in the Primary Aluminum Production Process Based on Neural Networks Using Clustering Methods“. In: *Sensors* 19.23 (Nov. 2019), S. 5255. DOI: 10.3390/s19235255.
- [STC+08] Marco A. Stam, Mark P. Taylor, John J.J. Chen, Albert Mulder und Renuka Rodrigo. „Common Behaviour and Abnormalities in Aluminium Reduction Cells“. In: *Light Metals 2008*. 2008, S. 309–314.
- [STC+09] Marco A. Stam, Mark P. Taylor, John J.J. Chen, Albert Mulder und Renuka Rodrigo. „Development of a multivariate process control strategy for aluminium reduction cells“. In: *Light Metals 2009*. 2009, S. 311–315.
- [TRI13] TRIMET Aluminium SE. *Grundlagen der Elektrolyse*. Interne Präsentation der TRIMET Aluminium SE. 2013.
- [TRI21] TRIMET Aluminium SE. *Geschäftsbericht 2020/21*. Abgerufen am 28.08.2024 um 08:24 Uhr. Essen, Deutschland, 2021. URL: <https://www.trimet.eu/fileadmin/downloads/de/geschaeftsberichte/trimet-geschaeftsbericht-2021.pdf>.
- [TRI23] TRIMET Aluminium SE. *Nachhaltigkeitsbericht 2021*. Abgerufen am 28.08.2024 um 08:28 Uhr. Essen, Deutschland, 2023. URL: <https://www.trimet.eu/fileadmin/downloads/de/Berichte/trimet-nachhaltigkeitsbericht-2021-de.pdf>.
- [TRI24] TRIMET Aluminium SE. *TRIMET Standorte*. Abgerufen am 11.04.2024 um 11:08 Uhr. Essen, Deutschland, Apr. 2024. URL: <https://www.trimet.eu/de/trimet/standorte>.
- [THSM+20] Priyanga Dilini Talagala, Rob J. Hyndman, Kate Smith-Miles, Sevvandi Kandanaarachchi und Mario A. Muñoz. „Anomaly Detection in Streaming Nonstationary Temporal Data“. In: *Journal of Computational and Graphical Statistics* 29.1 (2020), S. 13–27. DOI: 10.1080/10618600.2019.1617160. URL: <https://doi.org/10.1080/10618600.2019.1617160>.

- [THA23] Thiyanga S. Talagala, Rob J. Hyndman und George Athanasopoulos. „Meta-learning how to forecast time series“. In: *Journal of Forecasting* (Feb. 2023). DOI: 10.1002/for.2963. URL: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.2963>.
- [Tan10] Gaurav Tandon. „Causal Factors in the Variation of the Cell Voltage in Industrial Aluminium Smelting Cells“. Diss. The University of Auckland, 2010.
- [TCY13] Mark P. Taylor, John J. J. Chen und Brent Richmond Young. *Control for Aluminum Production and Other Processing Industries*. CRC Press, 2013.
- [TUV16] J. Thonstad, T. A. Utigard und H. Vogt. „On the Anode Effect in Aluminum Electrolysis“. In: *Essential Readings in Light Metals*. Springer International Publishing, 2016, S. 131–138. DOI: 10.1007/978-3-319-48156-2_18.
- [Umw24] Umweltbundesamt (UBA). *Zeitreihen zur Entwicklung der erneuerbaren Energien in Deutschland*. Abgerufen am 05.07.2024 um 19:41 Uhr. Dessau-Roßlau, Deutschland, Feb. 2024. URL: https://www.bmwk.de/Redaktion/DE/Downloads/Energie/zeitreihen-zur-entwicklung-der-erneuerbaren-energien-in-deutschland-1990-2023.pdf?__blob=publicationFile&v=6.
- [WTBW11] Xiangwen Wang, Gary Tarcy, Eliezer Batista und Geff Wood. „Active Pot Control using Alcoa STARprobe™“. In: *Light Metals 2011*. Springer International Publishing, 2011, S. 491–496. ISBN: 9783319481609. DOI: 10.1007/978-3-319-48160-9_87.
- [WSH06] Xiaozhe Wang, Kate Smith und Rob Hyndman. „Characteristic-Based Clustering for Time Series Data“. In: *Data Mining and Knowledge Discovery* 13.3 (Mai 2006), S. 335–364. DOI: 10.1007/s10618-005-0039-x.
- [Wib08] Nils Wiberg. *Lehrbuch der Anorganischen Chemie*. 102. Aufl. Berlin, Boston: De Gruyter, 2008. ISBN: 9783110206845. DOI: 10.1515/9783110206845.
- [WBSK+24] Choon-Jie Wong, Jie Bao, Maria Skyllas-Kazacos, Barry Welch, Jing Shi, Nadia Ahli, Amal Aljasmī, Mohamed Mahmoud und Mustafa Mustafa. „Predicting Electrolyte and Liquidus Temperatures of Aluminium Smelting Cells for Power Modulation Using Dynamic Model“. In: *Light Metals 2024*. Springer Nature Switzerland, 2024, S. 445–452. ISBN: 9783031503078. DOI: 10.1007/978-3-031-50308-5_57.
- [WTL14] David S. Wong, Alton Tabereaux und Pascal Lavoie. „Anode Effect Phenomena During Conventional AEs, Low Voltage Propagating AEs & Non-Propagating AEs“. In: *Light Metals 2014*. 2014, S. 529–534. DOI: 10.1007/978-3-319-48144-9_90.
- [ZXWZ18] Z. Zhang, G. Xu, H. Wang und K. Zhou. „Anode Effect prediction based on Expectation Maximization and XGBoost model“. In: *2018 IEEE 7th Data Driven Control and Learning Systems Conference*. 2018, S. 560–564.
- [ZJX+94] Qiu Zhuxian, Li Jingjiang, Cui Xiaoli, Kai Grjotheim, Halvor Kvande und Harald A. Øye. „Continuous bath temperature measurements in Al electrolysis cells“. In: *JOM* 46.8 (Aug. 1994), S. 28–30. ISSN: 1543-1851. DOI: 10.1007/bf03220771.

