Autoadaptives prädiktives Modell zur Quantifizierung von Gleichzeitigkeitseffekten in Lastverteilungen urbaner Energiesysteme

Von der Fakultät für Bauingenieurwesen der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Daniel Koschwitz

Berichter: Univ.-Prof. Dr.-Ing. habil. Christoph van Treeck
Univ.-Prof. Dr.-Ing. Christoph Nytsch-Geusen

Tag der mündlichen Prüfung: 15.01.2020

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Kurzfassung

Die vorliegende Arbeit beschreibt ein autoadaptives prädiktives Modell zur Quantifizierung von Gleichzeitigkeitseffekten in Lastverteilungen urbaner Energiesysteme mit dem Namen *AMSA* (*Auto-adaptive Model for Simultaneity Analysis*). Der mathematische Kern von *AMSA* beruht auf Techniken des maschinellen Lernens in Kombination mit Methoden der Energiedatenanalyse und ermöglicht eine Integration in intelligente vernetzte Energie-Informationssysteme. *AMSA* ist in Matlab implementiert und basiert auf einem modularen Aufbau, der durch verschiedene funktionale Teilblöcke definiert ist, wodurch einzelne Modifikationen von Schnittstellenfunktionen sowie modulare Erweiterungen nicht den numerischen Rechenkern beeinflussen.

Die bisher hauptsächlich auf Erfahrungswerten und empirischen Analysen basierende Quantifizierung von Gleichzeitigkeitsfaktoren zur ökonomisch optimierten Auslegung von Energieerzeugungs- und Energieverteilungssystemen beruht auf einer historischen Messdatenbasis. Die herkömmliche Ableitung von Kennlinien aus Messdatenbeständen impliziert einen mit Zunahme der Gebäudegruppengröße sinkenden Gleichzeitigkeitsfaktor, welcher für Neuplanungen zentraler Energieversorgungssysteme angesetzt wird. Im Sinne der Optimierung letzterer ist es jedoch notwendig, sowohl historische und gegenwärtige als auch zukünftige Informationen zum Energiebedarf von Gebäuden und Gebäudegruppen szenariobasiert sowie unter Berücksichtigung von Unsicherheit abzubilden. Ferner bedarf es der Analyse von Ähnlichkeiten in den Lastverläufen von Gebäuden, um gleichzeitig wiederkehrende Muster erkennen und für die Identifizierung geeigneter Energieversorgungsverbünde zu nutzen. Vor diesen Hintergrund dient das entwickelte Modell dem Wissensgewinn aus einer komplexen variierenden Datenbasis zur Identifizierung von Entscheidungskorridoren in der mittel- und langfristigen Planung im Rahmen der Entwicklung von Quartieren und urbanen Regionen sowie kurzfristig ableitbarer Lastmanagementstrategien innerhalb der Energieversorgung.

Grundlage für die Ableitung geeigneter Methoden zur Entwicklung von Teilmodellen bilden detaillierte Literaturrecherchen und -analysen zu Methodenfamilien mit ihren spezifischen Charakteristika sowie Erkenntnissen aus anwendungsbasierten Studien. Für die Abbildung zukünftiger Lastzustände dient ein Ensemble-Modell, das aus zwei Rekurrenten Neuronalen Netzen unterschiedlicher Tiefe und zwei Support Vector Machine Konfigurationen auf Basis verschiedener Kernel besteht. In Abhängigkeit von der Beschaffenheit der Eingangsdaten erfolgt eine automatisierte Auswahl der gebäudeindividuell geeigneten Prognosemethode auf Grundlage der Bewertung eines Fehlerindices in der jeweiligen Methodentestphase. Zur Analyse des Einflusses zukünftig variierender Wetterbedingungen und Gebäudesanierungen auf die Gleichzeitigkeitseffekte der urbanen Lastverteilung werden mathematisch-statistische Methoden zur Lastgangmodifikation eingeführt. Für die Mustererkennung in Lastverteilungen werden mit Self-Organizing Maps in Kombination mit Learning Vector Quantization kompetitive Neuronale Netze eingesetzt. In diesem Zusammhang erfogt im ersten Schritt eine automatisierte Gruppierung der Gebäude entsprechend ihrer charakteris-

tischen Lastprofile. Im zweiten Schritt wird die Gebäudegruppe, deren Lastprofil jenem des Gesamtquartiers am ähnlichsten ist, identifiziert. Darauf aufbauend erfolgt die Berechnung von Gleichzeitigkeitsfaktoren und Spitzenlastbeiträgen von Gebäuden und Gebäudegruppen.

Zur Modelldemonstration im abschließenden Teil dieser Arbeit dienen Monitoringdaten zum gebäudespezifischen Wärme- und Kälteverbrauch eines Forschungscampus. Im Rahmen dessen werden neben der statistischen Aufbereitung der Datenbasis eingangs Ergebnisse von Vorstudien zu Teilmodellen vorgestellt. Diese umfassen das Clustering thermischer Lastprofile, einen detaillierten Vergleich der verwendeten Prognosemethoden untereinander sowie langfristige Wärmelastprognosen auf Basis von Sanierungsszenarien. Darüber hinaus ermöglichen konventionell hergeleitete Kennlinien für den Gleichzeitigkeitsfaktor eine Einordnung und Evaluierung der mit *AMSA* erzielten Ergebnisse. Auf Gesamtmodellebene zeigen sich verschiedene Vorteile des Modells: Im Sinne eines kurzfristigen Lastmanagements muss durch die Einbeziehung des Gleichzeitigkeitsfaktors weniger Erzeugerleistung vorgehalten werden. Unter Berücksichtigung der ähnlichkeitsbasierten Gruppenbildungen lassen sich Schlüsselgebäude und die Gebäudegruppe identifizieren, deren temporäre Spitzenlastverschiebung den größten Anteil an dem zu erwartenden Spitzenleistungsbedarf auf Gesamtquartiersebene hat. Diese Informationen ermöglichen prädiktive Be- und Entladestrategien für Speichersysteme im Sinne eines Demand Side Managements. Hinsichtlich längerfristiger Planungszeiträume wird gezeigt, dass die Nutzung von Gleichzeitigkeitsfaktoren in Kombination mit ähnlichkeitsbasierten Gebäudegruppierungen zur Identifizierung möglicher Energieverbundsysteme eignen. Die Analyse zukünftiger Wetter- und Sanierungsszenarien demonstrieren ihren modifizierenden Einfluss auf die Ergebnisse der Gleichzeitigkeitsanalysen und unterstreichen den Bedarf eines autoadaptiven prädiktiven Modells. Das Ziel zukünftiger Forschungsarbeiten sollte in der Validierung von *AMSA* unter Berücksichtigung detaillierter Informationsgrundlagen liegen.

Abstract

The present work describes an auto-adaptive predictive model for quantifying simultaneity effects within load distributions of urban energy systems named *AMSA* (Auto-adaptive Model for Simultaneity Analysis). The mathematical structure of *AMSA* is based on Machine Learning techniques in combination with methods related to energy data analysis enabling the integration in intelligent interconnected energy information systems. *AMSA* is implemented in Matlab using a modular design including different functional components. Thus, individual modifications of interface functions and modular extensions do not influence the numerical calculation core.

Previous simultaneity factor quantification for economically optimized dimensioning of energy generation plants and networks is based on experience values and empirical analysis focusing on historical measurement data. Conventional derivations of characteristic curves using measurement databases imply a decreasing simultaneity factor in combination with increasing building group sizes, which is used for replanning energy supply systems. In order to optimize the latter, it is necessary to know historical and current as well as scenario-based future energy demand of single buildings and building groups taking uncertainty into account. Moreover, similarity analysis concerning building load profiles is required to recognise simultaneously recurring patterns in order to identify energy network systems. In this context, the developed model serves for knowledge gain from varying complex databases to identify decision ranges for long- and medium-term planning regarding the development of districts and urban regions. Furthermore, on operational level, it improves short-term load management strategies concerning energy supply.

The derivation of suitable methods for component model development is based on appropriate scientific literature research and analysis regarding method categories with their specific characteristics as well as findings from application-based studies. An ensemble model serves to calculate future load conditions, consisting of two recurrent neural networks of various depth and two support vector machine configurations using different kernels. According to input data characteristics, a building-specific suitable prediction method is automatically chosen based on an error evaluation during the method testing period. In order to analyse influences of varying weather conditions and building retrofits on simultaneity effects within urban load distributions, mathematical-statistical methods for load profile modifications are introduced. For pattern recognition of load distributions, self-organizing maps combined with learning vector quantification are used, which is assigned to competitive neural networks category. In this regard, initially, buildings are automatically grouped according to their characteristic load profile. Secondly, the building group is identified, which provides the most correlated load profile compared to the district load profile. Subsequently, simultaneity factors and peak load shares of buildings an building groups are calculated.

In the final part of this work, monitoring data of building-specific heating and cooling consumption of a research campus serve for model demonstration. In this context, besides statistical analysis of the database, results of preliminary studies concerning component models are presented. These include clustering of thermal load profiles, a detailed comparison concerning prediction models used as well as long-term heating load predictions based on retrofit scenarios. Furthermore, conventionally derived curves for simultaneity factors enable classification and evaluation of the results using *AMSA*. Taking the corporate model level into consideration, various model-related advantages may be shown: Regarding short-term load management, less generating capacities have to be provided. Regarding similarity-based grouping, key buildings and the building group may be identified, where temporal peak load shifting contains the largest share of the expected peak load on district level. This information enables predictive loading and unloading strategies for storage systems in terms of demand side management. With respect to longer-term planning periods, it is shown that simultaneity factors combined with similarity-based building classification suit for the identification of possible energy network systems. Future weather and retrofit scenario analysis demonstrates external influences on the results of simultaneity analysis, which emphasizes the demand for an auto-adaptive predictive model. The scope of future research work should focus on the validation of *AMSA* taking detailed information bases into account.

Für meine Familie

Vorwort

Die vorliegende Arbeit wurde während meiner Tätigkeit am Lehrstuhl für Energieeffizientes Bauen (E3D) der Fakultät Bauingenieurwesen an der RWTH Aachen University angefertigt. Vor diesem Hintergrund möchte ich Univ.-Prof. Dr.-Ing. habil. Christoph van Treeck für sein Vertrauen sowie seine Unterstützung und Förderung während der Vorbereitungs- und Umsetzungsphase dieser Arbeit danken. Ein großer Dank geht ebenfalls an Prof. Dr.-Ing. Christoph Nytsch-Geusen als Zweitgutachter dieser Arbeit und an Univ.-Prof. Dr.-Ing. Jörg Blankenbach als Vorsitzenden des Prüfungsausschusses.

Weiterhin möchte ich mich bei Dr.-Ing. Jérôme Frisch für das Feedback, die Diskussionen im Kontext der Vorstudien zur Modellentwicklung und der Vorbereitung auf die mündliche Prüfung bedanken. Bei meinen ehemaligen sowie gegenwärtigen Kolleginnen und Kollegen des Lehrstuhls für Energieeffizientes Bauen (E3D) bedanke ich mich für die gute und vertrauensvolle Zusammenarbeit in einer sehr angenehmen Arbeitsatmosphäre.

Bedanken möchte ich mich explizit auch bei meinen ehemaligen Kollegen Dipl.-Ing. Mark-Alexander Brüntjen sowie Dr.-Ing. Daniel Wölki, die mir mit dem einen oder anderen fachlichen und freundschaftlichen Rat sehr weiter geholfen haben. Im Rahmen meiner Masterarbeit bei Mark-Alexander ist das Software-Tool SCoTED entstanden, dessen Entwicklung mich für die Thematik dieser Arbeit begeistert hat. Mit Daniel habe ich viele Lehrveranstaltungen betreut und zahlreiche spannende interdisziplinäre Fragestellungen diskutiert, die meinen Blickwinkel auf verschiedene Forschungsthemen verändert haben.

Ein besonders großer Dank geht an meine Eltern, die mich immer unterstützen und mir den größtmöglichen Rückhalt und uneingeschränktes Vertrauen schenken. Gleiches verdanke ich auch meiner lieben Schwester, die gerade auch in schwierigeren Phasen immer ein offenes Ohr für mich hat. Der familiäre Zusammenhalt hat die notwendige Ruhe und Gelassenheit für das Gelingen dieser Arbeit vermittelt. Abschließend möchte ich meiner Partnerin, Katja, für ihre große Unterstützung, das Verständnis und den Mut in stressigen Zeiten gerade in der Endphase dieser Arbeit danken. Ich bin sehr glücklich, dich kennen gelernt zu haben.

Aachen, Januar 2020

Daniel Koschwitz
koschwitz@e3d.rwth-aachen.de

Nomenklatur

Symbole

A	Nettogrundfläche eines Gebäudes
B	Bias
C	Konstante
D	Grad einer polynomialen Funktion
E	Fehlerfunktion
J	Jakobi-Matrix
K	Kernel-Funktion
L	Verlustfunktion
M	Anzahl
N	Anzahl
P	Elektrische Leistung
$\dot{Q}$	Thermische Leistung
R	Korrelationsmatrix
S	Geometrischer Gebäudeschwerpunkt
V	Variationskoeffizient
W	Gewichtungsfaktor
X	Datensatz
a	Beliebiges initiales Cluster
c	Cluster
d	Dominierfaktor
f	Funktionswert
h	Nichtlineare Aktivierungsfunktion
k	Clusteranzahl
l	Level im hierarchischen Clustering
m	Referenzvektor
p	Parametervektor
q	Wasserverbrauch
r	Korrelationskoeffizient
s	Klasse
t	Zielwert
w	Gewichtungsfaktor
x	Eingangswert
y	Ausgangswert

z	Eingangswert eines Neurons in einer beliebigen Schicht

Griechische Symbole

Γ	Absolute Abweichung vom lokalen Median
Λ	Rekonstruktionsvektor
Ω	Summe der Ausgangsgrößen der verdeckten Schicht eines KNN
T	Endzeitpunkt einer Betrachtungsperiode
α^*	Lagrange Multiplikator
α	Lagrange Multiplikator
δ	Differenz
ε	Fehler
η	Lernrate
γ^*	Lagrange Multiplikator
γ	Lagrange Multiplikator
ι	Dämpfungsterm
κ	Gruppenspezifischer Spitzenlastanteil
λ	Zeitverzögerungsparameter
μ	Arithmetischer Mittelwert
ν	Summe der gewichteten Eingangswerte eines Neurons
ω	Koordinate
ϕ	Gleichzeitigkeitsfaktor
ψ	Ähnlichkeitsfaktor
ρ	Faktor für die Lernrate
σ	Standardabweichung
τ	Zeitvariable
θ	Temperatur
ξ^*	Schlupfvariable
ξ	Schlupfvariable

Indices

A	Außen
Geb	Gebäude
G	Gruppe
HG	Heizgrenze
K	Kälte
NA	Normaußentemperatur
NI	Norminnentemperatur
$Nenn$	Nennleistung
$Quart$	Quartier
RW	Raumwärme

TWW	Trinkwarmwasser
T	Endzeitpunkt des zeitlichen Betrachtungshorizonts
W	Wärme
ZE	zentrale Energieerzeugung
agg	aggregiert
c	Clusterzentrum
d	durchschnittlich
ges	gesamt
$hist$	historisch
inf	inferior
i	Laufvariable
j	Laufvariable
k	Laufvariable
max	maximal
med	Median
min	Minimum
$norm$	normiert
san	saniert
st	stündlich
sup	superior
wa	witterungsabhängig
wu	witterungsunabhängig
x	x-Koordninate
y	y-Koordninate

Abkürzungsverzeichnis

Akronyme

AMSA	Auto-adaptive Model for Simultaneity Analysis
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
AVA	All-Versus-All
BASICS	BATTELLE Scenario Inputs to Corporate Strategies
CART	Classification and Regression Trees
CDC	Climate Data Center
CHAID	Chi-squared Automatic Interaction Detector
DT	Decision Tree
DWD	Deutscher Wetterdienst
EN	Europäische Norm
FCM	Fuzzy Cognitive Map
GIS	Geoinformationssystem
IFS	Interactive Future Simulations
INTERAX	Interactive Cross-Impact Simulation
JSON	JavaScript Object Notation
KKT	Karush-Kuhn-Tucker
KML	Keyhole Markup Language
KNN	Künstliches Neuronales Netz
LDA	Linear Discriminant Analysis
LM	Levenberg-Marquardt
LVQ	Learning Vector Quantization

MAE	Mean Average Error
MSE	Mean Squared Error
NARX	Nonlinear Autoregressive Network with Exogenous Inputs
NGF	Nettogrundfläche
OVA	One-Versus-All
PLR	Peak Load Ratio
PMT	Probabilistic Modified Trends
RBF	Radial Basis Function
RESTful	Representational State Transfer
RF	Random Forests
RMSE	Root Mean Squared Error
RNN	Rekurrentes Neuronales Netz
SCoTED	Simplified Calculation of Thermal Energy Demand
SDHA	Swedish District Heating Association
SOM	Self-Organizing Map
SVM	Support Vector Machine
SVM-R	Support Vector Machine Regression
TGA	Technische Gebäudeausrüstung
TIA	Trend Impact Analysis
TRJ	Testreferenzjahr
VDI	Verein Deutscher Ingenieure

Inhaltsverzeichnis

Nomenklatur **i**
Symbole . . . i
Griechische Symbole . . . ii
Indices . . . ii

Abkürzungsverzeichnis **v**
Abkürzungsverzeichnis . . . v

1 Einleitung **1**
1.1 Hintergrund . . . 1
1.2 Zielsetzung dieser Arbeit . . . 5
1.3 Hypothesen . . . 5
1.4 Gliederung und struktureller Aufbau . . . 6

2 Gleichzeitigkeiten in der urbanen thermischen Lastverteilung **9**
2.1 Ermittlung des thermischen Energiebedarfs urbaner Regionen . . . 9
2.1.1 Theoretischer Hintergrund . . . 9
2.1.2 Studien . . . 14
2.1.3 Implikationen und Anforderungen an die Modellierung . . . 16
2.2 Ansätze zur Quantifizierung des Gleichzeitigkeitsfaktors . . . 17
2.2.1 Theoretischer Hintergrund . . . 17
2.2.2 Studien . . . 19
2.2.3 Implikationen und Anforderungen an die Modellierung . . . 25
2.3 Zusammenfassung . . . 26

3 Prognosemodelle und Szenariotechniken **29**
3.1 Prognosemodelle . . . 29
3.1.1 Theoretischer Hintergrund . . . 30
3.1.2 Studien . . . 35
3.1.3 Implikationen und Anforderungen an die Modellierung . . . 36
3.2 Szenariotechniken . . . 37
3.2.1 Theoretischer Hintergrund . . . 37
3.2.2 Studien . . . 40
3.2.3 Implikationen und Anforderungen an die Modellierung . . . 41
3.3 Zusammenfassung . . . 41

4 Klassifikation energetischer Lastprofile 45
4.1 Clustering . . . 45
4.1.1 Theoretischer Hintergrund . . . 45
4.1.2 Studien . . . 52
4.1.3 Implikationen und Anforderungen an die Modellierung . . . 53
4.2 Klassifikation . . . 54
4.2.1 Theoretischer Hintergrund . . . 54
4.2.2 Studien . . . 59
4.2.3 Implikationen und Anforderungen an die Modellierung . . . 60
4.3 Zusammenfassung . . . 60

5 Entwicklung eines autoadaptiven prädiktiven Modells 65
5.1 Methoden zur Quantifizierung von Gleichzeitigkeitseffekten . . . 65
5.1.1 Erweiterung der Definition des Gleichzeitigkeitsfaktors . . . 65
5.1.2 Einführung eines Ähnlichkeitsfaktors . . . 67
5.1.3 Identifizierung von gruppenspezifischen Spitzenlastanteilen . . . 67
5.2 Teilmodell Datenvorverarbeitung . . . 68
5.2.1 Identifizierung von Ausreißern . . . 68
5.2.2 Normalisierung der Daten . . . 70
5.3 Teilmodell Prognose-Ensemble . . . 70
5.3.1 Rekurrente Neuronale Netzwerke . . . 72
5.3.2 Support Vector Machine Regression . . . 81
5.3.3 Entscheidungslogik . . . 88
5.4 Teilmodell prädiktive Lastszenarien . . . 89
5.4.1 Zukünftige Wetterszenarien . . . 90
5.4.2 Lastgangbeeinflussungen durch Gebäudesanierungen . . . 91
5.5 Teilmodell Clustering und Klassifikation . . . 93
5.5.1 Self-Organizing Map . . . 95
5.5.2 Learning Vector Quantization . . . 98
5.6 Zusammenfassung . . . 99

6 Modellimplementierung 101
6.1 Anforderungen . . . 101
6.2 Softwarearchitektur . . . 102
6.2.1 Grafische Benutzeroberflächen . . . 103
6.2.2 Klassenhierarchie . . . 111
6.2.3 Datenstrukturen . . . 112
6.3 Eingliederung in ein Energie-Informationssystem . . . 113
6.4 Zusammenfassung . . . 115

7 Modelldemonstration 117
7.1 Datengrundlage . . . 117
7.2 Demonstration der Modellkomponenten . . . 124
7.2.1 Ergebnisse der Datenvorverarbeitung . . . 124

7.2.2 Ergebnisse zu thermischen Lastprognosen . . . 126
7.2.3 Ergebnisse des Clusterings von Lastprofilen . . . 144
7.3 Ergebnisse der Quantifizierung von Gleichzeitigkeitseffekten . . . 149
7.3.1 Konventionelle Ermittlung von Gleichzeitigkeitsfaktoren . . . 149
7.3.2 Quantifizierung von Gleichzeitigkeitseffekten mit AMSA . . . 155
7.4 Zusammenfassung . . . 166

8 Schlussbetrachtung 169
8.1 Zusammenfassung der wichtigsten Ergebnisse . . . 169
8.2 Ausblick . . . 172

Literatur 175

Anhang 209

A Methoden zur Datenanalyse 209
A.1 Statistische Kennwerte . . . 210
A.2 Korrelationsanalyse . . . 211
A.3 Konventionelle Herleitung von Kennlinien für den Gleichzeitigkeitsfaktor . . . 212
A.4 Ermittlung der Koordinaten zu den geometrischen Gebäudeschwerpunkten . . . 216

B Software- und Datenstrukturen 219
B.1 UML-Diagramm . . . 220
B.2 Beschreibung der wichtigsten Klassen und Funktionen . . . 221
B.3 Datenformate und Kennzeichnungsschlüssel . . . 225

C Gebäudeinformationen 227

D Charakteristische Heizlasten 233
D.1 Thermische Leistungen in Abhängigkeit von der Außentemperatur . . . 234
D.2 Ergänzende Ergebnisse der Gleichzeitigkeitsanalyse . . . 236

E Ergänzende Prognoseergebnisse 237

KAPITEL 1

Einleitung

Im Rahmen dieses einleitenden Kapitels erfolgt in Abschnitt 1.1 eine kurze Einführung in die Thematik und den Zusammenhang der Quantifizierung von Gleichzeitigkeitsfaktoren vor dem Hintergrund der thermischen Energieversorgung urbaner Regionen. Daran anknüpfend werden in Abschnitt 1.2 die Zielsetzung dieser Arbeit auf Grundlage des bestehenden Forschungsbedarfs in diesem Bereich dargelegt und in Abschnitt 1.3 mit den daraufhin abgeleiteten Hypothesen inhaltlich verknüpft. In Abschnitt 1.4 erfolgt die Übertragung der einleitenden Inhalte in die Gliederung und den strukturellen Aufbau dieser Arbeit.

1.1 Hintergrund

Die Energieversorgung von Gebäuden und urbanen Regionen steht im Kontext internationaler und nationaler Ziele vor der Frage nach geeigneten Systemen, die sowohl Erneuerbare Energien beinhalten als auch die Versorgungssicherheit gewährleisten. Gleichzeitig müssen diese Systeme wirtschaftlich sein. Die fortwährende Diskussion über Vor- und Nachteile dezentraler und zentraler Energieversorgungsvarianten führt zu keinem verallgemeinerbaren Ergebnis. Vielmehr muss im Einzelfall entschieden werden, welche Variante sich aus ökonomischer und ökologischer Sicht eignet. Die Realisierung zukünftiger nachhaltiger Energiesysteme ist eng mit der Weiterentwicklung von urbanen Wärme- und Kälteversorgungstechnologien verbunden, um die Energieeffizienz zu erhöhen, Kosten zu senken oder die Einführung von smarten Energiesystemen zu ermöglichen (Lund et al., 2014; Connolly et al., 2014).

In diesem Zusammenhang fokussieren sich Forschungsaktivitäten bezüglich der zentralen thermischen Energieversorgung von urbanen Regionen auf die Entwicklung effizienter und nachhaltiger Wärme- und Kälteversorgungssysteme. Während Fernwärmesysteme international schon seit den 1930er und in Europa seit den 1940er Jahren schrittweise etabliert wurden, weist die Einführung von Fernkältesystemen eine weitaus geringere traditionelle Verankerung in der Entwicklung neuer Energieversorgungssysteme auf. In Europa wird aktuell die Anzahl vorhandener Fernwärmesysteme auf ungefähr 6000 und die Anzahl vorhandener Fernkälteversorgungssysteme auf circa 150 geschätzt. (Werner, 2017) Zur Zeit besteht international und in Deutschland vornehmlich im Sektor der Nichtwohngebäude

Kühlbedarf (Werner, 2016). Aufgrund der weltweiten klimatischen Veränderungen mit hohen Temperaturniveaus im Sommer und gleichzeitig vergleichsweise milden Wintern sowie den steigenden Anforderungen an gebäudespezifische Dämmstandards, nimmt der Wärmebedarf zukünftig ab und der Kühlbedarf speziell im Wohngebäudebereich zu (Olonscheck et al., 2011; Christenson et al., 2006; Werner, 2016). In der Konsequenz reduziert sich auch der durch die modifizierte Charakteristik der thermischen Hülle induzierte Wärme- und Kühlbedarf, wobei gleichzeitig der Einfluss des nutzergesteuerten Energiebedarfs steigt (Guerra Santin et al., 2009).

Im Hinblick auf die urbane Energieversorgung differieren die Leistungsanforderungen an die Bereitstellung thermischer Energie zwischen den verschiedenen Gebäuden als Energieabnehmer, im Speziellen auch die zeitliche Verteilung des Spitzenlastbedarfs (Winter et al., 2001). In der Folge reduziert sich simultan die Gesamtleistungsanforderung der Abnehmergruppe, da in der Regel nicht alle Gebäude zum selben Zeitpunkt einen Energiebedarf auf Nennleistungsniveau aufweisen. Hinsichtlich der zentralen Energieerzeugerauslegung und der Netzdimensionierung ist dieser Effekt zu berücksichtigen, um Überdimensionierungen und damit einhergehende ökonomische und ökologische Ineffizienzen zu vermeiden. Die Menge der zu transportierenden thermischen Energie, welche ab dem Energieerzeuger vorgesehen werden muss, richtet sich nach dem Bedarf einzelner Abnehmergruppen, welche über Netztrassen mit dem Energieerzeuger verbunden sind. (Obernberger, 1997) Für die Dimensionierung einzelner Rohrabschnitte sind dagegen jeweils sämtliche Abnehmer eines hydraulischen Strömungspfades über den betreffenden Rohrabschnitt maßgebend für die Berechnung des Gleichzeitigkeitsfaktors (Winter et al., 2001). Letzterer wird in der Regel auf Grundlage der Auswertung historischer Messdaten bestimmt und in der Praxis als Erfahrungswert, ungeachtet der jeweiligen Charakteristika des betrachteten Energiesystems, weiter gegeben.

Im Zusammenhang mit der Nutzung von Gleichzeitigkeitseffekten auf Quartiersebene rückt zudem das Potenzial der temporären Verschiebung von energetischen Lasten in den Fokus. Hierbei werden gebäudespezifische Spitzenlasten mittels Speichern abgedeckt, sodass sich die Lasten für die Energieerzeuger gleichmäßig verteilen lassen (Frederiksen et al., 2014). Dadurch kann ineffizientes kurzes An- und Abschalten der Erzeugungseinheiten infolge unerwarteter Laständerungen vermieden werden. Die Erweiterung der Betrachtungsebene vom Einzelgebäude auf Gebäudegruppen, die zusammen versorgt werden, führt zu der Fragestellung, welche Gebäude zu welchem Zeitpunkt und in welcher Höhe zur Spitzenlast der gesamten Gebäudegruppe beitragen (Guan et al., 2016). Mit diesem Wissen lassen sich einerseits Strategien für den kurzfristigen Betrieb des Energieversorgungssystems infolge der Verschiebung von Energielasten einzelner Gebäude ableiten und andererseits die Planung neuer oder die Modifizierung bestehender Systeme optimieren. Dazu ist es notwendig, charakteristische Gleichzeitigkeiten in den Lastgängen zu identifizieren und zu quantifizieren, um Zeitpunkte und die Höhe der Lasten, welche verschoben werden sollen, bestimmen zu können. Ferner sind Schlüsselgebäude für die Planung der Energieversorgung zu lokalisieren, deren Lasten mittels der Nutzung von Speichern zukünftig vermehrt

verschoben werden können, um Spitzenlasten von Quartieren und urbanen Regionen zu reduzieren. Im Hinblick auf die Auslegung und Auswahl zukünftiger urbaner Energiesysteme stellt der Gleichzeitigkeitsfaktor daher einen zentralen Faktor dar, welcher wesentlich von der individuellen Charakteristik thermischer Lastprofile von Gebäuden innerhalb der zu versorgenden Energieabnehmergruppen beeinflusst wird. Ferner sind Dynamiken in der Zusammensetzung und strukturellen Beschaffenheit eines Quartiers in kurz-, mittel- und langfristigen Lastszenarien zu berücksichtigen, um bestmögliche Entscheidungen in Bezug auf die Systemwahl treffen zu können. Abbildung 1.1 visualisiert dazu exemplarisch die Transformation von thermischen Energiesystemen infolge digitaler und technischer Innovationen sowie der Integration erneuerbarer Energiequellen. Im Entscheidungszeitpunkt $\tau = 0$ stehen in der Planung historische energetische Lastprofile der Gebäude auf Basis der historischen Informationsgrundlage über das konventionelle Energiesystem aus $\tau = -1$ zur Verfügung. Die Aufgabe, ein geeignetes zukünftiges Energiesystem und mögliche Subsysteme in $\tau = 1$ zu definieren und im Rahmen eines Variantenvergleichs zu bewerten, bedarf der Quantifizierung energetischer Lastprofilszenarien. Letztere unterscheiden sich auf Gesamtsystemebene beispielsweise durch variierende Wetterentwicklungen in den kommenden Jahren, gebäudebezogene Sanierungspläne oder Rückbau sowie Neubau. Um diese Anforderungen in Kombination mit detaillierten Gleichzeitigkeitsanalysen erfüllen zu können, bedarf es eines autoadaptiven Modells, das sich an die jeweilige Informationsgrundlage einer Planungs- und Betriebsaufgabe anpasst. Aus dem abgeleiteten Forschungsbedarf erfolgt im Rahmen dieser Arbeit daher die Entwicklung des Software-Tools *AMSA* (Auto-adaptive Model for Simultaneity Analysis).

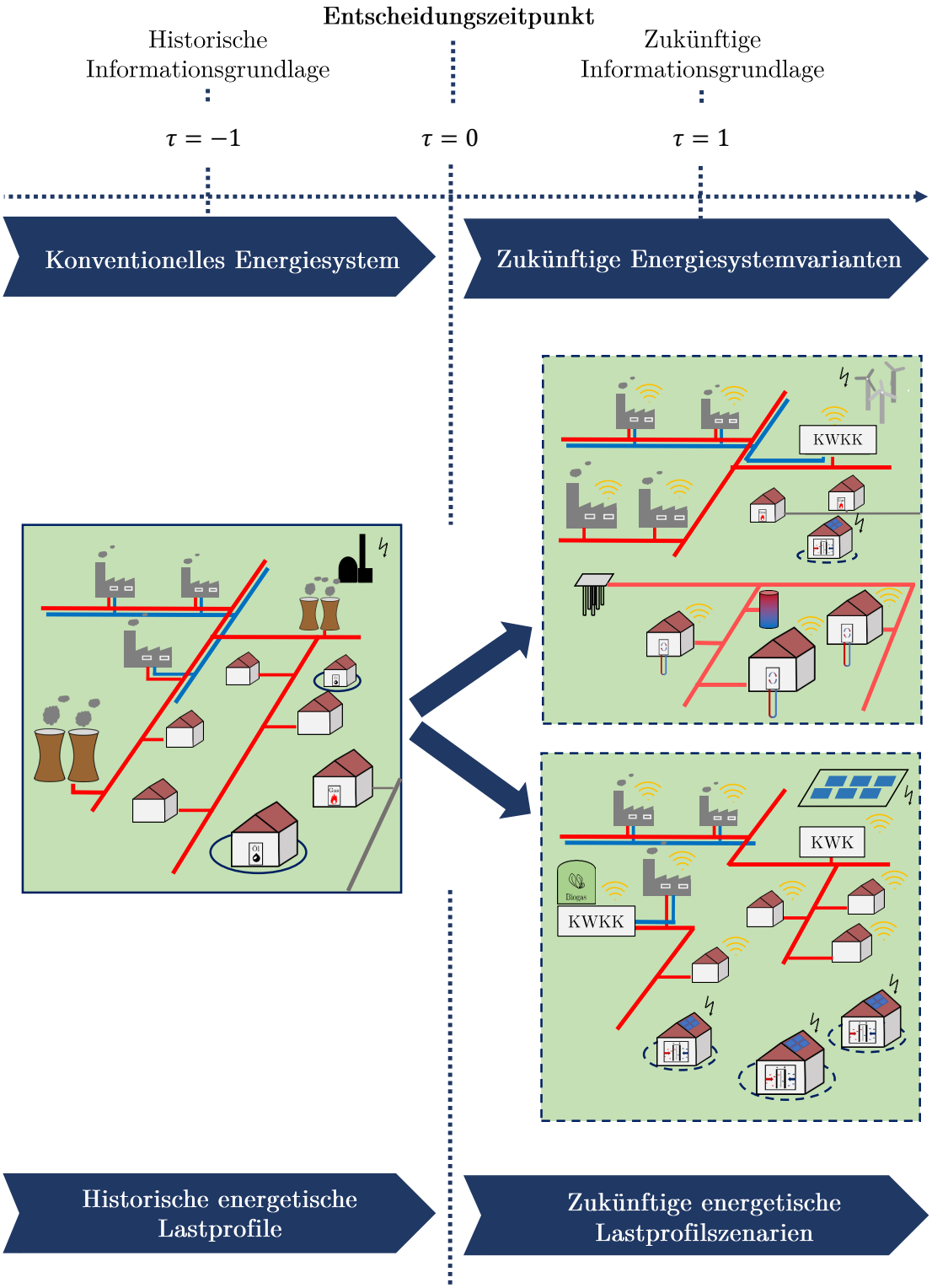


Abbildung 1.1: Transformation urbaner thermischer Energiesysteme

1.2 Zielsetzung dieser Arbeit

Vor dem Hintergrund der Entwicklung und Implementierung des Modells dieser Arbeit werden folgende Teilziele formuliert:

I. Detaillierte Aufbereitung des Stands der Technik und Wissenschaft zur Ermittlung von Gleichzeitigkeitsfaktoren

II. Analyse des Stands der Technik und Wissenschaft zu den methodischen Grundlagen der Modellentwicklung

   i. Kategorisierung von Methoden zur Prognose des thermischen Energiebedarfs von Gebäuden, Clustering und Klassifikation von energetischen Lastprofilen sowie Szenariotechniken

   ii. Identifizierung von charakteristischen Unterschieden verschiedener Methoden sowie Methodenkategorien und Ableitung von geeigneten Methoden für die Zielerreichung dieser Arbeit

III. Aufarbeitung der mathematischen Grundlagen zu den verwendeten Methoden im Rahmen der Modellentwicklung und -evaluierung sowie Herleitung geeigneter Parameter für die Modellkonfiguration

IV. Erarbeitung eines Softwarekonzepts im Rahmen der Implementierung des entwickelten Modells

   i. Berücksichtigung einer modularen Klassenstruktur

   ii. Entwicklung einer strukturierten, übersichtlichen Benutzeroberfläche

   iii. Integration von Schnittstellen zu bestehenden Komponenten eines Energie-Informationssystems

V. Dokumentation der Vorteile des entwickelten Modells sowie strukturierte Analyse der Ergebnisse anhand von Vorstudien und Teil- sowie Gesamtmodelldemonstrationen

1.3 Hypothesen

I. Mittels Prognosemethoden lassen sich sowohl kurz- und mittel- als auch langfristige Lastzustände hinreichend präzise simulieren, um zukünftige Gleichzeitigkeitseffekte in der thermischen Lastverteilung zu berücksichtigen.

II. Hinsichtlich der kurzfristigen, operativen Planung des Betriebs von urbanen Energieversorgungssystemen ergibt sich ein Mehrwert infolge der Berücksichtigung von Gleichzeitigkeitseffekten auf Basis prognostizierter zukünftiger Lastprofile.

III. Die Integration von gebäudespezifischen Sanierungsmaßnahmen in die langfristigen thermischen Lastprognosen ist für die energetische Planung von Quartiersentwicklungsmaßnahmen notwendig, um die Abbildung möglicher Lastszenarien realisieren und damit zukünftige Energieversorgungsvarianten sensitiv analysieren zu können.

IV. Die Gruppierung von gebäudebezogenen Lastcharakteristika im Sinne ähnlicher Lastgänge ermöglicht die Quantifizierung des Lastverschiebungspotenzials und liefert Informationen zum Einsatz von Speichertechnologien.

V. Für die automatisierte Ähnlichkeitsanalyse eignet sich insbesondere ein kombiniertes Modell aus Clustering- und Klassifikationsmethoden.

VI. Eine kombinierte Betrachtung aus Ähnlichkeitsfaktoren und Gleichzeitigkeitsfaktoren stellt ein Instrument zur Identifizierung möglicher Versorgungsinseln innerhalb eines Quartiers dar.

VII. Mit der Implementierung eines autoadaptiven Modells lässt sich eine breite Anwendbarkeit der Methodenkombination bei variierenden Randbedingungen und einer dynamischen Datengrundlage sicher stellen.

VIII. Schnittstellen zu bestehenden Software-Tools und Komponenten der Datenhaltung sowie des Datenaustauschs ermöglichen die Weiterentwicklung eines urbanen Energie-Informationssystems.

IX. Die Ergebnisse werden für operative und strategische Planungsaufgaben im Rahmen der Neu- und Weiterentwicklung von urbanen Energiesystemen leicht interpretierbar zur Verfügung gestellt.

1.4 Gliederung und struktureller Aufbau

Im Rahmen des anschließenden Kapitels 2 erfolgt die Abbildung von Methoden zur Quantifizierung von Gleichzeitigkeiten in der urbanen Lastverteilung. An dieser Stelle bildet die Ermittlung des thermischen Energiebedarfs die notwendige Informationsgrundlage, sodass in diesem Kapitel einleitend ein Überblick über dazugehörige Modelle und Methoden gegeben wird. Darauf aufbauend werden verschiedene Definitionen und Ansätze zur Berechnung von Gleichzeitigkeitsfaktoren vorgestellt. Kapitel 3 widmet sich der Thematik um Prognosen und Szenariotechniken vor dem Hintergrund der Identifizierung geeigneter Methoden für die Modellentwicklung. In Kapitel 4 werden Algorithmen für die Klassifikation energetischer Lastprofile eingeführt. Dadurch erfolgt die Adressierung des Ziels, Ähnlichkeiten in den Charakteristika thermischer Lastprofile automatisiert identifizieren zu können. Auf Grundlage der Aufarbeitung des aktuellen Stands der Technik und Wissenschaft zum thematischen Schwerpunkt und zu den Methoden zur Lösung der Problemstellung dieser Arbeit sind in Kapitel 5 die mathematischen und physikalischen Grundlagen für die Modellentwicklung und -evaluierung aufgeführt. Vor dem Hintergrund der Implementierung des Modells sind in Kapitel 6 eingangs die damit verbundenen Anforderungen auf Basis der zuvor erarbeiteten theoretischen Grundlagen formuliert. Daran anschließend wird das Softwarekonzept von *AMSA* inklusive der Benutzeroberfläche sowie die Umsetzung der Schnittstelle zur Datenbank und Webservern vorgestellt. Überdies erfolgt die Vorstellung vorgesehener Schnittstellen an bestehende Software-Tools und Applikationsentwicklungen. Zur Demonstration des entwickelten und implementierten Modells wird in Kapitel 7 exemplarisch ein Anwendungsfall analysiert. Als Grundlage dienen Datensätze eines Quartiers

mit Nichtwohngebäuden unterschiedlicher Nutzungsklassen, die mittels einer bestehenden Netzinfrastruktur mit Wärme und Kälte versorgt werden. Eingangs erfolgt neben der Vorstellung des Datensatzes und der Ergebnisse der Datenvorverarbeitung eine deskriptive Datenanalyse. Letztere dient der Einordnung von Ergebnissen der Modelldemonstration und ermöglicht bei zukünftig verbreiteter Anwendung von *AMSA* eine Vergleichbarkeit vor dem Hintergrund variierender Eingangsdaten. Im weiteren Verlauf des Kapitels werden die Ergebnisse der Vorstudien zu Teilmodellanwendungen analysiert, um die Gesamtmodellanalyse vorzubereiten. Letztere beinhaltet in der Folge die Ergebnisse der Quantifizierung historischer und zukünftiger Gleichzeitigkeitseffekte in der Lastverteilung des Quartiers. Im Rahmen der Schlussbetrachtung in Kapitel 8 erfolgt die Zusammenfassung der wichtigsten Ergebnisse dieser Arbeit im Zusammenhang mit der Bewertung der eingangs aufgestellten Hypothesen. Daran anschließend werden Potenziale des entwickelten Modells vorgestellt, deren Nutzung abschließend in einem Ausblick die Grundlage für die Formulierung des zukünftigen Forschungsbedarfs bilden.

KAPITEL 2

Gleichzeitigkeiten in der urbanen thermischen Lastverteilung

Der thermische Energiebedarf von Gebäuden wird durch zahlreiche Faktoren, wie beispielsweise die Baukonstruktion, die Technische Gebäudeausrüstung (TGA) oder das Nutzerverhalten beeinflusst (Magoulès et al., 2016). In diesem Zusammenhang können einige weitere Einflussfaktoren genannt werden, welche sich in diese Kategorien einordnen lassen. Auf urbaner Ebene gilt es neben diesen Effekten auf Einzelgebäudeebene auch Gleichzeitigkeiten in dem thermischen Leistungsabnahmeverhalten von Gebäudegruppen zu quantifizieren, um eine ökonomisch und ökologisch sinnvolle Dimensionierung von zentralen Erzeugungs- und Verteilungssystemen thermischer Energie zu ermöglichen (Winter et al., 2001). Die Grundlage zur Ermittlung des Gleichzeitigkeitsfaktors stellen jedoch die Lastprofile des thermischen Energiebedarfs der einzelnen Gebäude dar. Daran anknüpfend ist dieses Kapitel in die Abschnitte 2.1 *Ermittlung des thermischen Energiebedarfs urbaner Regionen* und 2.2 *Ansätze zur Quantifizierung des Gleichzeitigkeitsfaktors* untergliedert.

2.1 Ermittlung des thermischen Energiebedarfs urbaner Regionen

Dieser Abschnitt behandelt den theoretischen Hintergrund zur Ermittlung des thermischen Energiebedarfs urbaner Regionen und bildet den aktuellen Stand der Technik und Forschung ab. Auf Grundlage der herausgearbeiteten Modellcharakteristika und Ergebnissen anwendungsbasierter Studien wird die Wahl der Methoden begründet und implizite Anforderungen für die Teilmodellentwicklung hergeleitet und formuliert.

2.1.1 Theoretischer Hintergrund

Im Rahmen der Ermittlung des thermischen Energiebedarfs urbaner Regionen stehen sich mit einem Top-down- oder einem Bottom-up-Ansatz zwei verschiedene Ausgangspunkte für die urbane Energiebedarfsanalyse gegenüber (Reinhart et al., 2016). Ersterer umfasst die stadtgebietsbezogene Betrachtungs- und Analyseebene, welche sukzessive mittels Unterteilungen des Siedlungsraumes in eine kleinerskalige Ebene überführt wird (Howard et al., 2012). Ein Modellansatz für eine großskalige städtische Betrachtungsebene stellt beispielsweise die Siedlungstypmethode von Roth et al. dar, welche auf der Definition charakteristischer Siedlungsstrukturen und dem jeweils dafür abgeschätzten Wärmeleistungsbedarf beruht (Roth, 1980). Die Modellentwicklung mit einem Bottom-up-Ansatz

beginnt bei der Quantifizierung des Energiebedarfs von Einzelgebäuden. Letztere lassen sich wiederum zu Gebäudegruppen zusammenfassen, sodass auch die energetische Abbildung urbaner Räume mit diesem Ansatz möglich ist (Reinhart et al., 2016). Für die Untersuchung und Quantifizierung des Gleichzeitigkeitsfaktors wird der Bottom-up-Ansatz betrachtet, da die individuelle Charakteristik der einzelnen Gebäudelastprofile Grundlage für eine kleinteilige Analyseebene ist, welche eine möglichst genaue Ermittlung von Gleichzeitigkeitsfaktoren ermöglicht. Vordergründig geht es daher im Folgenden um die Ermittlung des thermischen Energiebedarfs einzelner Gebäude als Grundlage für die Betrachtungsebene von urbanen Räumen. Beginnend bei der Einzelgebäudeanalyse sind die folgenden beiden Modellkategorien zu unterscheiden:

I. Physikalische Modelle

II. Datenbasierte Modelle

Hinsichtlich der Zielstellungen in der Modellierung können ferner diagnostische von prognostischen Modellen unterschieden werden. In diesem Zusammenhang dienen diagnostische Modelle der Generierung von Wissen über Gesetzmäßigkeiten, wie beispielsweise Energiebilanzen oder Wärmedurchgangsprozesse, und prognostische Modelle der Vorhersage eines Systemverhaltens (Saltelli et al., 2008). Abbildung 2.1 visualisiert vereinfachend den Unterschied zwischen physikalischen und datenbasierten Modellen.

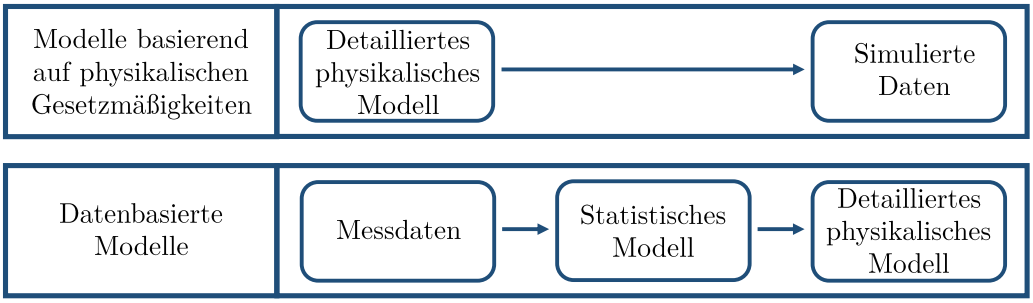


Abbildung 2.1: Schematische Darstellung der Unterschiede zwischen physikalischen und datenbasierten Modellen zur gebäudespezifischen Energieanalyse in Anlehnung an Coakley et al., 2014

Während bei physikalischen Modellen ein detailliertes Gebäudemodell zugrunde liegt, auf Basis dessen das energetische Verhalten des Gebäudes simuliert und dazugehörige Simulationsdaten ausgegeben werden, dienen bei datenbasierten Modellen Messdaten und ein statistisches Modell der Generierung von detaillierten physikalischen Modellen.

Physikalische Modelle

Diese Modellkategorie basiert auf physikalischen Zusammenhängen zur Quantifizierung der energetischen Zustände von Gebäuden, Gebäudeteilen oder einzelnen Komponenten

und ist seit über 50 Jahren Gegenstand zahlreicher Forschungsarbeiten (Magoulès et al., 2016). So lassen sich verschiedene Forschungsschwerpunkte im Kontext der energetischen Gebäudesimulation identifizieren, welche von Ansätzen der monatlichen oder jährlichen Energiebilanzen über die Simulation einzelner Gebäudekomponenten und Teile der Anlagentechnik bis hin zur umfassenden Lebenszyklusanalyse im Sinne der Bewertung von Ressourcenanforderungen und dem Primärenergieverbrauch von Gebäuden reichen. Mit energetischen Gebäudesimulationen können dabei mehrskalige Multiebenenprobleme modelliert und gelöst werden (van Treeck, 2004; van Treeck, 2010). Hinsichtlich der methodischen Ansätze lassen sich zwei verschiedene Kategorien unterscheiden:

i. Detaillierte ganzheitliche Methoden
ii. Vereinfachte Methoden

Die detaillierte Abbildung und Berechnung gebäudebezogener Energieströme bedarf einer umfangreichen Datengrundlage mit hoher gebäudespezifischer Informationsdichte, um physikalische Prozesse in der thermischen Dynamik eines Gebäudes im Detail modellieren und darauf aufbauend den Energiebedarf ermitteln zu können. Um die Umsetzung komplexer Berechnungsverfahren zur Abbildung physikalischer Prozesse in der thermisch-energetischen Gebäudeanalyse zu ermöglichen, wurden zahlreiche Software-Tools entwickelt, welche beispielsweise die Formulierung von gebäudebezogenen Energiestandards oder Energieanalysen unterstützen. In diesem Zusammenhang lassen exemplarisch *IDA Indoor Climate and Energy* (Björsell et al., 1999; Kalamees, 2004), *TRNSYS* (Klein et al., 2007) oder *EnergyPlus* (Crawley et al., 2001) aufführen (*Building Energy Software Tools Directory*). Weiterhin werden in der Forschung Modellbibliotheken wie *AixLib* (Müller et al., 2016) oder *BuildingsSystems* (Nytsch-Geusen et al., 2016) zur dynamischen Gebäudesimulation genutzt und weiterentwickelt (Nytsch-Geusen et al., 2014).

Im Bereich der deutschen Normen und Richtlinien ist zur Berechnung der Gebäudeheizlast die DIN EN 12831-1 anzuwenden, welche die Berechnungsvorschriften für die Ermittlung der benötigten Wärmezufuhr für die Beheizung von Einzelräumen, Gebäudeeinheiten und Gebäuden festlegt. Das detaillierte Standardverfahren erfordert umfangreiche Informationen über die thermischen Eigenschaften und die Gebäudegeometrie. Zur Bemessung der benötigten Wärmeerzeugerleistung für die Beheizung des Gebäudes werden Wärmeverluste infolge von Transmission und Lüftung sowie benötigte Aufheizleistungen den Wärmegewinnen im Gebäude gegenübergestellt (DIN EN 12831-1, 2017). Für die Kühllastberechnung steht die Richtlinie VDI 2078 zur Verfügung, welche die dafür benötigten Berechnungsvorschriften definiert. Als Basis dient ein detailliertes Mehrzonen-Gebäudemodell mit spezifischen anlagentechnischen und bauphysikalischen Eigenschaften sowie spezifischen Nutzungscharakteristika vor dem Hintergrund detailliert vorgegebener klimatischer Randbedingungen. Neben diesem umfangreichen Verfahren wird auch ein Abschätzverfahren für die raumweise Kühllastberechnung vorgeschlagen, welches ersteres jedoch nicht ersetzt (VDI 2078, 2015). Mit der DIN Norm 18599 steht im Rahmen der gebäudespezifischen Bewertung der Energieeffizienz weiterhin ein umfassendes Regelwerk zur statischen Ermittlung des Energiebedarfs für die Beheizung und Kühlung von Gebäuden, basierend auf detaillierten

raum- und zonenweisen Berechnungsverfahren, zur Verfügung (DIN V 18599-1, 2018). Für die Berechnung des Energiebedarfs zur Beheizung und Kühlung von Räumen und Gebäuden dient ferner die Richtlinie VDI 2067, welche auf dem Rechenkern der VDI 6007 beruht (VDI 2067 Blatt 10, 2013).

Detaillierte Modelle bedingen eine Datenbasis mit hoher gebäudespezifischer Informationsdichte, die selten gegeben ist, sodass sich im Laufe der Jahre auch vereinfachte Modellansätze etabliert und bewährt haben (Magoulès et al., 2016). Diese benötigen eine geringe Anzahl an Eingangsparametern und ermöglichen eine erste Einschätzung über den thermischen Energiebedarf eines Gebäudes. In diesem Zusammenhang lassen sich die folgenden Methoden unterscheiden (Al-Homoud, 2001).

i. Gradtagsmethode
ii. Modifizierte Gradtagsmethode
iii. Gradtagsmethode mit variabler Basis
iv. *Bin*-Methode
v. Modifizierte *Bin*-Methode

Die Gradtagsmethode stellt ein stationäres Berechnungsverfahren zur Approximation von Trends im gebäudespezifischen Energieverbrauch dar. Hierbei wird die Annahme getroffen, dass die Wärmeverluste im langjährigen Mittel durch solare und interne Energiegewinne im Gebäude ausgeglichen werden, sofern die mittlere tägliche Außenlufttemperatur 18,3 °C beträgt. Die Abschätzung des Energiebedarfs für die Beheizung des Gebäudes erfolgt über die Höhe des dazugehörigen Wärmeverlustkoeffizienten oder der Normheizlast in Abhängigkeit der berechneten Heizgradtage an dem spezifischen Standort (Al-Homoud, 2001). Die auf 18,3 °C fixierte Innenlufttemperatur wird bei der modifizierten Gradtagsmethode durch einen empirischen Korrekturfaktor angepasst. Die Gradtagsmethode und die modifizierte Gradtagsmethode bringen Nachteile mit sich. So lassen sich diese beiden Methoden beispielsweise nur für die Ermittlung der Heizlast verwenden. Weiterhin bedingen die vorgenommenen Annahmen und Vereinfachungen neben der unpräzisen Betrachtung von internen Gewinnen und verbesserten Dämmstandards auch Beschränkungen hinsichtlich der Anwendbarkeit für Nicht-Wohngebäude, deren Wärmebedarf nicht primär auf dem Ausgleich von Transmissions- und Lüftungswärmeverlusten beruht. Die Gradtagsmethode mit variabler Basis löst die Abhängigkeit von der Basistemperatur von 18,3 °C auf, indem eine gebäudespezifische Gleichgewichtstemperatur an Stelle der fixierten Basistemperatur eingeführt wird, welche einer Außentemperatur entspricht, die keine Beheizung oder Kühlung des Gebäudes erfordert. Damit wird unterschiedlicher klimatischer Bedingungen, Nutzungscharakteristika oder variierender Dämmstandards Rechnung getragen. Zudem lässt dich auch der witterungsinduzierte Kühlbedarf über die Ermittlung der Kühlgradtage abschätzen. Die *Bin*-Methode basiert auf der Heiz- und Kühllastberechnung zu verschiedenen Außentemperaturbedingungen (*Bins*). Anschließend werden die Ergebnisse mit der Länge des Zeitraums, in welcher diese Bedingungen auftreten, multipliziert. Mit der modifizierten *Bin*-Methode wurde dieser modelltheoretische Ansatz weiter verfeinert, indem

gemittelte Profile von solaren Wärmegewinnen in Gebäuden oder interne Nutzungsprofile integriert wurden. Die Kombination aus den verschiedenen Einflussparametern ermöglicht damit noch detailliertere Abschätzungen des thermischen Energiebedarfs von Gebäuden. (Al-Homoud, 2001)

Während sich die drei Gradtagsmethoden insbesondere für kleine Gebäude eignen, bei denen der Energieverbrauch größtenteils durch Energieverluste durch die Gebäudehülle verursacht werden, werden die *Bin*-Methoden für große Gebäude mit dominierenden internen Lasten angewendet (Magoulès et al., 2016). In den deutschen Normen findet sich die Gradtagsmethode in dem Beiblatt 2 der DIN EN 12831 zur Heizlastberechnung von Gebäuden wieder, welche vereinfachte Verfahren zur Heizlastberechnung beinhaltet. Letztere sind für energetische Bewertungen im Gebäudebestand anzuwenden, sofern keine ausreichende Datenbasis für detaillierte Heizlastberechnungen vorliegt (DIN 12831 Beiblatt 2, 2012). In umgekehrter Analogie zur Gradtagsmethode wird im Rahmen des Verbrauchsverfahrens die Heizlast eines Gebäudes auf Grundlage des jährlichen Endenergieverbrauchs für die Beheizung des Gebäudes abgeschätzt. Alternativ zur Berechnung der Gradtage kann anhand dieser Norm auch der mittlere Belastungsgrad des Wärmeerzeugers verwendet werden. Der Endenergieverbrauch für die Trinkwassererwärmung als Basis für die dazugehörige benötigte Erzeugerleistung ist entweder über die Flächenbelegung pro Person im Gebäude oder mittels des nutzerbezogenen Trinkwarmwasserbedarfs auf Basis empirischer Formeln zu ermitteln. Als weiteres vereinfachtes Verfahren zur Heizlastberechnung wird neben den Verbrauchsverfahren in dem Beiblatt das Hüllflächenverfahren vorgestellt, welches ein Ein-Zonen-Gebäudemodell als Bilanzierungsebene in den Mittelpunkt rückt. Die Gebäudeheizlast wird dabei über die Summe der Transmissions- und Lüftungswärmeverluste über die nach außen gerichtete thermische Hülle des Gebäudes approximiert. (DIN 12831 Beiblatt 2, 2012)

Datenbasierte Modelle

Alternativ zu physikalischen Modellen finden datenbasierte Modelle zur Analyse des gebäudebezogenen Energieverbrauchs Anwendung. In diesem Zusammenhang sind Black-Box und Grey-Box Methoden sowie Modellkalibrierungsmethoden voneinander zu unterscheiden. Black-Box Modelle basieren auf mathematischen oder statistischen Modellen, mit welchen ein Zusammenhang zwischen Einflussvariablen und gemessenem Zielgrößen hergestellt wird (Coakley et al., 2014). Dazu gehört die Regressionsanalyse, welche beispielsweise für die Korrelation zwischen Energieverbrauch und Klimavariablen eingesetzt wird (Magoulès et al., 2016) oder Künstliche Neuronale Netze (Coakley et al., 2014). Grey-Box Methoden unterscheiden sich von Black-Box Methoden hinsichtlich der Eingangsgrößen. Anders als Black-Box Methoden, benötigen Grey-Box Methoden aufgrund unvollständiger oder unsicherer Datensätze Systemparameter eines physikalischen Modells zur energetischen Analyse eines Gebäudes (Coakley et al., 2014; Magoulès et al., 2016). Modellkalibrierungsmethoden beruhen auf dem Ziel die Eingangsparameter eines vollständig beschriebenen Modells basierend auf physikalischen Gesetzmäßigkeiten mittels Messdaten als Zielgrößen zu opti-

mieren (Coakley et al., 2014). In der Literatur finden sich zudem verschiedene Modelle der vereinfachten Korrelationsanalyse wieder, welche nachfolgend beispielhaft aufgeführt sind (Bauer et al., 1998).

i. Energiesignatur-Methode
ii. *SLR*-Methode
iii. *H-m*-Methode
iv. *Eta*-Methode

Mit der Energiesignatur-Methode werden die Veränderungen im gebäudespezifischen Energieverbrauch berechnet. Es fehlt jedoch bei dieser Methode die Berücksichtigung von solaren Gewinnen (Bauer et al., 1998). Im Rahmen der *SLR*-Methode werden die Energiegewinne eines solaren Passivhauses ermittelt, welche sich in Relation zu einem Referenzgebäude ohne solaren Passivhausstandard ergeben (Bauer et al., 1998). Die *Eta*-Methode stellt eine Verfeinerung der *H-m*-Methode dar und ist für verschiedene Anwendungszwecke, wie dem Vergleich des Energieverbrauchs verschiedener Gebäude oder der Prognose des Energieverbrauchs bei variierenden klimatischen Randbedingungen geeignet. Sie basiert auf der Definition einer Heiz- und Kühlperiode und setzt mittels einer Regression einen modifizierten Leistungsbereitstellungsfaktor mit den internen und solaren Gewinnen ins Verhältnis, wobei die Steigung der Regressionsgerade den Nutzungsgrad der internen und solaren Gewinne impliziert. (Bauer et al., 1998)

In den deutschen Normen und Richtlinien finden sich datenbasierte Methoden zur Ermittlung des Wärmeverbrauchs und der Heizlast in der Richtlinie VDI 3807 bzw. dem Beiblatt 2 der Norm DIN EN 12831 wieder. Im Rahmen der VDI 3807 wurden unter anderem Energieverbrauchskennwerte für verschiedene Gebäudearten und Liegenschaften nach den anerkannten Regeln der Technik auf Grundlage statistischer Erhebungen zu Messwerten gebildet (VDI 3807 Blatt 1, 2013). Als Grundlage für die Ermittlung der Kennwerte dienten Energieverbrauchswerte, welche im Rahmen einer Regressionsanalyse mit der Außentemperatur in einen Zusammenhang gesetzt wurden. Ferner beschreibt die Richtlinie das allgemeine Vorgehen diesbezüglich für die eigene Kennwertberechnung. Für die vereinfachte Heizlastberechnung auf Grundlage von gemessenen Verbrauchsdaten sind nach der DIN EN 12831 Beiblatt 2 die Wärmeleistung zusammen mit der simultan erfassten Außenlufttemperatur aufzutragen, wobei die Regressionsgerade den Wärmetransferkoeffizienten beschreibt, auf Grundlage dessen die anzusetzende Heizlast zu berechnen ist (DIN 12831 Beiblatt 2, 2012).

2.1.2 Studien

Der thermische Energiebedarf von urbanen Regionen wird maßgeblich von den dazugehörigen Einzelgebäuden und ihren wechselseitigen thermischen Beeinflussungen geprägt. So untersuchten Liu et al. beispielsweise den Effekt der gegenseitigen thermischen Beeinflussung von Gebäuden urbaner Regionen im Hinblick auf den Energieverbrauch (Liu

et al., 2015a). In diesem Zusammenhang zeigten Bueno et al. den Einfluss von städtischen Wärmeinseln auf den Energieverbrauch von Einzelgebäuden (Bueno et al., 2012), wobei Mavrogianni et al. diesen Effekt am Beispiel der Stadt London demonstrierten (Mavrogianni et al., 2009). Um detaillierte physikalische Modelle infolge von Informationsknappheit auf urbaner Ebene vereinfachen zu können, sind die Anwendung und Validierung reduzierter Modelle Gegenstand zahlreicher Forschungsarbeiten (Reinhart et al., 2016). Die urbane Energieanalyse betrifft darüber hinaus jedoch noch eine Vielzahl von anderen Forschungsschwerpunkten. Keirstead et al. haben dazu in ihrer umfangreichen Literaturrecherche fünf Schlüsselbereiche identifiziert: Auslegung der Technologie, Gebäudekonzepte, Stadtklima, Systemkonzepte und politische Evaluierung. Ferner wurden mit dem Verständnis für Modellkomplexitäten, der Datenqualität und -unsicherheit, der Modellintegration sowie strategisch-politischen Relevanz vier Kernherausforderungen für diese Schlüsselbereiche herausgearbeitet (Keirstead et al., 2012).

In Bezug auf Studien im Bereich der gebäudebezogenen Energieanalyse sind verschiedene Forschungsrichtungen zur Verbesserung der Methoden für die Quantifizierung des gebäudebezogenen Energiebedarfs voneinander abzugrenzen. So steht beispielsweise die Identifizierung der Schlüsselvariablen für gebäudebezogene Energieanalysen im Rahmen von Sensitivitätsanalysen im Mittelpunkt zahlreicher Forschungsarbeiten (Tian, 2013). Coakley et al. geben in ihrer Studie einen Überblick über Forschungsarbeiten, die sich bisher mit dem Vergleich von energetischen Gebäudesimulationen und Messdaten fokussierten (Coakley et al., 2014). Dagegen zeigt Nguyen et al. die aktuellen Entwicklungen im Bereich der simulationsbasierten Optimierungsmethoden im Rahmen der energetischen Gebäudesimulation (Nguyen et al., 2014) auf.

Im Forschungsfeld der datenbasierten Modelle haben Pfafferott et al. auf Basis von Monitoringdaten die Energiebilanz eines Büroraumes über die Korrelation zwischen Innen- und Außenlufttemperatur bestimmt (Pfafferott et al., 2005). Darüber hinaus sind weitere vereinfachte statistische Methoden Gegenstand von Untersuchungen zu Modellkombinationen und -erweiterungen (Westergren et al., 1999; Bauer et al., 1998). Dhar et al. haben ein Modell auf Grundlage einer Fourierreihenentwicklung eingeführt, um die nichtlinearen Heiz- und Kühllasten quantifizieren zu können (Dhar et al., 1998; Dhar et al., 1999). Mit Hilfe eines Autoregressiven Modells mit zusätzlichen exogenen Einflussparametern analysierten Jiménez & Heras verschiedene Gebäudekomponenten (Jiménez et al., 2005). Hinsichtlich der Regressionsmethoden auf Basis unterschiedlicher Zeitskalen haben Studien die variierende Ergebnisgüte gezeigt (Cho et al., 2004). Im Bereich der Grey-Box und Black-Box Modelle wurden in den letzten zwei Dekaden zahlreiche Studien im Kontext der Prognose von Energieverbräuchen, der Betriebsoptimierung und Regelung von Komponenten der Anlagentechnik sowie der Ermittlung von Gebäudenutzungsparametern untersucht (Magoulès et al., 2016).

Im Hinblick auf die Frage nach dem am besten geeigneten Modelltyps für die Ermittlung des gebäudebezogenen thermischen Energiebedarfs gibt es keine allgemein gültige Antwort

auf Grundlage anwendungsbasierter Studien. Vielmehr beruht die Entscheidung für einen Modelltyp auf den gegebenen Randbedingungen, wie beispielsweise der vorhandenen Datengrundlage oder Restriktionen, die eine Begrenzung der Rechenzeiten bedingen (Magoulès et al., 2016). Beide Modellkategorien lassen sich jedoch auch miteinander kombinieren, wobei mit Künstlichen Neuronalen Netzen beispielsweise Gebäudeparameter abgeschätzt werden können (Olofsson et al., 2002; Lundin et al., 2004).

2.1.3 Implikationen und Anforderungen an die Modellierung

Für die Ermittlung des Energiebedarfs von Gebäudegruppen und urbanen Regionen unter Berücksichtigung von Gleichzeitigkeitseffekten stehen physikalische und datenbasierte Ansätze zur Verfügung. Detaillierte physikalische Modelle benötigen eine umfassende Datenbasis, welche in der Regel auf urbaner Ebene nicht vorhanden ist (Reinhart et al., 2016). Vereinfachte physikalische Modelle dienen der Abschätzung des Energiebedarfs von Gebäuden und ermöglichen bei Informationsknappheit eine vereinfachte Ermittlung des gebäudespezifischen Energiebedarfs, weisen jedoch keine universellen Anwendungseigenschaften auf. Datenbasierte Modelle erfordern wenige Eingangsparameter, gleichwohl müssen diese in Form von Messdaten vorliegen. In dieser Modellkategorie sind vornehmlich Black-Box und Grey-Box Modelle zu nennen, wobei Grey-Box Modelle im Gegensatz zu Black-Box Modellen wiederum teilweise auf physikalischen Parametern der Gebäude basieren.

Eine verstärkte Einführung von intelligenten Zählersystemen im Rahmen des Aufbaus smarter integrierter Energiesysteme (Lund et al., 2014) und der stetig steigenden Menge an Messdaten im Energiesektor (Zhou et al., 2016) begründet die aktuelle und zukünftige hohe Relevanz von datenbasierten Modellen und ermöglicht den weiteren zukünftigen Einsatz des in dieser Arbeit entwickelten Modells zur Modellierung des thermischen Energiebedarfs urbaner Regionen. Sofern detaillierte Gebäudedaten, beispielsweise für einen Neubau vorliegen, lässt sich dieser simulieren und die generierten Datensätze als Eingangsparameter für das datenbasierte Modell nutzen. Gleichzeitig können bei Informationsknappheit vereinfachte Modelle verwendet werden, um den gebäudespezifischen Energiebedarf abschätzen zu können, um diesen als Teil der Datengrundlage für das datenbasierte Modell zu nutzen. Auf Grundlage dieser Implikationen für die Modellierung bedarf es der Entwicklung eines datenbasierten Modells für die Ermittlung des gebäudespezifischen thermischen Energiebedarfs, der Zusammenführung der gebäudespezifischen Black-Box-Modelle zur Ermittlung des Energiebedarfs von Gebäudegruppen und urbanen Regionen sowie der Berücksichtigung von Schnittstellen für die Angliederung physikalischer Modelle.

2.2 Ansätze zur Quantifizierung des Gleichzeitigkeitsfaktors

Dieser Abschnitt knüpft inhaltlich an das Thema des thermischen Energiebedarfs urbaner Regionen an und behandelt den theoretischen Hintergrund zu Ansätzen der Definition und Ermittlung von Gleichzeitigkeitsfaktoren in der thermischen Lastabnahme von Gebäuden innerhalb von Versorgungsgruppen und bildet den aktuellen Stand der Technik und Forschung ab. Auf Grundlage der vorhandenen modelltheoretischen Ansätze erfolgt die Ableitung des Forschungsbedarfs in diesem Themengebiet, auf dessen Grundlage implizite Anforderungen für die Teilmodellentwicklung hergeleitet und formuliert werden. Im Hinblick auf die Vereinheitlichung der Nomenklatur erfolgt eine Übertragung der verwendeten Symbole aus den verschiedenen Forschungsarbeiten zu diesem Thema auf die Symbolbezeichnungen in dieser Arbeit.

2.2.1 Theoretischer Hintergrund

Die Diversität der Lastprofile in urbanen Räumen lässt sich im Kontext zweier Komponenten des Wärmebedarfs betrachten. Der Wärmebedarf kann nach Frederiksen et al. im Wesentlichen auf eine physikalische und eine soziale Komponente zurück geführt werden (Frederiksen et al., 2014). Die physikalische Komponente beschreibt dabei beispielsweise den transmissionsinduzierten Wärmebedarf, welcher von saisonalen Schwankungen der Außentemperatur abhängt. Die soziale Komponente spiegelt wiederum das menschliche Verhalten wieder, im Sinne von Anwesenheitszeiten oder Lüftungsverhalten. Die Diversität des sozial bedingten Wärmebedarfs ist aufgrund der hohen Individualität des Nutzerverhaltens als hoch einzustufen. Im Gegensatz dazu sind die Wärmebedarfsprofile auf Grundlage ausschließlich physikalischer Betrachtungen mit einer hohen Synchronität anzunehmen. (Frederiksen et al., 2014) Unterschiede können allerdings bei stark voneinander abweichenden Dämmstandards der Gebäude und damit einhergehend einer unterschiedlichen thermischen Trägheit entstehen (Rehau, 2012). Der Gleichzeitigkeitsfaktor ϕ wird in der Regel als abgenommene Leistung, aufkummuliert über alle N Abnehmer, zum Zeitpunkt der maximalen Leistungsanforderung $\dot{Q}\left(\tau_{max}\right)$ dieser Abnehmergruppe im Verhältnis zu der Summe aller Nennleistungen der Abnehmer $\dot{Q}_{Nenn}$ nach Gleichung (2.1) definiert (Winter et al., 2001).

$$\phi = \frac{\sum_{i=1}^{N} \dot{Q}_i\left(\tau_{max}\right)}{\sum_{i=1}^{N} \dot{Q}_{Nenn,i}} \tag{2.1}$$

Es existieren ferner jedoch auch andere Ansätze zur Quantifizierung des Gleichzeitigkeitsfaktors. So approximiert Glück den auf die benötigte Leistung für Raumwärme bezogenen Gleichzeitigkeitsfaktor für Wohngebäude über das Verhältnis zwischen der Summe der partiellen Gleichzeitigkeitsfaktoren aller eingeschalteter Heizkörpergruppen $\phi_{partiell}$ und

der Heizkörpergruppenanzahl N. Eine Heizkörpergruppe i besteht dabei aus Heizkörpern gleicher Leistung. (Glück, 1985)

$$\phi_{RW} = \frac{\sum_{i=1}^{N} \phi_{partiell,i}}{N} \tag{2.2}$$

Die partiellen Gleichzeitigkeitsfaktoren sind von den Einschaltwahrscheinlichkeiten der Heizkörper abhängig, welche abgeschätzt werden müssen, um die entsprechenden tabellarisch aufgeführten Kennwerte ablesen zu können. In seinen Ausführungen grenzt Glück Industrie- und Gesellschaftsgebäude von Wohngebäuden ab und weist den Nichtwohngebäuden aufgrund ihrer besonderen nutzungsspezifischen Charakteristika eine Sonderstellung zu, die es nicht erlaubt, pauschale Werte anzunehmen. (Glück, 1985)

In den deutschen Normen und Richtlinien findet der Gleichzeitigkeitsfaktor Eingang in die technischen Regeln für Trinkwasser-Installationen nach DIN 1988. In dieser Norm wird auf den Einfluss der Gleichzeitigkeit auf den anzusetzenden Spitzendurchfluss für die Dimensionierung von Trinkwasserleitungen in Abhängigkeit der Gebäudenutzung hingewiesen. Die Höhe des anzusetzenden Gleichzeitigkeitsfaktors wird jedoch nicht definiert, sondern ist im Einzelfall mit dem Betreiber festzulegen (DIN 1988-300, 2012). In dem vereinfachten Verfahren zur Bestimmung des Rohrinnendurchmessers von Trinkwasserrohrleitungen wird die Gleichzeitigkeit direkt in vorgegebenen Werten für den Spitzendurchfluss berücksichtigt (DIN EN 806-3, 2006). Die VDI-Richtlinie 2072 (VDI 2072, 2018) definiert einen Gleichzeitigkeitsfaktor für die Trinkwarmwasserversorgung. Ferner wurde von Sander in diesem Zusammenhang eine Kennlinie hergeleitet (Schramek et al., 2007), welche zusammen mit jener aus der VDI Richtlinie in Abbildung 2.2 dargestellt ist.

Der Gleichzeitigkeitsfaktor ϕ_{TWW} nach VDI 2072 ist in Abhängigkeit von der Anzahl an Wohneinheiten N entsprechend der empirischen Gleichung (2.3) definiert.

$$\phi_{TWW} = 0{,}03 + \frac{0{,}5}{\sqrt{N}} + 0{,}47 \cdot \frac{1}{N} \tag{2.3}$$

Sofern in dem System zur Trinkwassererwärmung eine Vorrangschaltung vorgesehen ist, wird der Gleichzeitigkeitsfaktor für die Raumwärme ϕ_{RW} nach Gleichung (2.4) berechnet.

$$\phi_{RW} = 1 - \phi_{TWW} \tag{2.4}$$

Neben diesen Definitionen wurden in anwendungsbasierten Studien empirische Gleichungen hergeleitet, welche den Gleichzeitigkeitsfaktor in Abhängigkeit von der Anzahl an Wärme-

abnehmern in zentralen Wärmeversorgungssystemen beschreiben. Diese werden im Rahmen der folgenden Abschnitte vorgestellt.

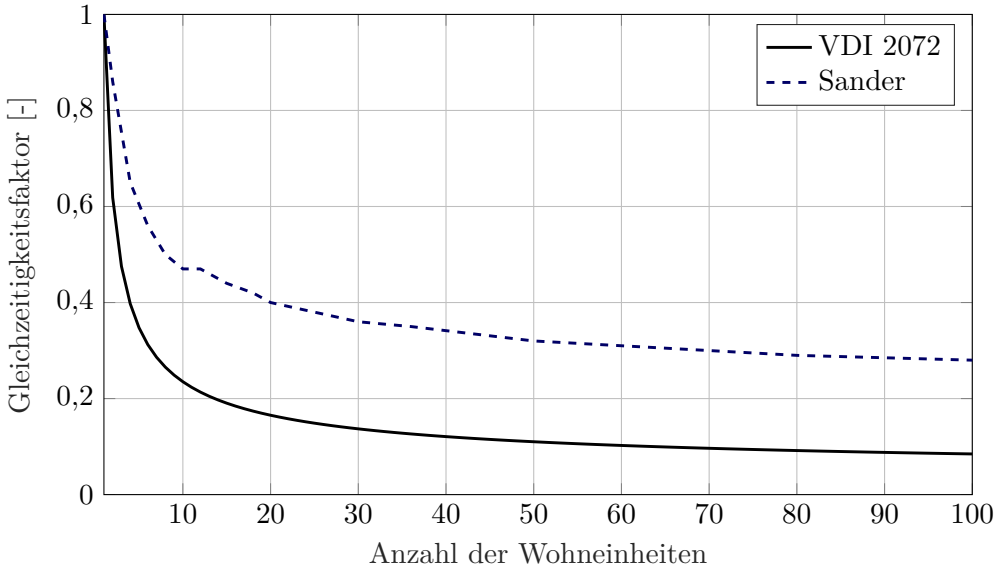


Abbildung 2.2: Vergleich der Kennlinien zum Gleichzeitigkeitsfaktor nach VDI 2072 und Sander in Abhängigkeit von der Abnehmeranzahl

2.2.2 Studien

Die Quantifizierung des Gleichzeitigkeitsfaktors beruht in der Regel auf betriebsbasierten Erfahrungswerten (Zhen et al., 2007). In diesem Zusammenhang wurden empirische Formeln, Kennlinien oder Kennwerte hergeleitet, welche selten im Mittelpunkt der Forschungsarbeiten stehen, jedoch eine große Relevanz für die Dimensionierung thermischer Energiesysteme aufweisen. Im Rahmen ihrer Analyse der Gleichzeitigkeit in österreichischen Nahwärmenetzen leiteten Winter et al. eine Näherungsfunktion zur Bestimmung des Gleichzeitigkeitsfaktors für Raumwärme ϕ_{RW} in Abhängigkeit der Abnehmerzahl N in einem Wärmenetz auf Grundlage von Gleichung (2.1) her (Winter et al., 2001).

$$\phi_{RW} = 0{,}45 + \frac{0{,}55}{1 + \left(\frac{N}{53{,}84}\right)^{1{,}76}} \tag{2.5}$$

In diesem Fall ist die Anzahl der Abnehmer als Eingangsparameter erforderlich, sodass sich die Komplexität der Abschätzung des Einflusses der Gleichzeitigkeit stark reduziert. Als Datenbasis für die Untersuchungen an beiden Nahwärmenetzen dienten stündliche Zählerdaten und Außentemperaturen, wobei jeweils circa 300 Abnehmer in die Untersuchung mit einbezogen wurden (Winter et al., 2001). Die Ermittlung der gebäudespezifischen

Nennleistungen der Wärmeabnehmer erfolgte mittels linearer Korrelation zwischen den aufgezeichneten mittleren stündlichen Leistungen und der mittleren stündlichen Außentemperatur zum jeweiligen Zeitpunkt. Über die Nutzung der Gebäude oder ihre Größe werden im einzelnen keine Angaben gemacht. Die Autoren widmen sich auch der Problematik verschiedener Nennleistungsklassen innerhalb von Abnehmergruppen, welche in bestimmten Fällen zu einer Unterdimensionierung des Gleichzeitigkeitsfaktors führt. So resultiert bei einer großen Anzahl an Kleinabnehmern und einem Großabnehmer eine Fehldimensionierung des Gleichzeitigkeitsfaktors, wodurch ein Versorgungsengpass möglich sei. Um dieses Risiko zu minimieren, schlagen die Autoren eine Aufteilung der gesamten Abnehmergruppe in Leistungsklassen vor, wobei das Verhältnis zwischen der Nennleistung des kleinsten und jener des größten Abnehmers maximal zwei betragen darf. Aufgrund dessen ist Gleichung 2.5 nur für Abnehmergruppen mit annähernd gleichen gebäudespezifischen Nennleistungen gültig. (Winter et al., 2001) Um diese Voraussetzung zu schaffen, wurde ein Algorithmus entwickelt, auf Basis dessen mittels Dominanzanalysen einzelner Leistungsabnehmer hinsichtlich ihrer Nennleistungen im Wärmenetz minimale und maximale Gleichzeitigkeitsfaktoren für verschiedene Abnehmergruppen bestimmt werden können (Gaderer, 2008).

In Studien von Tol und Svendsen (Tol et al., 2012b; Tol et al., 2012a) zu Niedertemperaturwärmenetzen für die Versorgung von Niedrigenergiehäusern in Dänemark wird ein empirisch ermittelter Gleichzeitigkeitsfaktor nach Gleichung (2.6) verwendet (Tol et al., 2012b).

$$\phi_{RW}\left(N\right) = 0{,}62 + \frac{0{,}38}{N} \tag{2.6}$$

Auch bei diesem Ansatz ist der Gleichzeitigkeitsfaktor ausschließlich von der kumulierten Anzahl der Wärmeabnehmer abhängig. Der für die Ermittlung der Heizlast infolge der Trinkwassererwärmung benötigte Gleichzeitigkeitsfaktor wird hierbei direkt in die empirische Gleichung zur Berechnung der Heizlast für die Trinkwassererwärmung integriert.

$$\dot{Q}_{TWW} = C_1 \cdot N + C_2 \cdot N^{0{,}5} + C_3 \tag{2.7}$$

Hierbei variieren die anzusetzenden Konstanten C_1, C_2 und C_3 in Abhängigkeit von der Kombination der Wärmeübergabestation mit einem vorgesehenen Pufferspeicher oder Wärmetauscher (Tol et al., 2012b). Abbildung 2.3 stellt den approximierten Verlauf der Kennlinien für Raumwärme und Trinkwarmwasser dar. Eine Gegenüberstellung der Kennlinien von Winter et al. (Winter et al., 2001) und Tol et al. (Tol et al., 2012b; Tol et al., 2012a) zeigt differierende Gleichzeitigkeitsfaktoren für die selbe Anzahl an Wärmeabnehmern im Netz. Abbildung 2.4 stellt in diesem Zusammenhang die Kennlinien für den auf die Raumwärme bezogenen Gleichzeitigkeitsfaktor für 300 Gebäude dar.

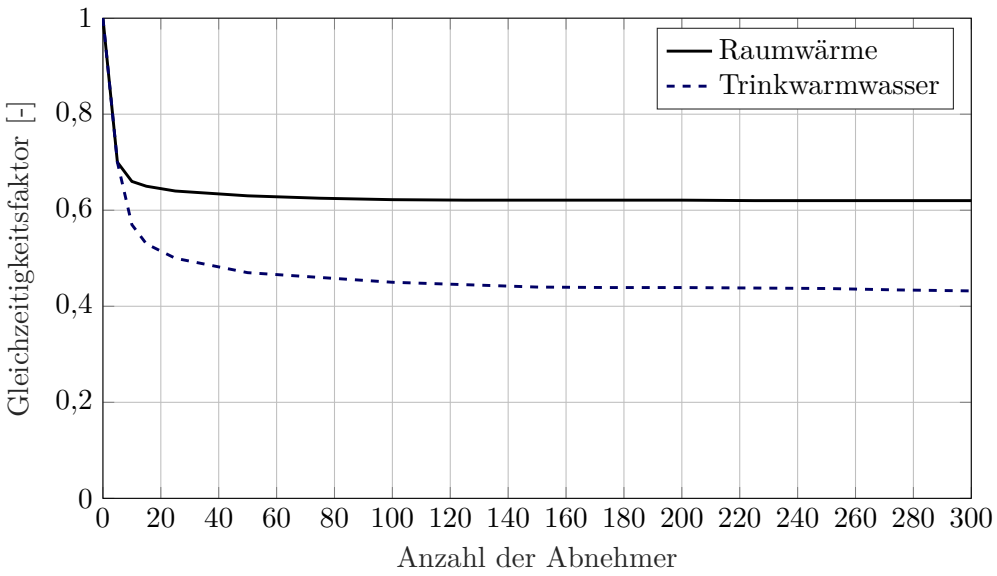


Abbildung 2.3: Kennlinien zur Änderung des Gleichzeitigkeitsfaktors für Raumwärme und Trinkwarmwasser mit zunehmender Abnehmerzahl in Anlehung an Tol et al, 2012

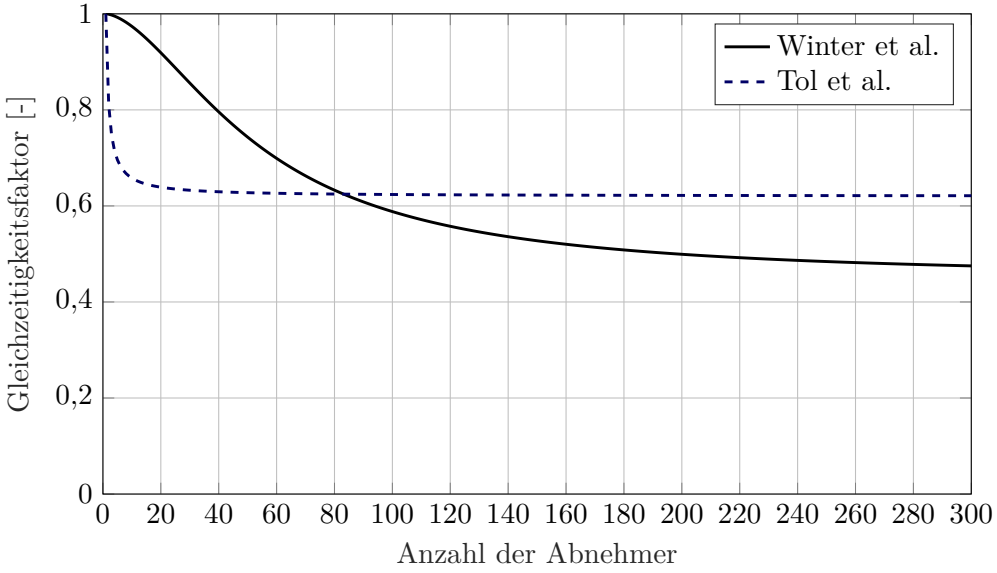


Abbildung 2.4: Vergleich der Kennlinien zum Gleichzeitigkeitsfaktor für Raumwärme von Winter et al. und Tol et al. in Abhängigkeit von der Abnehmeranzahl

Es zeigt sich, dass der Gleichzeitigkeitsfaktor für eine Gebäudegruppe mit ungefähr 75 Wärmeabnehmern sowohl nach Winter et al. als auch nach Tol et al. annähernd einen Wert von 0.62 erreicht. Demgegenüber wird von Tol et al. für eine Gebäudegruppe von ungefähr 20 Gebäuden ein Gleichzeitigkeitsfaktor von circa 0,64 vorgeschlagen, während nach Winter et al. ein Gleichzeitigkeitsfaktor von 0,95 vorgegeben wird. Für größere Abnehmergruppen

mit mehr als 200 Gebäuden differieren die Gleichzeitigkeitsfaktoren nach Winter et al. und Tol et al. um circa 0,1, wobei der degressive Verlauf der Kennlinie nach Winter et al. und der nahezu konstante Verlauf der Kennlinie nach Tol et al. ab ungefähr 100 Abnehmern darauf hinweist, dass sich die Differenz zwischen den Werten der beiden Kennlinien für große Gebäudegruppen weiter erhöht.

In der Fallstudie von Guan et al. (Guan et al., 2016) wird eine Analyse des Gleichzeitigkeitsfaktors für Elektrizität, Wärme und Wassernutzung im Rahmen der energetischen Planung eines Universitätscampus in Norwegen durchgeführt. Dabei liegt der Fokus auf der Quantifizierung von Gleichzeitigkeitsfaktoren auf Quartiersebene, der Ermittlung von Anteilen einzelner Gebäude an der Spitzenlast der gesamten Gebäudegruppe sowie auf einer Clusteranalyse. Letztere dient der Gruppierung von Gebäuden anhand verschiedener Kennwerten, wie den zuvor genannten, um Lastverschiebungspotential einzelner Gruppen zu identifizieren. Der Gleichzeitigkeitsfaktor ϕ ist dabei nach Gleichung (2.8) definiert (Guan et al., 2016).

$$\phi = \frac{P_{ges,max}}{\sum_{i=1}^{N} P_{i,max}} \tag{2.8}$$

Diese Definition beruht auf dem Verhältnis aus der maximal benötigten elektrischen Leistung P eines Gebäudes i zu der benötigten Leistung des Campus mit N Gebäuden und gibt den prozentualen Anteil der maximalen Leistungsabnahme eines Gebäudes an jener der gesamten Abnehmergruppe an (Guan et al., 2016). Weissmann et al. untersuchten den Einfluss gebäude- und nutzerbezogener Charakteristika auf den Gleichzeitigkeitsfaktor eines Wohngebietes und definierten dazu den Peak Load Ratio (PLR) gemäß Gleichung (2.9), wobei $Q_{ges,Nenn}$ die Summe der Nennleistung der einzelnen Abnehmer und $Q_{ZE,max}$ die zentral nachgefragte Wärmeleistung darstellt (Weissmann et al., 2017).

$$PLR = \frac{Q_{ges,Nenn} - Q_{ZE,max}}{Q_{ges,Nenn}} \tag{2.9}$$

Frederiksen und Werner (Frederiksen et al., 2014) schlagen zwei verschiedene Kennlinien für den Gleichzeitigkeitsfaktor vor, deren Grundlage die Europäische Norm (EN) 806-3:2006 und die schwedische Richtlinie F:101 der Swedish District Heating Association (SDHA) (Swedish District Heating Association, 2004) bilden. Abbildung 2.5 visualisiert den Verlauf der beiden Kennlinien. Beide Kennlinien weisen einen ähnlichen funktionalen Verlauf auf, wobei die Höhe des Gleichzeitigkeitsfaktors speziell für den Bereich von 5 bis 50 Energieabnehmern ungefähr um 0,1 differiert.

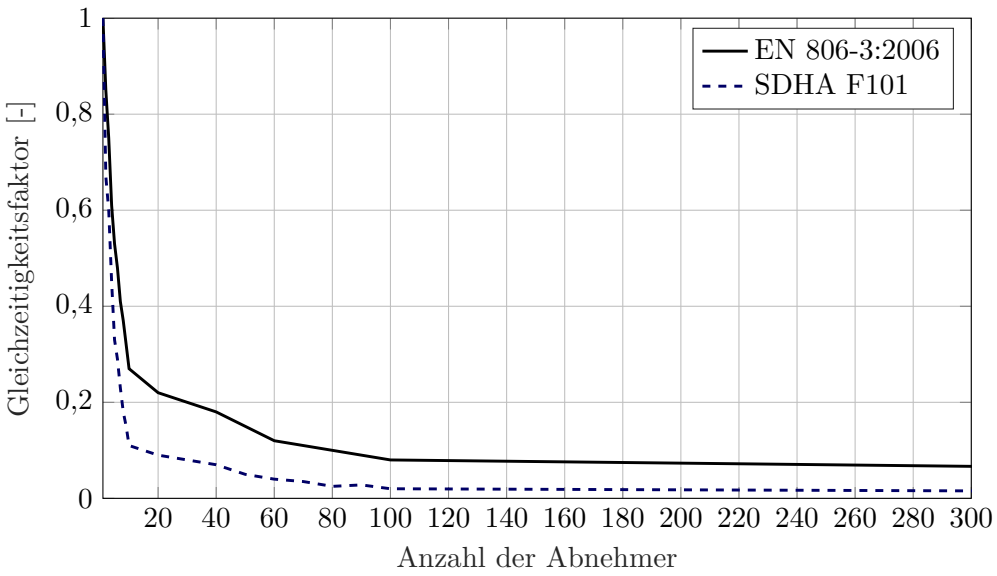


Abbildung 2.5: Vergleich norm- und richtlinienbasierter Kennlinien zu Gleichzeitigkeitsfaktoren in Anlehnung an Frederiksen et al. 2014

Zhang et al. untersuchten den Einsatz eines zentralen Wärmeerzeugungs- und Wärmeverteilungssystems zur Versorgung eines mehrstöckigen Gebäudes in China und verglichen im Rahmen dessen verschiedene Kennlinien für Trinkwarmwasser aus Dänemark, Schweden, Deutschland und China (Zhang et al., 2016). Letztere basieren auf der Anzahl an Wohnungen. Während sich die Kennlinien aus Dänemark, Schweden und Deutschland in ihrem Verlauf ähnlich sind, wird in China ein nahezu konstanter Gleichzeitigkeitsfaktor von 0,2 in diesem Kontext angesetzt. Abbildung 2.6 visualisiert die Verläufe der Kennlinien aus Dänemark, Schweden und Deutschland. Die Unterschiede zwischen den europäischen Kennlinien und dem Ansatz aus China liegen in der unterschiedlichen Berechnungsgrundlage begründet. In China erfolgt die Quantifizierung der Gleichzeitigkeit in der thermischen Lastabnahme mittels des Kehrwerts des stündlichen Variationskoeffizienten V_h nach Gleichung (2.10). Dabei steht q_{st} für den maximalen Wasserverbrauch am Spitzenlasttag und q_d für den durchschnittlichen Wasserverbrauch pro Person und Tag. (Zhang et al., 2016)

$$V_h = \frac{q_{st}}{q_d} \tag{2.10}$$

Damit ähnelt die Definition jener nach Gleichung (2.1). Die Autoren begründen damit die Vergleichbarkeit der Kennlinien. Vielmehr stellt der mehrheitlich vorhandene Gebäudetyp, welcher in China vorwiegend in Form eines mehrstöckigen Gebäudes mit zahlreichen Wohneinheiten und in Europa häufiger in Form von Einfamilienhäusern vorhanden ist, einen Grund für den abweichenden Verlauf der chinesischen Kennlinie im Vergleich zu den europäischen Kennlinien dar. Durch vielfältige individuelle Nutzungen der einzelnen Wohneinheiten, auf Grundlage diversifizierter Komfortansprüche, sind voneinander abwei-

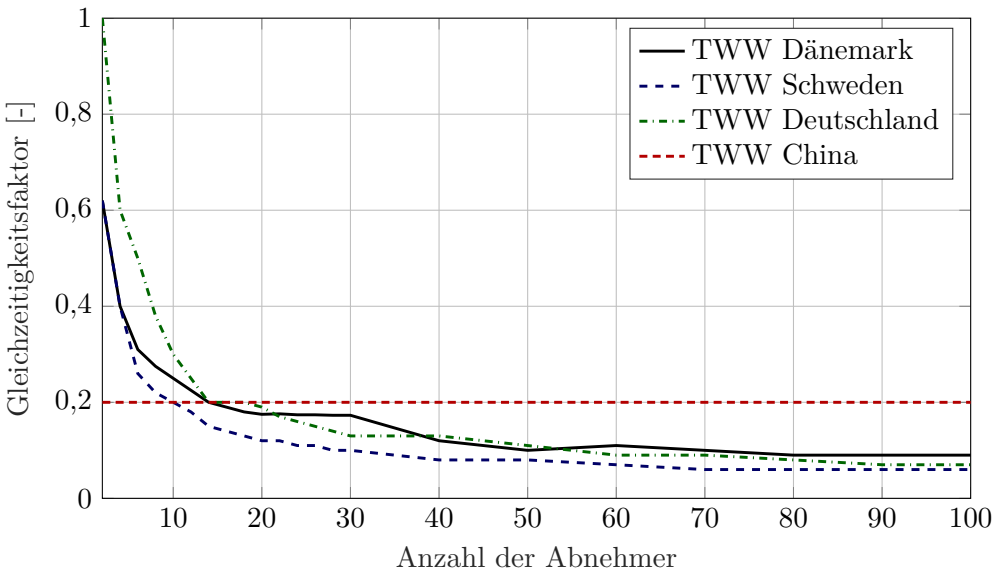


Abbildung 2.6: Vergleich länderspezifischer Kennlinien zu Gleichzeitigkeitsfaktoren in Anlehnung an Zhang et al. 2016

chende Trinkwarmwasserzapfprofile zu erwarten, welche den geringen Gleichzeitigkeitsfaktor aus China erklären. Demgegenüber implizieren Einfamilienhäuser im Vergleich dazu eher homogene Zapfprofile, wodurch eine höhere Gleichzeitigkeit anzunehmen ist. Gleichwohl wird auch auf die Möglichkeit von Fehlern bei der Berechnung der Auslegungsbedingungen für den Wärmebedarf und auf den empirischen Charakter des Gleichzeitigkeitsfaktors hingewiesen. (Zhang et al., 2016)

Im Rahmen seiner Dissertation stellt Blesl charakteristische Größen für die Fernwärmeversorgung in Abhängigkeit von Siedlungstypen vor (Blesl, 2002). In diesem Zusammenhang werden Kennwerte für Gleichzeitigkeitsfaktoren vorgeschlagen, welche Tabelle 2.1 zu entnehmen sind. Konstantin nimmt dahingehend für die Fernwärmeversorgung von Mischgebieten Gleichzeitigkeitsfaktoren in einem Bereich zwischen 0,7 und 0,8 sowie für reine Wohngebiete Werte zwischen 0,8 und 0,95 an (Konstantin, 2013). Mit der Fragestellung, inwieweit der Peak-Strompreis eines Forschungszentrums mittels einer Gleichzeitigkeitsanalyse reduziert werden kann, motivieren Yarbrough et al. (Yarbrough et al., 2015) die Untersuchung des Verhältnisses zwischen des individuellen Spitzenstrombedarfs eines Gebäudes und jenem des gesamten Campus. Als Grundlage dienen fünfzehnminütige Zählerdaten (Yarbrough et al., 2015). Die Autoren unterscheiden im Rahmen ihrer Studie zwischen vier Szenarien:

i. Hoher Bedarf mit geringer täglicher Fluktuation
ii. Geringer Bedarf mit geringer täglicher Fluktuation
iii. Tägliche Fluktuationen, die gleichzeitig zum Campus-Profil auftreten
iv. Tägliche Fluktuationen, die sich gegensätzlich zum Campus-Profil verhalten

Tabelle 2.1: Kennwerte für Gleichzeitigkeitsfaktoren in Abhängigkeit von dem Siedlungstyp nach Blesl, 2002

Siedlungstyp	**GLF**
Lockere offene Bebauung (Streusiedlung)	0,90
Einfamilien- und Doppelhäuser-Siedlung	0,85
Dorfkern	0,80
Reihenhäuser	0,85
Siedlung kleiner Mehrfamilienhäuser	0,80
Zeilenbebauung mit kleinen und großen Mehrfamilienhäusern	0,75
Zeilenbebauung mit großen Mehrfamilienhäusern oder Hochhäusern	0,70
Blockbebauung niedriger Dichte	0,80
Blockbebauung hoher Dichte	0,80
Citybebauung	0,70
Historische Altstadt	0,80
Öffentliche Sonderbauten (groß)	0,70
Gewerbliche Sonderbauten / Dienstleistungsbauten	0,70

Hinsichtlich der Frage, wann ein Gebäude in welcher Höhe zum Spitzenbedarf des Campus beiträgt, liegt ein zusätzliches Augenmerk auf der saisonalen Komponente der Lastprofile. Aufgrund der variierenden Gebäudenutzerzahlen eines Universitätscampus beeinflusst das Nutzerverhalten heizungs-, kühlungs- und raumlufttechnische Anlagen erheblich. Im Rahmen dieser Arbeit dienen Heatmaps der Visualisierung von Gebäudebedarfsprofilen, wobei der Vergleich einzelner Heatmaps zur Campus Heatmap aufgezeigt werden kann. Mittels dieser Methodik lassen sich Schlüsselgebäude und charakteristische Gebäudenutzungen identifizieren, die den größten Einfluss auf den Spitzenbedarf des Campus haben. In Kombination mit dem Gleichzeitigkeitsfaktor wird in dieser Studie ein Modell vorgestellt, welches das Energiemanagement unterstützt. (Yarbrough et al., 2015)

Bezugnehmend auf die Quantifizierung des Gleichzeitigkeitsfaktors für die Dimensionierung von Kältenetzen sind kaum Studien vorhanden, welche Erfahrungswerte oder Kennlinien beinhalten. Shu et al. approximieren den Gleichzeitigkeitsfaktor in ihrer Forschungsarbeit auf einen Bereich zwischen 0,8 und 0,9 Shu et al., 2010.

2.2.3 Implikationen und Anforderungen an die Modellierung

Die mathematische Formulierung des Gleichzeitigkeitsfaktors variiert in der Forschung, wobei weitestgehend Einigkeit über die inhaltliche Aussage besteht. An dem Vergleich zwischen den hergeleiteten Kennlinien zeigt sich vornehmlich der empirische Charakter des Gleichzeitigkeitsfaktors, da die Wertebereiche jener bei unterschiedlicher Abnehmergruppengröße wesentlich voneinander abweichen. Dadurch ist die Übertragbarkeit der Ergebnisse auf andere Anwendungsfälle begrenzt. Es existieren dabei unterschiedliche Kennlinien

für Raumwärme und Trinkwarmwasser auf Wohnungs- und Gesamtgebäudeebene. Die verwendeten statistischen Methoden zur Herleitung der Kennzahlen und Kennlinien werden selten erläutert, mit Ausnahme der umfassenden Studie von Winter et al., im Rahmen derer auch ein Lösungsvorschlag für etwaige Unter- und Überdimensionierungen des Gleichzeitigkeitsfaktors formuliert wird. Die Ermittlung der Nennleistung erfolgt dabei mittels einer linearen Regressionsanalyse, welche jedoch auch bei geringer linearer Korrelation zwischen Lastprofil und Außentemperatur an methodische Grenzen stößt (Gaderer, 2008). Zukünftige Veränderungen in der charakteristischen Leistungsabnahme, beispielsweise durch Sanierungen oder Änderungen der Versorgungsvarianten, finden noch keine Berücksichtigung in den bisher verwendeten Methoden. Gleichwohl sind diese Faktoren überaus relevant für die nachhaltige Planung urbaner Energiesysteme. Es bedarf daher eines Modells, welches zukünftige Lastszenarien berücksichtigt, komplexere Zusammenhänge zwischen thermischen Lasten und Nutzungs- bzw. Klimavariablen berücksichtigt und simultan das Risiko einer Unter- und Überdimensionierung des Gleichzeitigkeitsfaktors verringert. Um verschiedene Planungs- und Betriebszustände sowie eine variierende Informationsdichte fortlaufend berücksichtigen und damit eine optimierte Quantifizierung des Gleichzeitigkeitsfaktors ermöglichen zu können, bedarf es eines selbstlernenden, adaptierenden Modells. Letzteres ermöglicht es auch, urbane Regionen von differierender Größe unter Berücksichtigung extern variierender Einflüsse, wie Sanierungsreihenfolgen, im Hinblick auf die Gleichzeitigkeiten in der thermischen Leistungsabnahme zu untersuchen. Guan et al. haben in ihrer Arbeit mittels Clusteranalysen Gebäudegruppen identifiziert, welche sich für Lastverschiebungen im besonderen Maße eignen, da sie den größten Einfluss auf die Spitzenlast auf Quartiersebene haben (Guan et al., 2016). Hierbei stehen einzelne Kenngrößen und dazugehörige Wertespektren sowie eine vordefinierte Clusteranzahl im Fokus. Um diesen Ansatz weiterzuentwickeln, sollten Gebäudegruppen auf Grundlage ihrer individuellen Lastcharakteristika klassifiziert und die Ähnlichkeit zum Lastprofil auf Quartiersebene analysiert werden, um stetiges Lastverschiebungspotential zu identifizieren. Daran anknüpfend lassen sich ferner auch weitere Erkenntnisse, beispielsweise zu möglichen Versorgungsinseln, gewinnen. Auf Grundlage des abgeleiteten Forschungsbedarfs in diesem Abschnitt wird ein autoadaptives prädiktives Modell zur numerischen Abbildung von Lastszenarien sowie ähnlichkeitsbasierten Lastprofilanalysen für die Quantifizierung von Gleichzeitigkeitseffekten innerhalb dynamischer kurz- mittel- sowie langfristiger Planungsprozesse benötigt.

2.3 Zusammenfassung

Die Quantifizierung von Gleichzeitigkeiten in der urbanen thermischen Lastverteilung setzt Kenntnisse über den thermischen Energiebedarf der Gebäude voraus. Hinsichtlich der Ermittlung des letzteren ist zwischen physikalischen und datenbasierten Modellen zu unterscheiden, welche auf vereinfachten oder detaillierten Methoden basieren können. Die Entscheidung für eine Modellkategorie ist von der zugrundeliegenden Informationsbasis abhängig, wobei auch physikalische und datenbasierte Modelle miteinander kombiniert werden können. Detaillierte Methoden zur Berechnung des thermischen Energiebedarfs weisen gegenüber vereinfachten Methoden bessere Verallgemeinerungseigenschaften auf und

ermöglichen dadurch universellere Einsatzmöglichkeiten. Gleichwohl liefern vereinfachte Modelle eine erste Abschätzung des gebäudespezifischen thermischen Energiebedarfs, der im Einzelfall von vergleichbarer Genauigkeit wie jener Auf Grundlage der Berechnung mit detaillierten Methoden sein kann. Eine verstärkte Einführung von intelligenten Zählersystemen im Rahmen des Aufbaus smarter integrierter Energiesysteme und der stetig steigenden Menge an Messdaten im Energiesektor begründet die aktuelle und zukünftige hohe Relevanz von datenbasierten Modellen und ermöglicht den weiteren zukünftigen Einsatz des in dieser Arbeit entwickelten Modells zur Modellierung des gebäudespezifischen thermischen Energiebedarfs. Die Aufarbeitung und Analyse der vorhandenen Literatur zur Quantifizierung des Gleichzeitigkeitsfaktors ergibt einen Forschungsbedarf insbesondere im Bereich der Integration selbstlernender Algorithmen, welche komplexere Zusammenhänge zwischen thermischen Lasten der Gebäude als Energieabnehmer identifizieren sowie die gesamte Lastcharakteristik der Gebäude in die Berechnung des Gleichzeitigkeitsfaktors zum aktuellen Lastzustand und im Hinblick auf zukünftige veränderliche Bedarfsszenarien berücksichtigt. Weiterhin ist es notwendig Unsicherheitsanalysen zur Abbildung von Entscheidungskorridoren und zur Reduzierung der Ergebnisabhängigkeit von eventuell unrealistischer Anomalien in der Datengrundlage vorzusehen.

KAPITEL 3

Prognosemodelle und Szenariotechniken

Für die Abbildung zukünftiger thermischer Lastzustände und -szenarien urbaner Regionen bedarf es Prognosemethoden und Szenariotechniken. Letztere dienen der Modellierung von Unsicherheiten hinsichtlich zukünftiger Einflussparameter auf die Prognosen. Für die automatisierte Ähnlichkeitsanalyse gebäudespezifischer Lastgänge eignen sich insbesondere Clustering-Methoden, mit welchen eine Gruppierung von Lastprofilen im Sinne der Abbildung von Gleichzeitigkeiten in Leistungsabnahmecharakteristika innerhalb von Gebäudegruppen möglich ist. Damit wird die Abhängigkeit der Quantifizierung des Gleichzeitigkeitsfaktors von einzelnen Lastspitzen in historischen Messdaten aufgelöst, indem normierte funktionelle Verläufe der Lastgänge einer Ähnlichkeitsanalyse unterzogen werden. Dadurch kann der Einfluss von Messfehlern in den historischen Datensätzen oder individueller nutzungsbedingter Leistungsspitzen reduziert und eine verbesserte Verallgemeinerbarkeit des kollektiven Auftretens von Leistungsspitzen gewährleistet werden. Im Rahmen dieses Kapitels erfolgt daran anknüpfend in Abschnitt 3.1 die Darstellung des Stands der Technik und Wissenschaft im Zusammenhang mit der Entwicklung von Lastprognosemodellen. Abschnitt 3.2 behandelt verschiedene Szenariotechniken, wobei die Einordnung von Szenariotechniken zur quantifizierbaren Abbildung von Einflussparametern und disruptiven Ereignissen vor dem Hintergrund von Lastprognosen in das Forschungsfeld der Szenarioanalyse vorgenommen wird. Abschließend sind in Kapitel 3.3 die wesentlichen Implikationen und Anforderungen an die Modellentwicklung zusammengefasst.

3.1 Prognosemodelle

Dieser Abschnitt behandelt den theoretischen Hintergrund zu Lastprognosemodellen und bildet den aktuellen Stand der Forschung auf Grundlage von Studien über Methoden zur Vorhersage gebäudebezogener energetischer Lasten ab. Auf Grundlage der herausgearbeiteten Modellcharakteristika und Ergebnissen anwendungsbasierter Studien wird die Wahl des Prognosemodells begründet und implizite Anforderungen für die Teilmodellentwicklung hergeleitet und formuliert.

3.1.1 Theoretischer Hintergrund

Die Abschätzung des Zeitpunktes der maximalen Leistungsanforderung des Quartiers und der tatsächlich auftretenden Höchstlast bedarf Prognosen, die in ihrer Genauigkeit maßgebend für die Ermittlung eines realistischen Gleichzeitigkeitsfaktors sind. In diesem Zusammenhang ist zwischen einem kurzfristigen und langfristigen Prognosehorizont zu unterscheiden. Während ersterer zeitliche Granularitäten von Minuten, Stunden oder Tagen impliziert, betrifft letzterer einen Zeitraum von mindestens einem Jahr (Amasyali et al., 2018). In Forschung und Praxis finden verschiedene Modelle und Modellkombinationen Anwendung. Häufig verwenden Energieversorgungsunternehmen Standardlastprofile auf Grundlage eines normierten Verbraucherverhaltens innerhalb von Kundengruppen im Zuge des Vertriebs von Strom (Schellong, 2016). Darüber hinaus findet die Referenzmethode Anwendung, welche auf einer historischen Datenbasis zu Lastprofilen repräsentativer Tage beruht. Liegen Vergleichstage vor, welche ähnliche charakteristische Merkmale aufweisen, so erfolgt die Annahme, dass der entsprechende historische Lastgang des repräsentativen Tages gleich ist. Der abgeleitete Prognosewert lässt sich zusätzlich mit einflussspezifischen Korrekturwerten manipulieren, sodass gezielt beispielsweise klimabedingte Einflüsse berücksichtigt werden können. Die Referenzmethode findet in der Praxis häufig auch in Kombination mit komplexeren Prognosemodellen Anwendung. (Schellong, 2016)

Neben diesen Methoden zur Annahme zukünftiger energetischer Lasten, lassen sich die folgenden drei Hauptkategorien zur Quantifizierung des zukünftigen Energiebedarfs identifizieren.

I. Zeitreihenmodelle
II. Kausalmodelle
III. Modelle basierend auf maschinellem Lernen

Bei Zeitreihenmodellen wird die prognostizierte energetische Last als Funktion in Abhängigkeit der dazugehörigen historischen Funktionswerte modelliert. Kausalmodelle demgegenüber definieren zukünftige Lasten mittels exogener Faktoren, wie beispielsweise Wettervariablen oder nutzerbezogenen Variablen (Hippert et al., 2001). In diesem Zusammenhang wird beispielsweise die Heizlast mittels dieser exogenen Einflussfaktoren mit Hilfe historischer Daten beschrieben und darauf aufbauend ein Regressionsmodell entwickelt. Der daraus resultierende funktionale Zusammenhang zwischen der Heizlast und den beiden exogenen Variablen dient der Prognose zukünftiger Heizlasten (Dotzauer, 2002). Unter Modellen des maschinellen Lernens werden lernende Methoden zusammengefasst (Amasyali et al., 2018), welche oftmals auch Eigenschaften von Zeitreihen- und Kausalmodellen aufweisen. Mit ihnen lassen sich komplexe nicht-lineare Probleme lösen (Magoulès et al., 2016). Das Ergebnis eines Algorithmus des maschinellen Lernens lässt sich beispielsweise als Ausgabevektor y ausdrücken, welcher durch den Eingangsvektor x und adaptiven Gewichtungsfaktoren beschrieben werden kann (Bishop, 2006). Im Modelltrainingsprozess (Lernphase) erfolgt die sukzessive Anpassung letzterer, um die Differenz zwischen y und dem Zielvektor t zu

minimieren. Auf Basis des dadurch trainierten Modells und einem neuen Eingangsvektor lässt sich ein neuer Ausgabevektor prognostizieren.

Zeitreihenmodelle

Mit einem Zeitreihenmodell wird der Energiebedarf als ausschließlich zeitabhängig modelliert und prognostiziert. Hierbei gilt die Annahme, dass sich die Daten aus verschiedenen Komponenten zusammensetzen, deren Kombination einen Trend beschreibt, welcher sich in die Zukunft fortsetzt. Die Komponenten und ihr funktionaler Zusammenhang hängen von der Modellierung ab. So kann beispielsweise zwischen einem additivem und einem hybriden Zeitreihenmodell unterschieden werden. Das additive Zeitreihenmodell wird durch die Summe aus Trendkomponente, Saisonkomponente, Zykluskomponente und irregulärer Komponente definiert. Dagegen ist das hybride Zeitreihenmodell mittels des Produkts aus Trend- und Saisonkomponente, in Addition mit der irregulären Komponente, beschrieben. Die mathematisch einfach gehaltene Struktur erlaubt eine leichte Implementierung. Demgegenüber sind jedoch keine zusätzlichen zeitunabhängigen Einflussgrößen in das Modell integrierbar. (Schellong, 2016) Zu den Zeitreihenmodellen gehören Autoregressionsmodelle. Die Grundlage für Autoregressive (AR) Modelle bilden Kombinationen aus Bestandteilen historischer Komponenten, addiert zu Eingangsgrößen zum aktuellen Zeitpunkt. Die historischen Komponenten werden mit einem Autoregressionskoeffizienten jeweils gewichtet, wobei letztere Approximationen auf Basis von Messwerten darstellen. Die Integration gleitender Mittelwerte in das Autoregressionsmodell führt zum Autoregressive Integrated Moving Average (ARIMA) Modell. (Schellong, 2016)

Daran anknüpfend untersuchten beispielsweise Mbamalu und El-Hawary ein multiplikatives Autoregressionsmodell (Mbamalu et al., 1993) und Huang ein grenzwertbasiertes Autoregressionsmodell (Huang, 1997). Neben den Autoregressionsmodellen werden im Zusammenhang mit Zeitreihenmodellierungen auch Kalman Filter verwendet (Infield et al., 1998; Park et al., 1991; Sargunaraj et al., 1997). Letztere stellen dynamische Schätzer für Messdaten (Kalman, 1960) zur Lösung linearer Problemstellungen dar, wobei das Modell auch mit Hilfe von Linearisierungen zur Modellierung nicht-linearer Probleme verwendet werden kann (Ko et al., 2013).

Kausalmodelle

Im Rahmen von Kausalmodellen finden Regressionsmodelle zur Lastprognose Anwendung, sofern der Energiebedarf in Abhängigkeit von einer geringen Anzahl an Einflussfaktoren beschrieben werden soll. In diesem Kontext sind aufgrund ihres vermehrten Einsatzes lineare Regressionsmodelle zu nennen (Haida et al., 1994; Papalexopoulos et al., 1990; Soliman et al., 1997). Varianten dieser Modelle stellen Regressionsmethoden dar, welche aus einer witterungsunabhängigen und einer witterungsabhängigen Komponente bestehen (Fan et al., 1994; Hyde et al., 1997; Park et al., 1991; Dotzauer, 2002).

Modelle basierend auf maschinellem Lernen

In dieser Kategorie sind vornehmlich Künstliche Neuronale Netze (KNN), Support Vector Machines (SVM) und Decision Trees (DT) aufgrund ihrer breiten Anwendung zu nennen. Auch statistische Modelle, welche hier beispielsweise in die Kategorie der Zeitreihenmodelle fallen, können teilweise dem maschinellen Lernen zugeordnet werden (Amasyali et al., 2018). SVM stellt einen Kernel-basierten Algorithmus dar, welcher für die Lösung von Regressions- und Klassifikationsproblemen anwendbar ist (Wu et al., 2008). SVM zählt zu den robustesten Algorithmen des maschinellen Lernens und zeigt sich in zahlreichen Anwendungsfällen gegenüber anderen Algorithmen des maschinellen Lernens überlegen (Wu et al., 2008; Amasyali et al., 2018). Eine SVM lässt sich dabei entweder als Klassifikator oder Regressor anwenden und ermöglicht damit eine universelle Einsetzbarkeit. Das Prinzip der SVM basiert auf der Suche einer optimalen Entscheidungsgrenze entweder für die Klassifikation von Objekten oder die Modellierung von Beziehungen zwischen Objekten. Hierzu dienen Support Vectors als definierende Trainingstupel, wobei ihr größtmöglicher Abstand zur Entscheidungsgrenze gesucht wird, um die optimale Lösung für das Klassifikations- bzw. Regressionsproblem zu finden (Han et al., 2012). Um den Rechenaufwand der Modelle zu reduzieren wurden der Least Squares SVM (LS-SVM) Algorithmus (Edwards et al., 2012) und parallele SVM (Zhao et al., 2010) vorgeschlagen (Amasyali et al., 2018). KNN dagegen stellen dem Gehirn nachempfundene Modellierungen von Neuronen und Verbindungen zwischen diesen dar (Amasyali et al., 2018), welche lernfähig sind und mithilfe von Daten trainiert werden können. Diese nicht-linearen Berechnungsmodelle finden in verschiedenen Bereichen Anwendung und lenken den Fokus der Forschung auf sich (Zhang et al., 1998). Häufig verwendete modelltheoretische Konfigurationen der KNN sind nachfolgend aufgeführt (Magoulès et al., 2016).

i. Einschichtiges Perzeptron Netz
ii. Rekursives deterministisches Perzeptron Netz
iii. Mehrschichtiges Perzeptron Netz
iv. Neuronales Netz mit radialer Basisfunktion
v. Rekurrentes Netz

Das einschichtige Perzeptron-Netz besteht ausschließlich aus einer Ausgabeschicht, dessen Ausgabewert y den gewichteten Eingangswerte x_i entspricht. Die jeweiligen Gewichtungsfaktoren w_i werden dabei im Rahmen des Modelltrainings modifiziert und optimiert. Abbildung 3.1 visualisiert diesen Netzarchitekturtypus exemplarisch. Das einschichtige Perzeptron-Netz folgt in der Regel dem Prinzip des vorwärtsgerichteten Informationstransfers innerhalb des Netzes und wird daher auch als *Feed Forward* Netz bezeichnet. (Magoulès et al., 2016) Eine Verallgemeinerung des einschichtigen Perzeptron Modells stellt das rekursive deterministische Perzeptron-Netz dar (Tajine et al., 1998), welches während des Trainingsprozesses dynamisch wächst und damit zu einem mehrschichtigen Perzeptron-Netz mutiert (Magoulès et al., 2016).

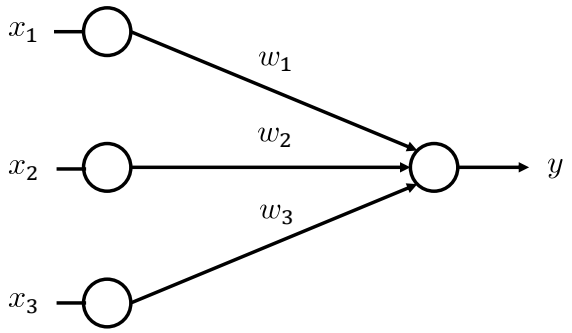


Abbildung 3.1: Exemplarische Darstellung eines einschichtigen Perzeptron-Netzes in Anlehnung an Magoulès et al., 2016

Im Sinne eines konventionellen mehrschichtigen Perzeptron-Netzes wird die Netzarchitektur um eine verdeckte Schicht erweitert, wobei die Neuronen einer Ebene mit den jenen einer anderen Ebenen mittels gewichteter Kanten verbunden sind. Abbildung 3.2 veranschaulicht ein exemplarisches mehrschichtiges Perzeptron-Netz mit jeweils drei Neuronen in der Eingangsschicht und in der verdeckten Schicht sowie zwei Neuronen in der Ausgabeschicht.

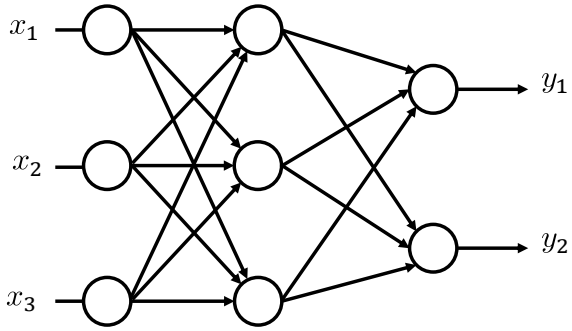


Abbildung 3.2: Exemplarische Darstellung eines mehrschichtigen Perzeptron Modells in Anlehnung an Magoulès et al., 2016

Jedes Neuron besteht aus zwei Units. Das erste Unit bildet jeweils das Produkt aus Eingangs- und Gewichtungsfaktor. Das zweite Unit des Neurons generiert die dazugehörige nicht-lineare Aktivierungsfunktion $f(.)$, wobei der Ausgangswert y gleichzeitig wieder als Eingangswert für ein anderes Unit dienen kann. (Magoulès et al., 2016) Das Netzwerktraining entspricht der sukzessiven Anpassung und Optimierung der Gewichtungsfaktoren für die Verbindungen zwischen den Neuronen. Hierbei wird jeder mittels eines Trainingstupels prognostizierte Ausgabewert des Netzwerkes dem bekannten Zielwert gegenüber gestellt und die quadratische Abweichung bestimmt. Letztere ist zu minimieren, indem die Gewichtungsfaktoren modifiziert werden. Dieser Anpassungsprozess findet, ausgehend von der Ausgabeschicht des Netzwerkes, rückwärts gerichtet statt (Backpropagation). (Han et al.,

2012) KNN mit radialer Basisfunktion f bestehen stets genau aus drei Schichten mit einer verdeckten Schicht (Magoulès et al., 2016). Abbildung 3.3 visualisiert diesen Netztypus.

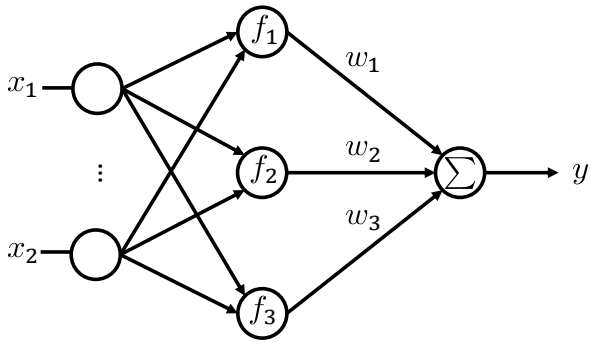


Abbildung 3.3: Exemplarische Darstellung eines KNN mit radialer Basisfunktion in Anlehnung an Magoulès et al., 2016

Rekurrente Neuronale Netzwerke (RNN) unterscheiden sich von klassischen Feed Forward Modellen insofern, als dass sich das Netzwerk selbst beeinflusst, indem bereits generierte Ausgangswerte in den folgenden Berechnungen berücksichtigt werden. In diesem Fall beinhaltet die verdeckte Schicht ein Neuron, welches den aktuellen Wissensstand des KNN behält (Magoulès et al., 2016). Dahingehend veranschaulicht Abbildung 3.4 diese Netztopologie exemplarisch.

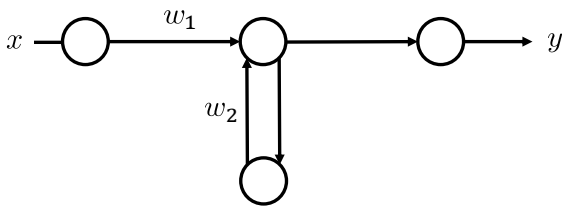


Abbildung 3.4: Exemplarische Darstellung eines Rekurrenten Netzes in Anlehnung an Magoulès et al., 2016

Im Sinne der Auslegung und des prädiktiven Betriebs gebäudetechnischer Anlagen ermöglichen RNN beispielsweise die Modellierung der thermischen Trägheit eines Gebäudes, indem Ausprägungen beschreibender Variablen wie der Außentemperatur zu vorherigen Zeitschritten für die Prognose zukünftiger thermischer Lastzustände berücksichtigt werden. Somit kann die Einschwingphase von Gebäuden in Abhängigkeit ihrer thermischen Masse abgebildet werden. Neben SVM und KNN stellen DT eine weitere Methodenkategorie dar, welche mittels Baumstrukturen Entscheidungsregeln modellieren. Dabei wird an jedem Knotenpunkt ein Attribut abgefragt, bis ein Blatt erreicht wird, welches eine Prognose darstellt. Die Entscheidungsbäume wachsen flexibel mit der Menge an Daten (Amasyali et al., 2018). Im Zusammenhang mit dieser Kategorie sind weitere Algorithmen, wie beispielsweise

Classification and Regression Trees (CART), Chi-squared Automatic Interaction Detector (CHAID) oder Random Forests (RF) als Prognosemethoden für den Energieverbrauch hervorzuheben (Amasyali et al., 2018).

3.1.2 Studien

Auf Grundlage der bisherigen Forschungsarbeiten lassen sich im Wesentlichen zwei Modelltypen für die Prognose thermischer Lastprofile unterscheiden:

I. Physikalische Modelle

II. Datenbasierte Modelle

Erstere bedingen ausreichend Informationen zu den gebäudespezifischen thermischen Eigenschaften, Materialien oder dem Nutzerverhalten (Al-Homoud, 2001). Dieser Umstand führt häufig zu Problemen in der Parametrierung von physikalischen Gebäudemodellen, welche durch komplexe und kostenintensive Datensammlungen gelöst werden müssen. In diesem Zusammenhang sind datenbasierte Modelle vorteilhaft, deren Anwendung jedoch eine detaillierte historische Datenbasis voraussetzen. In der Studie Neto et al. werden Ergebnisse von Gebäudesimulationen auf Grundlage physikalischer Modellierungen jenen eines datenbasierten Modells, basierend auf einem KNN, gegenüber gestellt (Neto et al., 2008). Dabei konnten ähnliche Ergebnisse erzielt werden, obwohl lediglich eine geringe Menge an Parametern zur Verfügung stand, um das KNN zu trainieren.

Kurzfristige datenbasierte Prognosen beruhen häufig auf Autoregressive Moving Average (ARMA) oder (ARIMA) Modellen (Fan et al., 1994). Auch vereinfachte Regressionsmodelle, wie abschnittsweise lineare Regressionsmodelle finden Anwendung. So prognostizierte Dotzauer Heizlasten beispielsweise mittels einer witterungsabhängigen Komponente sowie einer sozialen Komponente und modelliert damit eine abschnittsweise lineare Zielfunktion (Dotzauer, 2002). Mehr denn je konzentriert sich die Forschung jedoch auf Prognosemodelle, basierend auf Methoden der künstlichen Intelligenz. Diese beruhen dabei auf Experten-Systemen oder Methoden des maschinellen Lernens (Geysen et al., 2018). In diesem Zusammenhang haben Sakawa et al. Prognosen mittels eines dreischichtigen KNN mit multiplen Output Units untersucht (Sakawa et al., 2001). Catalina et al. legten den Fokus ihrer Studie auf ein multiples Regressionsmodell und betrachteten verschiedene Eingangsparameter (Catalina et al., 2013). Im Kontext der Prognose von Heizlasten im Gebäudebereich untersuchten Shamshirband et al. Neuro-Fuzzy-Methoden (Shamshirband et al., 2015). Aktuell beschäftigen sich zahlreiche Studien mit SVM als Methode zur Prognose von Lastprofilen. So untersuchten beispielsweise Li et al. stündlich prognostizierte Kühllasten eines Bürogebäudes (Li et al., 2009). Im Rahmen eines ganzheitlichen gebäudebezogenen Monitoringkonzepts integrierten Wu et al. ein Support Vector Machine Regression (SVM-R) Modell (Wu et al., 2012). Magoulès et al. kombinierten ein SVM-Modell mit einem KNN zur Identifizierung von Fehlern in den Vorhersagen (Magoulès et al., 2016). Prognosen mittels SVM haben sich in zahlreichen Studien bewährt und zeigten sich gegenüber anderen Methoden überlegen. Häufig konnten mittels SVM-Modellen Prognosen

mit geringeren Fehlerraten erzielt werden als mit KNN auf Grundlage unterschiedlicher Netzwerk-Architekturen (Amasyali et al., 2018).

Im Kontext der Prognose von Zeitreihen zeigen Nonlinear Autoregressive Networks with Exogenous Inputs (NARX) als spezielle Konfigurationen von RNN vielversprechende Ergebnisse. In Mena et al. wurde eine NARX-Architektur gewählt um den Strombedarf von Nicht-Wohngebäuden vorherzusagen (Mena et al., 2014). In der Studie von Powell et al. wurden verschiedene Netzwerk-Architekturen im Rahmen der Prognose von Strom-, Kälte- und Heizlastgängen untersucht sowie einander vergleichend gegenüber gestellt. Es zeigte sich dabei, dass sich die NARX-Architektur am besten für diesen Anwendungsfall eignet (Powell et al., 2014). Weitere Studien zur Modellierung von Zeitreihen und nicht-linearen Systemen unterstützen dieses Ergebnis (Connor et al., 1992; Chen et al., 1990). Weiterhin konnte die Vorteilhaftigkeit dieser Architektur hinsichtlich der Prognose chaotischer Zeitreihen und der Reduzierung von Problemen bezüglich der Langzeit-Abhängigkeiten heraus gearbeitet werden (Diaconescu, 2008; Lin et al., 1996). In einer komparativen Studie von Koschwitz et al. konnte auf urbaner Ebene ferner gezeigt werden, dass sich NARX RNN für datenbasierte Prognosen von Heiz- und Kühllasten eignen. Im Rahmen dieser Studie wurde jedoch auch aufgezeigt, dass SVM-R Modelle mit radialer Basisfunktion und polynomialen Kernel dritten Grades im Vergleich zu NARX RNN für Kühllastberechnungen und Lastgänge, die starkes Rauschen aufweisen, in Einzelfällen zu besseren Ergebnissen führen. Im Weiteren ermöglichte die Vertiefung der Netzwerkarchitektur des NARX RNN in Einzelfällen genauere Prognosen des gebäudebezogenen thermischen Energieverbrauchs, jedoch zeigte sich demgegenüber auch vermehrt eine starke Tendenz zur Überanpassung in der Trainingsphase, wodurch größere Abweichungen zwischen den prognostizierten Werten und den Zielwerten in der Testphase resultierten. Hinsichtlich der Rechenzeiten wurde der Einfluss verschiedener Modellkonfigurationen analysiert und die Bewertung der einzelnen Modelle für den Anwendungsfall miteinbezogen. Im Ergebnis zeigte sich, dass NARX RNN mit einer verdeckten Schicht den anderen Modellen mehrheitlich überlegen war, bei einem vergleichsweise geringen Rechenaufwand. (Koschwitz et al., 2018)

3.1.3 Implikationen und Anforderungen an die Modellierung

Das modelltheoretische Spektrum an Methoden zur Prognose energetischer Lasten umfasst sowohl reduzierte Modelle, basierend auf wenigen mathematischen und physikalischen Zusammenhängen, als auch hoch komplexe Optimierungen auf Grundlage umfangreicher Modellarchitekturen und einer Vielzahl an Eingangsparametern. In Einzelfällen können mit vereinfachten Regressions- und Autoregressionsmodellen Lastprognosen mit hoher Genauigkeit erzielt werden. Methoden des maschinellen Lernens ermöglichen dagegen als universelle Approximatoren eine bessere Modelladaption hinsichtlich variierender Datensätze und damit implizierte Generalisierungseigenschaften. Im Speziellen ermöglichen KNN- und SVM-Modelle Ergebnisse mit der größten Prognosegenauigkeit. Im Bereich der datenbasierten Prognosen konnte gezeigt werden, dass NARX RNN aufgrund ihrer rekursiven Eigenschaften, sehr gute Ergebnisse liefern. Ein Vergleich zwischen SVM-R-Modellen und

NARX RNN-Konfigurationen hat gezeigt, dass sich mit NARX RNN Prognoseergebnisse mit einer höheren Genauigkeit erzielen lassen, jedoch impliziert ein Deep Learning-Ansatz auch Probleme bezüglich eines möglichen Overfittings. In Einzelfällen stellen SVM gelegentlich die bessere Wahl dar. Auf Grundlage des erarbeiteten theoretischen Hintergrunds zu Prognosemodellen für die Modellierung zukünftiger Lastprofile und der Aufbereitung veröffentlichter Studien in der Literatur lassen sich Vorteile von SVM sowie NARX RNN gegenüber anderen Methoden des maschinellen Lernens begründen, wodurch beide Methodentypen für die Modellentwicklung in Frage kommen. Infolgedessen werden die folgenden Anforderungen an die Modellierung zusammengefasst:

I. Entwicklung eines Prognosemodell-Ensembles aus SVM und NARX RNN
   i. Modelle: NARX RNN unterschiedlicher Tiefe und SVM-R Modelle mit unterschiedlichen nicht-linearen Kerneln
   ii. Automatisierte Modellauswahl: Die Prognosegüte in der Testphase des Modells entscheidet darüber, welches Mitglied des Modellensembles für die thermische Lastprognose eines Gebäudes zum Einsatz kommt.

II. Variabilität der Modellkomplexität durch das Hinzuziehen und Ausgrenzen von Prognosemodellen innerhalb des Ensembles zur hardwarespezifischen Anpassung des Rechenaufwands

3.2 Szenariotechniken

Im Kontext dieses Abschnitts wird der theoretische Hintergrund zu Methoden der Generierung von gebäudebezogenen energetischen Lastszenarien vorgestellt und der aktuelle Stand der Technik und Forschung abgebildet. Zusammen mit den thematisch daran anknüpfenden Studien und Forschungsrichtungen erfolgt die Formulierung der Implikationen und Anforderungen an die Teilmodellentwicklung im Bereich der Methoden zur numerischen Abbildung zukünftiger Entwicklungspfade.

3.2.1 Theoretischer Hintergrund

Seit mehr als 40 Jahren stehen Szenariotechniken und -methoden im Mittelpunkt zahlreicher wissenschaftlicher Publikationen und praktischer Arbeiten. Während verschiedene Methoden entwickelt und Szenariokategorien definiert wurden, besteht vielfach keine Einigkeit in Bezug auf letztere (Bradfield et al., 2005). Es haben sich im wesentlichen drei Schulen etabliert (Amer et al., 2013).

i. Intuitive Logics School
ii. Probabilistic Modified Trends (PMT) School
iii. French School

Während im Sinne der *Intuitive Logics School* hypothetische Abfolgen von Ereignissen (Burt, 2007) auf Grundlage überwiegend nicht-mathematischer Methoden (Huss et al., 1987a) in Bezug auf Unternehmensentscheidungen für die Erstellung von Szenarien dienen, basieren Methoden der *PMT School* auf der Kombination aus quantitativen Prognosetechniken und qualitativen Einflussfaktoren (Bradfield et al., 2005; Chermack et al., 2001). Die *French School* rückt demgegenüber die Formulierung idealistischer Visionen in den Fokus der Szenarioentwicklung, welche politischen Entscheidungsträgern zukünftigen Handlungsbedarf aufzeigen soll (van Vught, 1987). Im Zusammenhang mit der nachhaltigen Planung von Energiesystemen stehen die Entscheidungsträger in der Regel vor der Herausforderung, verschiedene Kriterien und Zielstellungen zu berücksichtigen (Pohekar et al., 2004). Letztere bedingen in der Folge multikriterielle Analysen, welche bereits als Gegenstand zahlreicher Untersuchungen im Bereich von Energie und Umwelt fungierten (Huang et al., 1995; Lahdelma et al., 2000; Salminen et al., 1998). Mit der Integration von Szenarioanalysen in die Modellierung zukünftiger Planungszustände lässt sich die Unsicherheit hinsichtlich Entwicklungsvarianten repräsentieren und quantifizieren. Es ist in diesem Zusammenhang zwischen qualitativen und quantitativen Szenarioanalysen zu unterscheiden (Amer et al., 2013), wobei der Fokus im Rahmen dieser Arbeit im Wesentlichen auf den quantitativen Methoden liegt. Letztere dienen der Abbildung langfristiger gebäudespezifischer thermischer Lastzustände auf Grundlage veränderlicher Randbedingungen vor dem Hintergrund der prognosebasierten Quantifizierung von Gleichzeitigkeitsfaktoren zur optimierten strategischen Planung urbaner Energieversorgungsvarianten. Dennoch werden qualitative Annahmen getroffen, welche die Generierung quantitativer Szenarien begründen. Im Hinblick auf die genannten Schulen gründet die Szenarioanalyse dieser Arbeit auf den Grundgedanken der *PMT School*. Amer et al. unterscheiden auf Basis ihrer umfangreichen Recherchen in diesem Kontext zwischen vier Hauptkategorien der szenarienbasierten Planung (Amer et al., 2013).

i. Interactive Cross-Impact Simulation (INTERAX)

ii. Interactive Future Simulations (IFS)

iii. Trend Impact Analysis (TIA)

iv. Fuzzy Cognitive Map (FCM)

INTERAX beschreibt eine Prognosetechnik mit einem Szenariohorizont von einem Jahr, die auf einer Kombination aus analytischen Modellen sowie Expertenbeurteilungen beruht und dabei verschiedene Umwelteinflüsse berücksichtigt. Häufig werden Prognose- oder Simulationsmethoden integriert und beispielsweise Delphi-Techniken zur Analyse der Expertenbeurteilungen verwendet (Enzer, 1981; Huss et al., 1987a). Im Sinne der IFS-Technik, welche auch unter BATTELLE Scenario Inputs to Corporate Strategies (BASICS) bekannt ist (Amer et al., 2013), erfolgt die Identifizierung und Ableitung von Einflussfaktoren bezüglich zukünftiger Ereignisse. Sie unterscheidet sich von INTERAX darin, keine Monte Carlo Simulation oder individuelle Prognosen von Schlüsselvariablen mit einzubeziehen (Huss et al., 1987a). Die TIA-Technik basiert auf der Prognose einer Schlüsselvariablen unter Berücksichtigung historischer Trends, welche anschließend aufgrund von beeinflussenden zukünftigen Ereignissen angepasst wird (Gordon, 1994; Huss et al., 1987b). Mit der FCM

lassen sich komplexe Zusammenhänge zwischen Konzepten mittels gewichteter Kanten auf Grundlage eines Fuzzy-Graphs quantifizieren (Kosko, 1986; Jetter et al., 2011).

Neben der Kategorisierung von Szenario-Techniken hinsichtlich ihrer qualitativen und quantitativen Eigenschaften, lassen sich auch Unterscheidungen bezüglich der hinterlegten Zielstellungen vornehmen. Bishop et al. fassen verschiedene Ansätze aus der Literatur zusammen (Bishop et al., 2007). Beispielsweise unterscheiden Börjeson et al. sechs verschiedene Szenario-Typologien, welche in Abbildung 3.5 visualisiert sind. Mit prädiktiven Szenarien wird der Frage nachgegangen, *was passieren wird*, wobei explorativen Szenario-Techniken die Frage, *was passieren kann* zugrunde liegt und normative Szenarien die Frage beantworten, *wie ein spezifisches Ziel erreicht werden kann.* (Börjeson et al., 2006)

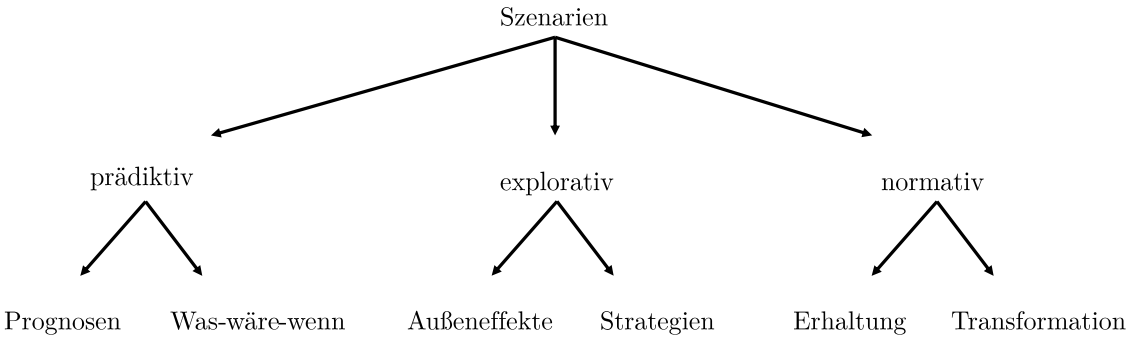


Abbildung 3.5: Szenario-Typologien nach Börjeson et al., 2006

Erstere basieren auf Prognosen oder Was-wäre-wenn-Szenarien, wobei zweitere durch Außeneffekte und Strategien beeinflusst werden. Letztere werden durch Szenarien der Erhaltung oder Transformation bestimmt. (Bishop et al., 2007) Im Rahmen dieser Arbeit soll vornehmlich im Sinne prädiktiver Szenarien die Frage beantwortet werden, in welcher Höhe der Gleichzeitigkeitsfaktor auf Grundlage der Variation von Randbedingungen anzusetzen ist, um eine nachhaltige Dimensionierung zentraler Energieversorgungssysteme zu ermöglichen. Die Anzahl der betrachteten Szenarien variiert in der Literatur in der Regel zwischen drei und sechs, wobei vornehmlich die Betrachtung von drei Szenarien vorgeschlagen wird, um einerseits die Tendenz zur Fokussierung auf zwei Extremszenarien oder einem Trendszenario zu vermeiden und andererseits das Verhältnis zwischen Rechenaufwand und Informationsertrag zu optimieren. Gleichwohl besteht das Risiko der Konzentration auf das Trendszenario. (Amer et al., 2013) Abbildung 3.6 veranschaulicht die Szenarioentwicklung über die Zeit unter Berücksichtigung wachsender Unsicherheit je weiter sich diese in die Zukunft erstrecken, am Beispiel zweier Extremszenarien und einem Trendszenario.

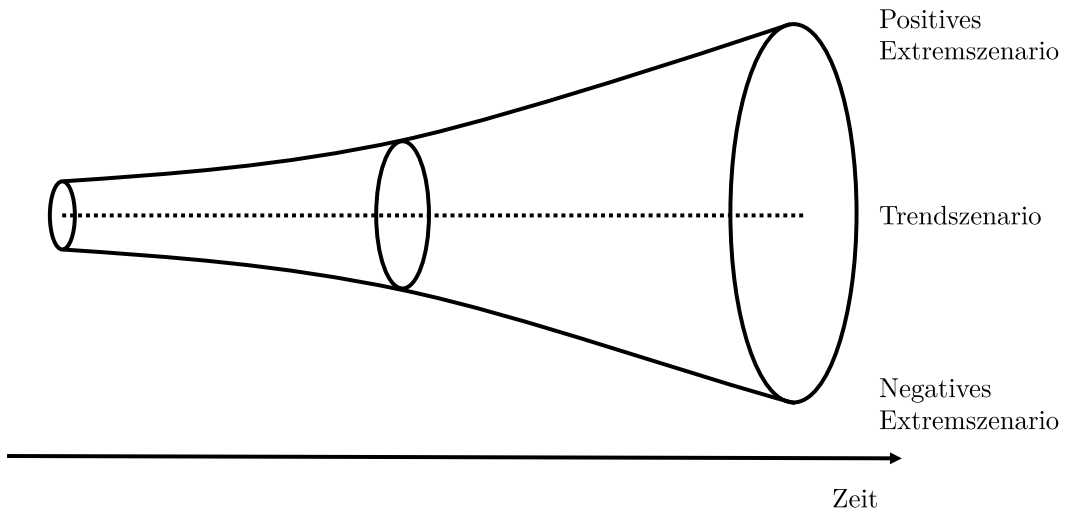


Abbildung 3.6: Visualisierung eines positiven und negativen Extremszenarios sowie eines Trendszenarios

3.2.2 Studien

Szenarioanalysen finden vielfältig Eingang in Analysen zur Entwicklung zukünftiger urbaner Energiesysteme in Form von Roadmaps (Grubler et al., 2012). Vor dem Hintergrund der Ziele der Europäischen Union bis 2050 die CO_2 Emissionen um 80% zu reduzieren, indem der Wärmesektor elektrifiziert wird, untersuchten Connolly et al. das Potenzial eines vermehrten Einsatzes zentraler Wärmeversorgungssysteme als Alternative (Connolly et al., 2016). In diesem Zusammenhang wurde ein Szenario basierend auf dem flächendeckenden Austausch alter Heizkessel und Wärmeerzeuger sowie der Installation von zentralen Wärmeerzeugungssystemen einem Referenzszenario gegenübergestellt. Neben der Analyse globaler und europäischer Szenarien dienen einzelne Regionen oder Energiesektoren als Grundlage weitergehender Untersuchungen. In diesem Kontext zeigten Cartalis et al. die potentiellen Veränderungen des Energiebedarfs urbaner Regionen infolge zukünftiger klimatischer Veränderungen im südöstlichen mediterranen Raum (Cartalis et al., 2001), wobei Asare-Bediako et al. den Einfluss des jahreszeitlichen Nutzungsgrades gebäudespezifischer Anlagentechnik auf zukünftige elektrische Lastprofile im Wohngebäudesektor untersuchten (Asare-Bediako et al., 2014). Im Rahmen der Studie von Yin et al. wurde die Auswirkung urbaner Strukturänderungen auf das Energieverbrauchsverhalten der Gebäudenutzer analysiert (Yin et al., 2015). Die Generierung von Szenarien beruht vermehrt auf Prognosen mittels Fuzzy-Methoden (Saez et al., 2015) oder KNN (Vagropoulos et al., 2016). Ferner finden probabilistische Prognosemodelle zur Modellierung unsicherer zukünftiger energetischer Lasten Anwendung (Hong et al., 2014; Liu et al., 2015b; Hong et al., 2016).

Im Zusammenhang mit der Abbildung unterschiedlicher Wetterszenarien als Eingangsvariablen für Lastprognosen mit KNN lassen sich Wetter-Ensembles verwenden, deren Szenarien unterschiedliche Entwicklungspfade der zukünftigen klimatischen Bedingungen darstellen (Taylor et al., 2002). Auf Basis dessen resultieren unterschiedliche Lastprognosen. Al Hamadi & Soliman korrelierten kurzfristige Lastprognosen zusammen mit jährlichen Steigerungsraten zur Abbildung langfristiger Lastszenarien (Al-Hamadi et al., 2005).

3.2.3 Implikationen und Anforderungen an die Modellierung

Hinsichtlich der Ziele dieser Arbeit folgt die Teilmodellentwicklung zur Generierung verschiedener Lastszenarien der *PMT School*, wobei der Fokus auf der Erstellung quantitativer Szenarien liegt. Aufgrund der Abhängigkeit thermischer Energielasten von den Witterungsverhältnissen, werden Wetterszenarien erstellt, deren Einfluss auf die zukünftigen gebäudespezifischen Lastprofile auch indirekt die Höhe des Gleichzeitigkeitsfaktors bestimmt. Um neben Klimavariablen auch Strukturänderungen in urbanen Räumen abbilden zu können, finden zusätzlich Sanierungsszenarien Eingang in die Teilmodellentwicklung. Damit ergibt sich eine Szenarioanzahl, die höher ausfällt, als mehrheitlich empfohlen wird. Dadurch kann einerseits die Tendenz zur Wahl des Trendszenarios verringert werden, gleichzeitig bedingt dies eine Erhöhung des Rechenaufwands. Aus diesem Grund werden bei der Modellierung Freiheitsgrade in Bezug auf die zu Szenarien bei der Ermittlung zukünftiger Gleichzeitigkeitsfaktoren berücksichtigt. Zusammengefasst ergeben sich für die Teilmodellentwicklung der Szenarioabbildung die folgenden Anforderungen an die Modellierung.

I. Generierung von Szenarien auf Grundlage klimatischer Veränderungen und struktureller Modifikationen urbaner Räume in Form von Gebäudesanierungen
II. Berücksichtigung von Freiheitsgraden in der Modellierung zur Ermöglichung variabel wählbarer Szenarien

3.3 Zusammenfassung

Im Kontext der Prognose des zukünftigen thermischen Energieverbrauchs von Gebäuden lassen sich im Wesentlichen Zeitreihen- von Kausalmodellen unterscheiden, wobei Modelle basierend auf maschinellem Lernen teilweise die methodischen Grundlagen letzterer vereinen und darüber hinaus erweitern. Der Fokus der Wissenschaft liegt ferner in diesem Kontext mehrheitlich auf KNN und SVM, da diese sich im Rahmen zahlreicher Studien anderen Methoden als überlegen erwiesen haben. Hinsichtlich ersterer wurden verschiedene Netzarchitekturen untersucht, wobei in einer Vorstudie die Stärken einer NARX-Konfiguration für die Prognose thermischer Lasten auf Grundlage historischer Daten infolge ihrer rekursiven Eigenschaften herausgestellt werden konnten. Darüber hinaus eignen sich SVM-R-Modelle aufgrund der hohen erreichbaren Prognosegüte. Für die Lösung nicht-linearer Optimierungsprobleme lassen sich im Zusammenhang mit dieser Methodenkategorie verschiedene Kernel-Funktionen einsetzen, wobei insbesondere der RBF-Kernel

in wissenschaftlichen Arbeiten Ergebnisse von hoher Güte lieferte. Ein direkter Vergleich zwischen NARX RNN und SVM-R-Modellen im Zusammenhang mit ihrer Eignung für datenbasierte, gebäudespezifische Lastprognosen zeigte, dass keine der beiden Modellkategorien auf urbaner Betrachtungsebene ausschließlich vorteilhaft hinsichtlich der Ergebnisgüte der gebäudespezifischen Lastprognosen ist. Bezug nehmend auf die Modellentwicklung wird daher ein Ensemble, bestehend aus zwei NARX RNN unterschiedlicher Tiefe und zweier SVM-R Modelle auf Grundlage unterschiedlicher nicht-linearer Kernel, entwickelt, wobei die Fehlerrate der Lastprognosen in der Modelltestphase automatisiert über den Einsatz eines trainierten Modells für die Lastprognose eines Gebäudes entscheidet. In Tabelle 3.1 sind die Inhalte dieses Abschnitts abschließend zusammengefasst. Die Thematik der Szenarioanalyse umfasst ein breites Spektrum an qualitativen und quantitativen Methoden sowie Kombinationen aus beiden Kategorien und findet in variierender Ausprägung Eingang in die Entwicklung globaler, nationaler und lokaler Energiesysteme, wodurch differenzierte, multiple Analyseebenen zugänglich sind. Es haben sich mit der *Intuitive Logics School*, der *PMT School* und der *French School* im Wesentlichen drei Schulen etabliert, welche ein unterschiedliches Verständnis der Szenariomodellierung vertreten. Im Rahmen dieser Arbeit wird dem Verständnis der *PMT School* gefolgt. In diesem Zusammenhang sind verschiedene Ansätze der Kategorisierung von Methoden zur Generierung von Szenarien identifizierbar. Vor dem Hintergrund einer Integration und Abbildung zusätzlicher Unsicherheiten in den langfristigen urbanen thermischen Lastprognosen finden in dieser Arbeit prädiktive Szenariotechniken Anwendung, welche Variationen in den Prognosen bedingen und die Frage *Was-wäre-wenn* beantworten sollen. Letzteres bezieht sich auf zukünftige klimatische Veränderungen und strukturelle Modifikationen der urbanen Regionen infolge Gebäudesanierungen. Damit kann der Einfluss zukünftiger Entwicklungen auf den Gleichzeitigkeitsfaktor gezeigt und quantifiziert werden, wodurch der Fokus von der ausschließlichen Auswertung historischer Messdaten auf Grundlage zahlreicher Annahmen auf eine szenarienbasierte ganzheitliche Betrachtungsebene verschoben wird. Um den Rechenaufwand für die Einbeziehung verschiedener Szenarien variabel zu halten, werden ferner Freiheitsgrade in Modellentwicklung berücksichtigt.

Tabelle 3.1: Zusammenfassung der Analyseergebnisse zur Eignung der Prognosemethodenkategorien für die Modellbildung in dieser Arbeit

Kategorie	**Algorithmen**	**Besonderheiten**	**Performanz**	**Eignung**
Zeitreihenmodelle	AR, ARMA, ARIMA,...	basieren im Wesentlichen auf historischen Zeitreihenwerten, keine Berücksichtigung von zeitunabhängigen Einflussfaktoren, gute Ergebnisse bei kurzfristigem Zeithorizont	o	-
Kausalmodelle	lineare Regressionsmodelle, abschnittsweise lineare Regressionsmodelle,...	auf Basis geringer Informationsdichte anwendbar, geringe Generalisierungseigenschaften	o	-
Modelle basierend auf maschinellem Lernen	KNN, SVM, DT,...	gute Generalisierungseigenschaften, im Vergleich bessere individuelle Modelladaption, NARX RNN vereinen alle drei Modellkategorien, SVM-R aufgrund Kernel-Funktion sehr robust, Modelltrainingsaufwand	+	+

KAPITEL 4

Klassifikation energetischer Lastprofile

Dieses Kapitel dient der Abbildung des Stands der Technik und der Wissenschaft zur Mustererkennung in Datensätzen. Darauf aufbauend werden geeignete Methoden zur Quantifizierung von Ähnlichkeiten in der Charakteristik gebäudebezogener thermischer Lastgänge auf Stadtquartiersebene identifiziert. Die automatisierte Ableitung von Lastprofilgruppen stellt die Grundlage für die Quantifizierung des Ähnlichkeitsfaktors dar. Darüber hinaus lässt sich das charakteristische aggregierte Lastprofil einer Versorgungsgruppe klassifizierter Gebäudesubgruppen zuordnen. In Kombination mit den Prognosemodellen und vor dem Hintergrund historischer Lastzustände können für vergangene und zukünftige Zeitpunkte Gebäude und Gebäudegruppen ermittelt werden, die zu ähnlichen Zeiten wie die gesamte Versorgungsgruppe einen Spitzenbedarf aufweisen. Dadurch lassen sich die Energieabnehmer(-gruppen) mit dem größten Lastverschiebungspotenzial identifizieren. Abschnitt 4.1 adressiert im Kontext dieses Kapitels Clustering-Lösungen und Abschnitt 4.2 Klassifikationsmodelle, wobei in Abschnitt 4.3 die Ergebnisse zusammengefasst werden.

4.1 Clustering

Im Rahmen dieses Abschnitts wird der theoretische Hintergrund zu Methoden des Clusterings und daran anknüpfend der aktuelle Stand der Forschung im Bereich des Clusterings von Lastprofilen vorgestellt. Beide Aspekte bilden im Folgenden die Grundlage für die Anforderungen an die Teilmodellentwicklung.

4.1.1 Theoretischer Hintergrund

Clustering ist den Methoden des unüberwachten Lernens im Bereich des maschinellen Lernens zugeordnet und stellt ein dynamisches interdisziplinäres Forschungsfeld dar. Mittels Clustering kann *a priori* Wissen aus finiten Datensätzen generiert werden. Die dazugehörigen Algorithmen finden thematisch Eingang in der dazugehörigen Fachliteratur (Hartigan, 1975; Jain et al., 1988; Kaufman et al., 2009) oder im Rahmen des Data Minings (Han et al., 2012), der Mustererkennung sowie des maschinellen Lernens (Bishop, 2006). Weiterhin ermöglichen zusätzliche Studien einen zusammenfassenden Überblick zu verschiedenen Techniken des Clusterings (Jain et al., 1999; Parsons et al., 2004; Jain, 2010). In der Li-

teratur finden eine Vielzahl an Clustering-Algorithmen Anwendung, wobei sich diese in folgende sechs Hauptkategorien einordnen lassen.

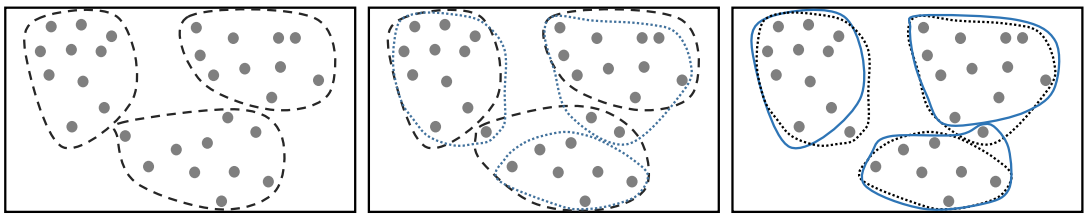
I. Partitionierende Methoden
II. Hierarchische Methoden
III. Dichtebasierte Methoden
IV. Gitterbasierte Methoden
V. Topologiebasierte Methoden
VI. Probabilistische modellbasierte Methoden

Partitionierende Methoden basieren auf der Annahme, dass Objekte in einer vorzugebenden Anzahl an nicht-hierarchisch strukturierter Gruppen zusammengefasst werden können (Han et al., 2012). Ein iterativer Veränderungsprozess der Gruppenzusammensetzungen zielt im Optimierungsprozess auf die Verbesserung der Objektpartitionierung ab. Hierarchische Methoden adressieren die Anordnung von Objekten mittels eines Clusterbaums, wobei beispielsweise die euklidische Distanz zwischen den zu gruppierenden Objekten als Kriterium für den Aufbau einer Hierarchie dient (Han et al., 2012). Auf der untersten Ebene des Clusterbaums bilden die Objekte jeweils ein Cluster und werden beispielsweise auf Grundlage ihrer geringen Distanz zu anderen Objekten sukzessive zu größeren Clustern auf den darüber liegenden Ebenen zusammengefasst. Dichtebasierte Methoden ermöglichen im Gegensatz zu den beiden letztgenannten Methodenkategorien die Bildung von Clustern beliebiger Formen auf Grundlage der Objektdichte im Raum. Hierbei ist die maximale Objektanzahl für die Cluster vorzugeben. Gitterbasierte Methoden beruhen auf der Segmentierung des Objektraumes mittels einer zu definierenden Gitterstruktur. Diese ermöglicht die effiziente Suche nach Clustern mit beispielsweise dichtebasierte Methoden, da die Rechenzeit nicht mehr von der Anzahl an Objekten, sondern von der Anzahl an Zellen innerhalb der Gitterstruktur abhängt. (Han et al., 2012) Topologiebasierte Methoden zielen auf die Darstellung hochdimensionaler Datensätze im zwei- bzw. dreidimensionalen Raum ab, wobei Referenzvektoren für Vektorgruppen als Clusterrepräsentanten auf Basis von Abstandsmaßen bestimmt werden (Xu et al., 2015). Probabilistisch modellbasierte Methoden erlauben multiple Clusterzugehörigkeiten von Objekten (Han et al., 2012). Dabei lässt sich der Zugehörigkeitsgrad beispielsweise im Sinne des partitionierenden Konzepts mittels der euklidischen Distanzen eines Objekts zu beliebig gewählten Clusterzentren bestimmen.

Partitionierende Methoden

Mittels partitionierender Methoden werden n Objekte eines Datensatzes X in k Gruppen unterteilt, wobei jeder Gruppe mindestens ein Objekt zugeordnet sein muss, sodass die Bedingung $k \leq n$ erfüllt ist. Methoden dieser Kategorie basieren in der Regel auf Distanzmaßen einzelner Objekte zu den Gruppen, welche auf Grundlage von zu bestimmenden Anfangswerten für k berechnet werden. In einem iterativen Prozess der Verschiebung von

Objekten in andere Gruppen erfolgt die sukzessive Verbesserung des Clusterings. Abbildung 4.1 veranschaulicht diesen Ablauf. (Han et al., 2012)



i. Initiales Clustering ii. Iteration iii. Finales Clustering

Abbildung 4.1: Prinzip des partitionierenden Clusterings am Beispiel des k-Means Algorithmus' in Anlehnung an Han et al., 2012

Für die Beurteilung der Cluster-Güte sind zahlreiche Kriterien zu identifizieren. Im Allgemeinen ist ein Clustering hierbei als gut einzustufen, wenn die Objekte eines Clusters nah beieinander liegen, während andere Cluster weit entfernt sind oder sich die dazugehörigen Datenpunkte deutlich unterscheiden. Partitionierende Methoden sind vornehmlich für Datensätze kleiner und mittlerer Größe geeignet, um kugelförmige Cluster zu identifizieren. (Han et al., 2012) Im Ergebnis wird ein lokales Optimum erreicht, wobei unter der Voraussetzung erheblichen Rechenaufwands auch ein globales Optimum möglich ist. Als Beispiele für partitionierende Clustering-Methoden sind der k-Means (MacQueen et al., 1967), k-Medoids (Estivill-Castro et al., 2000) oder k-Modes (Huang, 1998) Algorithmus zu nennen. Für die Anwendung partitionierender Clustering-Algorithmen ist das *a priori* Wissen um die Clusteranzahl Voraussetzung (Xu et al., 2015). Eine Abwandlung des k-Means-Algorithmus stellt die *Adaptive Vector Quantification* Methode dar (Tsekouras et al., 2007). Letztere zählt zu den kompetitiven KNN und ähnelt der SOM Methode, welche im weiteren Verlauf dieses Kapitels noch vorgestellt wird.

Hierarchische Methoden

Hierarchische Clustering-Methoden ermöglichen die Gruppierung von Daten in Hierarchien oder Cluster-Bäumen, wodurch Strukturen in Datensätzen sichtbar werden. Häufig dient dabei ein Dendrogramm als Visualisierung. Abbildung 4.2 veranschaulicht dazu ein Beispiel mit sechs zu clusternden Objekten a_1 bis a_6, wobei ein Ähnlichkeitsmaß, wie beispielsweise die minimale euklidische Distanz, für die Gruppierung der Objekte zu wählen ist. Weiterhin existieren in algorithmischen, probabilistischen und Bayes'schen Methoden Unterkategorien des hierarchischen Clusterings. Mittels algorithmischer Methoden werden die Datenobjekte und die Distanzen zwischen diesen als deterministisch betrachtet. Dabei werden den algorithmischen Ansätzen agglomerative, trennende und mehrphasige Ansätze zugeordnet, welche sich in ihrer Strategie zur sukzessiven Fusionierung von Objekten zu Clustern unterscheiden.

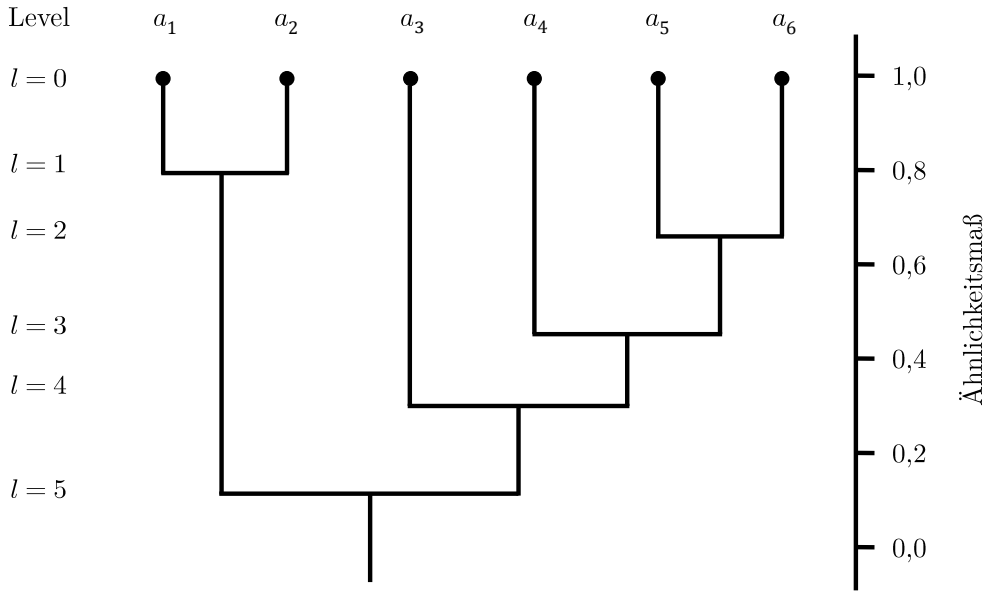


Abbildung 4.2: Prinzip des hierarchischen Clusterings in Anlehnung an Han et al., 2012

Während mit agglomerativen Methoden eine bottom-up-Strategie verfolgt wird, indem die Objekte, welche zunächst ihr eigenes Cluster bilden, iterativ zu Clustern zusammen gefasst werden, bis ein Abbruchkriterium erfüllt ist, basieren trennende Methoden darauf, das initialisierte eine Cluster, welches alle Objekte enthält, rekursiv zu partitionieren. (Han et al., 2012) Beide Methoden vereint der Bedarf an einer vorgegebenen Clusteranzahl oder einer minimal zulässigen Distanz zwischen Clustern als Abbruchkriterium (Chicco et al., 2006). Nicht zuletzt muss ein geeignetes Distanzmaß gewählt werden, wodurch sich verschiedene Varianten des hierarchischen Clusterings ableiten lassen (Lance et al., 1967; Murtagh, 1983; Jain et al., 1988; Yager, 2000) Darüber hinaus finden auch mehrphasige Methoden, wie *BIRCH* (Zhang et al., 1996), *Chameleon* (Karypis et al., 1999) oder modifizierte Algorithmen basierend auf dem traditionellen hierarchischen Clustering, wie *CURE* (Guha et al., 1998), Anwendung. Eine weitere Variante des hierarchischen Clusterings stellt die Methode *Follow-the-Leader* dar. Hierbei wird ein zentrales Objekt des Datensatzes als *Leader* einer Gruppe definiert. Objekte, deren Abstand zum *Leader* am geringsten ist, werden als *Follower* dieser Gruppe zugeordnet. Dieser Algorithmus unterscheidet sich von den traditionellen hierarchischen Clustering Methoden in der Eigenschaft, kein *a priori* Wissen bezüglich der Clusteranzahl vorauszusetzen. Gleichwohl muss eine Grenze für die Zuordnung weiterer Mitglieder eines Clusters durch die Festlegung ihres Abstands zum *Leader* vorgegeben werden. (Wu et al., 2013) Als Nachteile des hierarchischen Clusterings sind primär die Bestimmung der optimalen Clusteranzahl und die hohen Rechenzeiten im Vergleich zu

den partionierenden Clustering-Methoden angeführt (Xu et al., 2015). Demgegenüber sind sehr detaillierte Clusterzuordnungen und -abgrenzungen zwischen den Clustern möglich (Chicco et al., 2006).

Dichtebasierte Methoden

Während sich mit partionierenden und hierarchischen Clustering-Methoden kugelförmige Cluster in Datensätzen identifizieren lassen, ermöglichen dichtebasierte Clustering-Algorithmen die Ermittlung von Clustern mit beliebigen Formen (Han et al., 2012). Die Idee hinter dem dichtebasierten Clustering beruht auf der sukzessiven Vergrößerung eines Clusters und damit der Verschiebung der Clustergrenze bis eine Abbruchbedingung in Form einer maximalen Objektanzahl innerhalb des Clusters erreicht ist. Letztere entspricht in diesem Zusammenhang der Objektdichte innerhalb der betrachteten Datenpartition. (Han et al., 2012) Dabei muss sich in der Umgebung eines Objektes eine initial zu bestimmende minimale Anzahl an anderen Objekten befinden, damit ein dichtebasiertes Cluster definiert werden kann (Xu et al., 2015). Abbildung 4.3 veranschaulicht das Prinzip des dichtebasierten Clusterings. Bezugnehmend auf Abbildung 4.3 und das prinzipielle Vorgehen, werden ausgehend von den Datenpunkten a_1 und a_4 als Startpunkte bei vorgegebenem Radius um diese Punkte herum weitere Datenpunkte in der Nachbarschaft gesucht, deren Distanz zu a_1 und a_4 maximal dem Radius um letztere entsprechen. Erfüllen beispielsweise in diesem Fall mindestens drei Datenpunkte in der Nachbarschaft von a_1 und a_4 diese Bedingung, werden in der Nachbarschaft von letzteren weitere Datenpunkte gesucht, welche zu dem Cluster gehören könnten. Falls die ermittelte minimale Datenpunktanzahl nicht der Vorgabe entspricht, erfolgt die Bildung eines neuen Clusters für a_1. (Han et al., 2012) Als Beispiele für dichtebasierte Clustering-Algorithmen sind *DBSCAN* (Ester et al., 1996) oder *CLARANS* (Ng et al., 2002) zu nennen. Dabei konnte die Vorteilhaftigkeit dieser Clustering-Methoden im Vergleich zu partionierenden oder hierarchischen Clustering-Algorithmen in Bezug auf die Identifizierung von Gruppen innerhalb großer Datensätze nachgewiesen werden (Ester et al., 1996). Ein weiterer Algorithmus stellt *DENCLUE* dar, welcher im Vergleich zu den beiden letztgenannten Methoden vornehmlich geringere Rechenzeiten für das dichtebasierte Clustering benötigt (Hinneburg et al., 1998). Mit *OPTICS* wurde von Ankerst et al. ein weiterer Algorithmus eingeführt (Ankerst et al., 1999). Ein Forschungsschwerpunkt liegt in der Effizienzsteigerung der Algorithmen, um beispielsweise Overfitting Effekte zu vermeiden (Xiao-Feng Wang et al., 2009). Dichtebasierte Methoden benötigen die Vorgabe einer minimalen Anzahl an Objekten in der Umgebung des Clusterzentrums und nicht das *a priori* Wissen um die Clusteranzahl (Xu et al., 2015).

Gitterbasierte Methoden

Bei gitterbasierten Methoden erfolgt die Überführung des Objektraums in eine Gitterstruktur, welche durch eine finite Zellenanzahl aufgespannt wird, wobei das Clustering stets auf der Gitterstruktur beruht. Abbildung 4.4 visualisiert das Prinzip dahingehend. Die Clusterobjekte fallen dabei in die Zellen, welche das Intervall aufweisen, in dem sich die

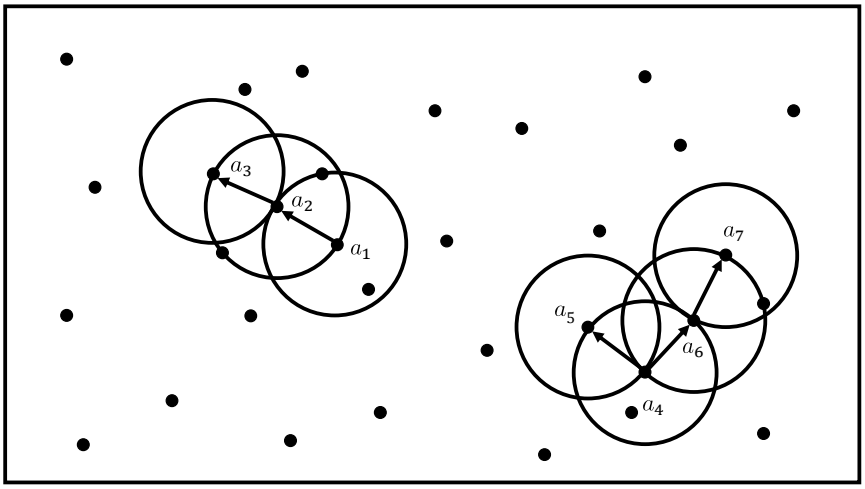


Abbildung 4.3: Prinzip des dichtebasierten Clusterings in Anlehnung an Ester et al., 1996

Objektwerte befinden. Über diese Zellen wird das Cluster, häufig mit einem dichtebasierten Ansatz, gebildet. Mit gitterbasierten Methoden lässt sich eine geringe Rechenzeit erreichen, da diese wiederum ausschließlich von der gewählten Anzahl an Gitterzellen und nicht mehr von der Anzahl der Objekte abhängt. Dahingehend sind gitterbasierte Methoden auch für die kombinierte Anwendung mit dichtebasierten und hierarchischen Clustering Methoden geeignet. (Han et al., 2012) Zu gitterbasierten Methoden gehören weiterhin beispielsweise *STING* (Wang et al., 1997), *STING+* (Wei Wang et al., 1999), *GDILC* (Zhao Yanchang et al., 2001), *CLIQUE* (Agrawal et al., 1998) oder *WaveCluster* (Sheikholeslami et al., 2000). Ferner haben Ma et al. eine gitterbasierte Clustering-Methode mit sich verschiebenden Gittern eingeführt (W.M. Ma et al., 2004).

Topologiebasierte Methoden

Topologiebasierte Clustering-Algorithmen beruhen auf Techniken zur Datenvisualisierung im zwei- und dreidimensionalen Raum (Bunte et al., 2010; van der Maaten et al., 2008; Si et al., 2000). Dieser Kategorie wird neben der statistischen multivariaten Technik *MANOVA* (Verdu et al., 2006) der SOM-Algorithmus von Teuvo Kohonen (Kohonen, 1981) zugeordnet, welcher hochdimensionale Datensätze mittels eines niedrigdimensionalen kompetitiven KNN auf planarer Ebene visualisiert und repräsentiert (Wu et al., 2012). Abbildung 4.5 veranschaulicht dahingehend das Prinzip der SOM. Jeder beliebige Datenpunkt x_j eines hochdimensionalen Eingangsraumes lässt sich durch einen der k Referenzvektoren (Neuronen) repräsentieren. Mittels der im meist zweidimensionalen Raum gitterförmig angeordneten Neuronentopologie erfolgt ein Adaptionsprozess, ausgelöst durch Signale. Dazu wird in jedem Iterationsschritt zufällig ein Eingangsvektor gewählt, dessen Distanz

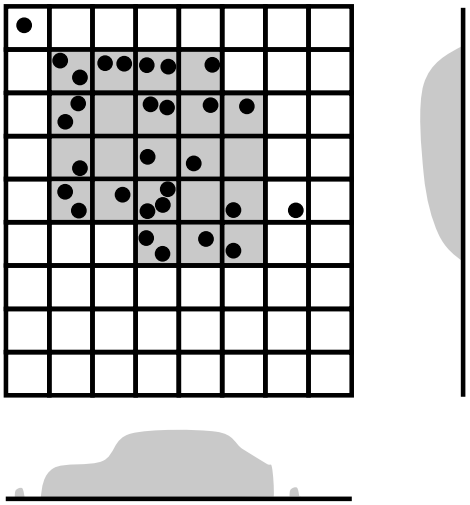


Abbildung 4.4: Prinzip des gitterbasierten Clusterings in Anlehnung an Han et al., 2012

zu dem Gewicht jedes Neurons ermittelt wird. Die Neuronen stehen dahingehend in einem Wettbewerb zueinander, wobei jenes Neuron mit dem geringsten Abstand zum Eingangsvektor gewinnt. Anschließend werden das Gewicht des Gewinnerneurons und die Gewichte der dazugehörigen Nachbarneuronen aktualisiert. Dadurch lässt sich der Eingangsraum in verschiedene Regionen aufteilen. (Kohonen, 1990b; Wu et al., 2004)

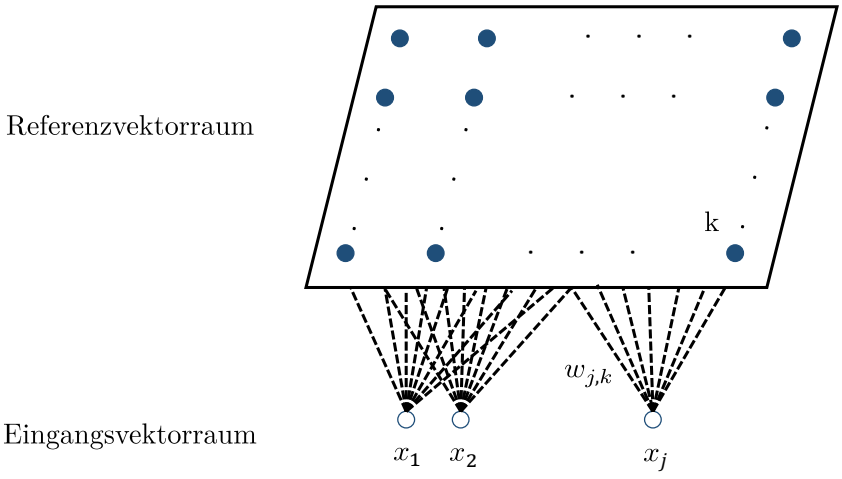


Abbildung 4.5: Prinzip der SOM von Teuvo Kohonen in Anlehnung an Enami et al., 2010

Probabilistische modellbasierte Methoden

Im Gegensatz zu bisher vorgestellten Clustering-Algorithmen ermöglichen probabilistische modellbasierte Methoden multiple Clusterzugehörigkeiten. Dieser Kategorie lassen sich beispielsweise Fuzzy Clustering-Algorithmen zuordnen, welche eine *weiche* Allokation von Objekten zu Clustern erlauben. Im Speziellen gehören alle Objekte gleichzeitig jedem Cluster zu einem bestimmten Grad an, wodurch Objekte nicht mehr zwangsläufig exklusiv einem Cluster zugeordnet werden. (Han et al., 2012) Daher sind diese Methoden für Datensätze geeignet, welche überlappende Clustergrenzen aufweisen (Xu et al., 2015). Mit Fuzzy-Methoden lassen sich somit Sensitivitäten in die Ergebnisse der Ermittlung von Clusterexistenzen und -abgrenzungen integrieren (Dunn, 1973). In diesem Zusammenhang ist der Fuzzy-c-Means-Algorithmus (Dunn, 1973) zu nennen, welcher auf die Identifizierung von Datenpartitionen auf Grundlage der Minimierung einer Kostenfunktion abzielt. Letztere ist durch die Zugehörigkeitskoeffizienten auf Grundlage der Annahme einer anzunehmenden Clusteranzahl definiert (Xu et al., 2015).

Die verschiedenen Clustering-Algorithmen zeichnen sich durch unterschiedliche Einsatzmöglichkeiten aus, wobei eine für die individuelle Analyseaufgabe geeignete Methode gewählt werden muss. Um verschiedene Modellcharakteristika zu kombinieren, finden auch Zwei-Ebenen-Ansätze Anwendung. Rodrigues et al. haben beispielsweise den SOM- mit dem k-Means-Algorithmus verknüpft. Mittels SOM erfolgte die Bestimmung der Clusteranzahl, welche im Bereich zwischen 6 und 9 angenommen wurde. Die detaillierte Cluster-Zuordnung basierte auf den Ergebnissen des Clusterings mit k-Means. (Rodrigues et al., 2003) Darüber hinaus wurde die Kombination aus SOM und hierarchischen Clustering-Algorithmen untersucht (Wu et al., 2004; Taşdemir et al., 2011).

4.1.2 Studien

Die Gruppierung von Lastprofilen mittels Methoden des Clusterings findet eine breite Anwendung in Studien zur Charakterisierung von Energieversorgungsgruppen. Dabei sind im Wesentlichen die folgenden zwei Kategorien zu unterscheiden.

I. Hierarchisches Clustering

II. Nicht-hierarchisches Clustering

Die Ähnlichkeitsanalysen zielen auf die Identifizierung von charakteristischen Abnehmergruppen ab. In diesem Zusammenhang erfolgt eine Unterscheidung in Jahreszeiten, Wochentage, Wochenenden oder Ferienzeiten (Chicco et al., 2001; Bobric et al., 2009; McLoughlin et al., 2015). In zahlreichen Untersuchungen dienen die Ergebnisse des Clusterings der Informationsanreicherung von Klassifikationsalgorithmen (In Hyeob Yu et al., 2005) und Prognosemodellen (Duan et al., 2011; Mori et al., 2001; Jain et al., 2009; Chaouch, 2014; Nazarko et al., 2005).

Weiterhin werden verschiedene Clustering-Methoden einander vergleichend gegenüber gestellt (Chicco, 2012; Labeeuw et al., 2013). Dabei finden k-Means (Bobric et al., 2009; Granell et al., 2015; McLoughlin et al., 2015; Chicco, 2012) und Fuzzy-C-Means-Algorithmen (Duan et al., 2011; Sathiracheewin et al., 2011; Wang et al., 2013; Zakaria et al., 2006) eine breite Anwendung. Gleiches ist für die Nutzung des *Follow-the-Leader*-Algorithmus festzustellen (Chicco et al., 2004; Chicco et al., 2005; Chicco et al., 2006). SOM werden in den Studien von Chicco et al. und McLoughlin et al. hinsichtlich ihrer Eignung zum Clustering von Verbrauchslastprofilen untersucht (Chicco et al., 2006; Chicco et al., 2001; McLoughlin et al., 2015). Hinsichtlich der Ergebnisgüte zeigt sich ein uneinheitliches Bild, jedoch mit einzelnen erkennbaren Tendenzen. So kann aus den anwendungsbasierten Studien festgehalten werden, dass mit dem k-Means-Algorithmus eine sinnvolle und robuste Gruppierung von Lastprofilen möglich ist (Bobric et al., 2009; Granell et al., 2015). Die Anwendung des SOM-Algorithmus ermöglicht zuverlässige Clustering Ergebnisse (Chicco et al., 2001; McLoughlin et al., 2015). Demgegenüber existieren auch Forschungsarbeiten, welche dies lediglich teilweise bestätigen und den SOM-Algorithmus für das initiale prototypische Clustering vorschlagen, um anschließend andere Clustering-Methoden für finale Gruppenzuordnungen zu verwenden (Chicco et al., 2006). Diesbezüglich zeigte sich beispielsweise, dass eine Kombination aus k-Means- und SOM-Algorithmus bessere Ergebnisse liefert als bei einer isolierten Anwendung der beiden Methoden (Rodrigues et al., 2003).

Hinsichtlich der Verwendung des hierarchischen Clusterings zur Gruppierung von energetischen Lastprofilen sind sowohl Ergebnisse zu finden, welche eine Überlegenheit dieser Methoden gegenüber anderen Clustering-Algorithmen implizieren (Chicco et al., 2006) als auch gegensätzliche Ergebnisse (Granell et al., 2015). Zahlreiche Studien basieren auf der Untersuchung von Fuzzy-Methoden (Zakaria et al., 2009). Gleichwohl wurde im Zusammenhang mit letzteren die starke Abhängigkeit der Ergebnisse von Fehlern im Datensatz herausgestellt (Zalewski, 2006), welche jedoch beispielsweise durch die Abwandlung des Fuzzy-C-Means-Algorithmus (Sathiracheewin et al., 2011) oder durch eine spezielle Datenauswahl (Semeraro et al., 2014) reduziert werden kann.

4.1.3 Implikationen und Anforderungen an die Modellierung

Die Vielzahl an Clustering-Methoden und Kombinationen aus jenen bedingt detailliertere Analysen der Algorithmen sowie jeweiliger Eingangs- und Ausgangsparameter. Im Sinne der Automatisierung von Datenanalysen stellt sich vordergründig die Frage, welche Eingangsparameter von Algorithmen selbstständig bestimmt werden können. Um die Ähnlichkeit von Lastprofilen automatisiert quantifizieren zu können, bedarf es der Ermittlung der Clusteranzahl, welche zunächst unbekannt ist. Clustering-Algorithmen wie k-Means benötigen jedoch die Clusteranzahl zur Initialisierung. Hierarchisches Clustering basiert auf einer Baumstruktur und bedarf einer Abbruchbedingung, auf dessen Grundlage definiert wird, welche Subcluster wieder zu übergeordneten Clustern zusammengefasst werden sollen. Mit der Festlegung auf ein bestimmtes Abbruchkriterium verliert der Modellansatz an dieser Stelle seine Generalisierungseigenschaft, da das gewählte Kriterium nicht für jeden

Datensatz plausibel sein muss. Demgegenüber sind bei SOM die Anzahl der Neuronen vorzugeben, um die Dimension des Kohonen-Netzes bestimmen zu können. Die Anzahl der Neuronen korreliert dabei nicht zwangsläufig mit der Clusteranzahl, sodass Variationen der Eingangsdaten auch mit Variationen der Clusteranzahl einhergehen können. Dennoch sind in der Modellierung verschiedene Netzdimensionen vorzusehen, um Unterschiede in der Clustergranularität analysieren zu können.

4.2 Klassifikation

Im Anschluss an das Clustering von unbekannten Lastprofilen und der damit verbundenen Vorgruppierung dienen Klassifikationsmodelle der Bildung von charakteristischen Gruppen auf Grundlage der ermittelten Cluster und der Zuordnung eines neuen Lastprofils in die Klasse mit der größten Ähnlichkeit. Im Folgenden wird der theoretische Hintergrund und der aktuelle Stand der Technik sowie der Wissenschaft zu verschiedenen Methodenkategorien vorgestellt.

4.2.1 Theoretischer Hintergrund

Klassifikationsmethoden sind den überwachten Lernverfahren zuzuordnen und beruhen auf dem Ziel, Daten zu kategorisieren. Auf Grundlage der trainierten Klassifikationsmodelle lassen sich überdies Prognosen in Bezug auf neue Datengrundlagen stellen. (Han et al., 2012) Im Sinne einer methodischen Abgrenzung verschiedener Klassifikationsalgorithmen voneinander sind nachfolgende Hauptkategorien definierbar.

I. Hierarchische Methoden
II. Perzeptron-Methoden
III. Support Vector Machines
IV. Statistische Methoden
V. Ensemble Methoden

Hierarchische Methoden basieren in der Regel auf Baumstrukturen mit Knoten, Kanten und einer Wurzel als Ausgangspunkt. An jedem Knotenpunkt eines Entscheidungsbaums, welcher kein Blatt und damit ein Endpunkt eines Pfades dargestellt, ist ein zu definierendes Attribut zu bewerten. Das Ergebnis entscheidet über den weiteren Pfad durch die Struktur bis ein Blatt und damit eine Entscheidung bezüglich der Klassifikation des Eingangsdatensatzes erreicht ist. Perzeptron Methoden folgen dem Prinzip des KNN, das zusammen mit dem Konzept der Support Vector Machine bereits in Kapitel 3 vorgestellt wurde. Mit statistischen Methoden erfolgt im Gegensatz zu den zuvor genannten Kategorien die Quantifizierung von Wahrscheinlichkeiten von Klassenzugehörigkeiten gegebener Eingangsdaten. Ensemble Methoden basieren auf der Annahme, dass mehrere schwache Klassifikationsmodelle, welche mit einer geringeren Genauigkeit Klassifikationen von Daten vornehmen, kombiniert miteinander robuste Klassifikationsergebnisse ermöglichen. (Han et al., 2012)

Hierarchische Methoden

Im Zusammenhang mit hierarchischen Klassifikationsmodellen sind Entscheidungsbäume zu nennen. Abbildung 4.6 visualisiert diese exemplarisch.

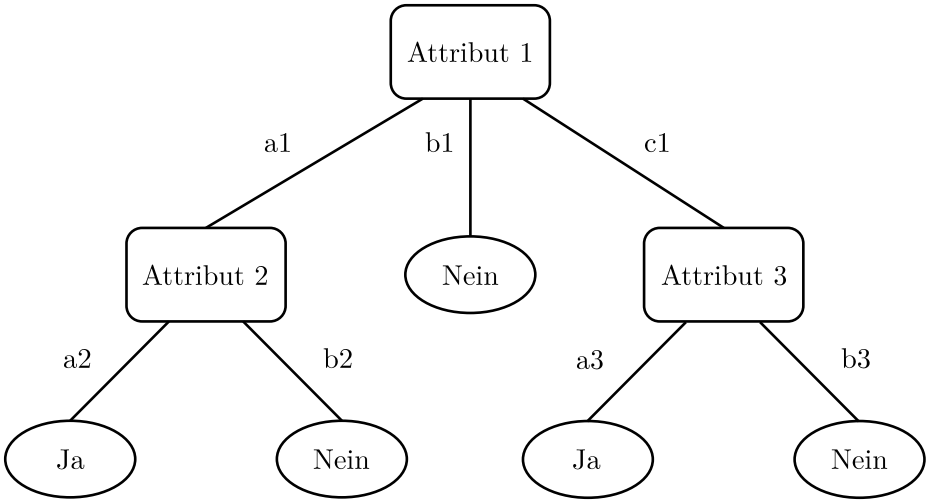


Abbildung 4.6: Prinzip der Klassifikation mit Entscheidungsbäumen in Anlehnung an Kotsiantis et al., 2007

Jedes Blatt innerhalb der Baumstruktur stellt ein *Feature* innerhalb einer zu klassifizierenden Instanz dar und jedem Ast ist ein Wert zugeordnet (Kotsiantis et al., 2007). Im Zusammenhang mit dieser Modellkategorie existieren verschiedene Algorithmen, wie *CART*, *ID3*, *C4.5* oder *C5.0* (Jadhav et al., 2016). Entscheidungsbäume sind leicht zu interpretieren und bieten eine hohe Klassifikationsgenauigkeit. Als Nachteil sind unter anderem die hohe Modelltrainingsdauer und Probleme bei diagonalen Klassengrenzen zu nennen (Bhavsar et al., 2012).

Perzeptron-Methoden

Perzeptron-Methoden umfassen KNN und beinhalten verschiedene Architekturen, deren Kategorien im Rahmen der verschiedenen Prognosemodelle in Kapitel 3 bereits vorgestellt wurden. Beginnend bei den Perzeptron-Modellen mit einer verdeckten Schicht sind beispielsweise die Algorithmen *WINNOW* (Littlestone, 1988) oder *voted-perceptron* (Freund et al., 1999) zu nennen. Mehrschichtige Perzeptron-Architekturen bedingen in der Regel hohe Rechenzeiten für die Klassifikation von Datensätzen, wodurch verschiedene Ansätze zur Verringerung der Trainingszeiten infolge der Optimierung von Strukturen und Parametern untersucht wurden (Kotsiantis et al., 2007). Yen & Lu haben einen hierarchischen genetischen Algorithmus eingeführt, um Parameter und die Netztopologie zu bestimmen

(Yen et al., 2000). Ferner existieren auch verschiedene Bayessche Methoden für das Training von KNN (Vivarelli et al., 2001). Im Kontext der mehrschichtigen Perzeptron-Modelle stehen auch Netzwerke mit radialer Basisfunktion für die Klassifikation von Datensätzen im Fokus der Untersuchung verschiedener Netzarchitekturen (Howlett et al., 2013). Perzeptron-Methoden zeichnen sich insbesondere durch eine hohe Robustheit gegenüber Datensätzen mit Rauschen aus und sind vorteilhaft bei kontinuierlichen Eingangs- und Ausgangsdaten. Nachteile ergeben sich bezüglich der geringen Interpretationsmöglichkeiten der Ergebnisse, beispielsweise der einzelnen trainierten Gewichtungsfaktoren der Kanten des Netzwerkes. (Han et al., 2012)

Support Vector Machines

Mittels SVM lassen sich lineare und nicht-lineare Klassifikationsprobleme lösen, wobei Hyperebenen verschiedene Klassen voneinander trennen und der Abstand der Hyperebene zu den einzelnen Instanzen einer Klasse zu maximieren ist (Kotsiantis et al., 2006). Abbildung 4.7 zeigt diesen Zusammenhang.

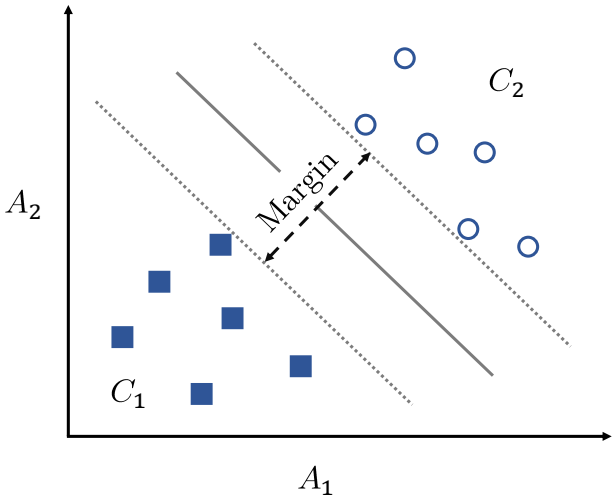


Abbildung 4.7: Prinzip der Klassifikation mit SVM in Anlehnung an Han et al., 2012

Infolge der Abbildung des Eingangsdatensatzes in einer ausreichend hohen Dimension, lässt sich für jedes Klassifikationsproblem eine Klassengrenze definieren (Han et al., 2012). Vorteile der SVM ergeben sich durch die Robustheit gegenüber Ausreißern in Datensätzen und durch die Möglichkeit der Abbildung komplexer nichtlinearer Klassengrenzen. Demgegenüber wird viel Zeit für das Modelltraining benötigt (Han et al., 2012). Zur Lösung nicht-binärer Klassifikationsprobleme existieren überdies erweiterte SVM-Modelle. Ein intuitiver Ansatz basiert auf der Lösung multipler binärer Klassifikationsprobleme, wobei jede Klasse gegen die verbleibenden Klassen (*One-Versus-Rest*) oder einzelne Klassen einander gegenüber gestellt werden (*One-Versus-One*) (Weston et al., 1999). Eine Modifi-

kation des klassischen *One-Versus-One*-Ansatzes stellt das *DAGSVM*-Modell dar, welches, abweichend von der herkömmlichen Methode, in der Testphase auf einem gerichteten azyklischen Graphen beruht, dessen Knoten binäre SVM repräsentieren (Platt et al., 2000). Ein Vorteil dieser Methode liegt in verbesserten Generalisierungseigenschaften und einer verringerten Modelltestzeit begründet (Hsu et al., 2002).

Statistische Methoden

Statistische Methoden für die Klassifikation basieren auf Wahrscheinlichkeitsmodellen, wodurch Wahrscheinlichkeiten für Klassenzugehörigkeiten vorliegen (Kotsiantis et al., 2007). In diesem Zusammenhang sind Linear Discriminant Analysis (LDA) (Friedman, 1989) und Fisher's linear discriminants aufzuführen (Mika et al., 1999). Auf Grundlage beider Techniken werden lineare Trennfunktionen für die Separierung zweier oder multipler Klassen voneinander gesucht (Bishop, 2006). Im Rahmen dieser Methodenkategorie ist überdies das Konzept der Learning Vector Quantization (LVQ) zur Definition von Klassengrenzen zu nennen (Kohonen, 2001). Hierbei werden Referenzvektoren ähnlichen Klassen und somit ähnlichen Quantisierungsregionen zugeordnet, wobei Hyperebenen Quantisierungsregionen voneinander trennen. Die euklidische Distanz zwischen einem Eingangsvektor und dem dichtesten Referenzvektor dient dabei als Maß für die gleiche Klassenzugehörigkeit. Im Zusammenhang mit LVQ existieren mit LVQ1, LVQ2 oder LVQ3 verschiedene Algorithmen für die Mustererkennung in Datensätzen, welche unterschiedlichen Philosophien unterliegen. Im Hinblick auf die Ergebnisgüte unterscheiden sich die Algorithmen kaum voneinander. Während LVQ1 auf der Prämisse beruht, im Lernprozess jeweils einen Referenzvektor zu modifizieren, werden im Sinne der LVQ2 und LVQ3 Konzepte zwei Referenzvektoren im Hinblick auf überlappende Klassengrenzen angepasst. (Kohonen, 2001) Abbildung 4.8 visualisiert diesen Zusammenhang exemplarisch. LVQ bietet zahlreiche Vorteile gegenüber Perzeptron-Methoden und Support Vector Machine (SVM) (Nova et al., 2014).

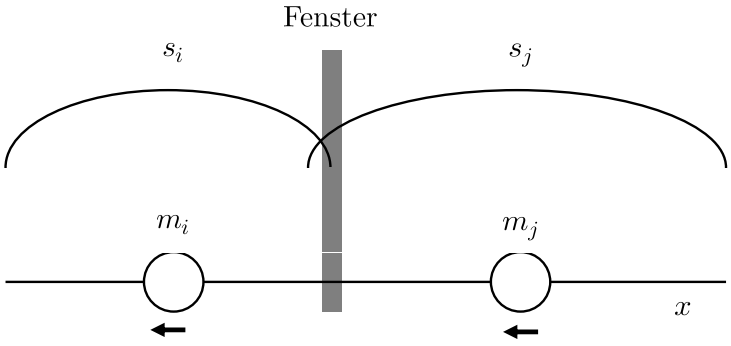


Abbildung 4.8: Schematische Darstellung von Referenzvektoren mit den dazugehörigen Klassen im Rahmen von LVQ2 in Anlehnung an Kohonen, 1990a

Während letztgenannte aufgrund ihrer Black-Box-Eigenschaft schwierig zu interpretieren sind, lassen sich unter Anwendung von LVQ repräsentative Vektoren mit dazugehörigen Klassenregionen (Voronoi-Partitionen) identifizieren. Diese Vektoren stellen für einen gegebenen Datensatz maßgebliche Vektoren dar, wohingegen beispielsweise Stützvektoren (SVM) Extremvektoren innerhalb eines Datensatzes entsprechen (Hofmann et al., 2012; Hammer et al., 2004). Im Hinblick auf den Rechenaufwand ist LVQ als vorteilhaft einzustufen, da dieser von der vorzugebenden festen Anzahl an repräsentativen Vektoren abhängt, während die Rechenzeit der SVM-Modelle von den Dimensionen des Trainingsdatensatzes abhängt (Nova et al., 2014).

Ensemble Methoden

Im Rahmen von Ensemble Methoden werden mehrere Klassifikatoren in einem Modell miteinander kombiniert. Jeder Klassifikator gibt dabei eine *Stimme* für eine Klassenzugehörigkeit ab, sodass eine Stimmensammlung im Ergebnis entsteht. (Han et al., 2012) In Abbildung 4.9 ist dieses Prinzip auf Grundlage eines Eingangsdatensatzes X mit n Trainingsdatensätzen dargestellt. Im Zusammenhang mit dieser Methodenkategorie ist der Algorithmus *Bagging* von Breiman zu nennen (Breiman, 1996).

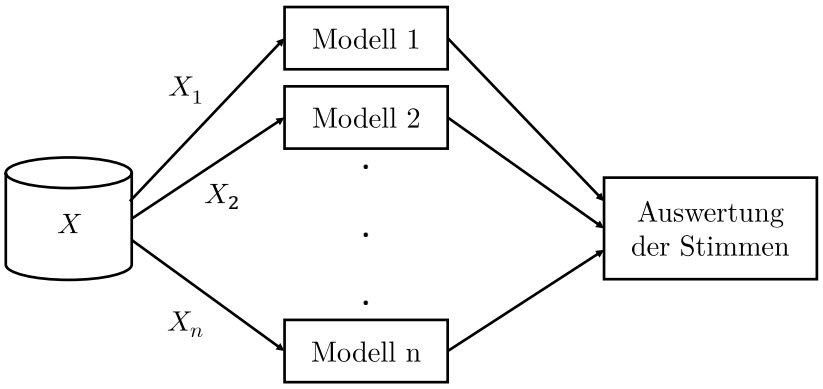


Abbildung 4.9: Prinzip der Ensemble Methoden in Anlehnung an Han et al., 2012

Dieser kombiniert verschiedene Klassifikatoren miteinander und gibt die Klasse des Eingangsdatensatzes in Abhängigkeit von der Anzahl an individuellen Stimmen für die Klasse innerhalb des Ensembles an (Han et al., 2012). Eine weitere Technik in diesem Kontext stellt *Boosting* dar, wobei *AdaBoost* (Freund et al., 1997) eine sehr breite Anwendung findet (Bishop, 2006). *Boosting* folgt wie *Bagging* der Idee, n Klassifikatoren zu trainieren. Im Unterschied zu *Bagging* werden beim *Boosting* alle Datenpunkte gewichtet, wobei die Gewichtungsfaktoren von der Performanz der zuvor trainierten Klassifikatoren abhängt. (Bishop, 2006) Die Gewichtungen der falsch klassifizierten Datenpunkte werden erhöht, wodurch der Fokus für den folgenden Trainingsdurchlauf auf die falsch klassifizierten

Datenpunkte gelegt wird (Han et al., 2012). Im Bereich der hierarchischen Methoden stellen *Random Forests* eine Erweiterung des Konzepts der Entscheidungsbäume dar. Erstere definieren sich durch eine Menge an unkorrelierten Entscheidungsbäumen, wobei im Klassifikationsprozess jeder Entscheidungsbaum für eine Klasse stimmt. Die endgültige Entscheidung für eine Klasse basiert auf dem Mehrheitsprinzip. (Han et al., 2012)

Im Hinblick auf die Auswahl geeigneter Klassifikationsmethoden sind ferner die Anwendungslimitationen der Algorithmen zu berücksichtigen. Viele Konzepte wurden für die Lösung binärer Klassifikationsprobleme entwickelt. Mehrklassen-Klassifikationen erfordern dagegen Modellerweiterungen. In diesem Kontext ist zwischen One-Versus-All (OVA) und All-Versus-All (AVA) zu unterscheiden (Han et al., 2012). Farid et al. haben in diesem Zusammenhang hybride Entscheidungsbäume und naive Bayes Klassifikationsmethoden entwickelt (Farid et al., 2014). Für Mehrklassen-Klassifikationsprobleme eignet sich darüber hinaus insbesondere auch LVQ, da bei dieser Methode implizit alle Klassengrenzen gleichzeitig modelliert werden (Lloyd et al., 2007).

4.2.2 Studien

Klassifikationstechniken werden im Rahmen von anwendungsbasierten Studien in verschiedenen Bereichen eingesetzt. Molina-Solana et al. geben dazu in ihrer Studie zu Data Science Methoden für das Gebäudeenergiemanagement (Molina-Solana et al., 2017) einen umfassenden Überblick. In diesem Kontext sind vornehmlich hierarchische Methoden zu nennen, die in Form von Entscheidungsbäumen Anwendung finden (Wu et al., 2008). Mit der Attribuierung verschiedener Energieabnehmerklassen wird das Ziel verfolgt, neue Abnehmer ihrer Lastprofile entsprechend der ähnlichsten Klasse zuzuordnen (Figueiredo et al., 2005). Diese Herangehensweise entspringt energiewirtschaftlichen Untersuchungen hinsichtlich der Fragestellungen, wann und in welcher Höhe Energieabnehmer als Kunden eines Energieversorgungsunternehmens Energie nutzen (Molina-Solana et al., 2017). In der Literatur werden in diesem Themenzusammenhang Clustering- und Klassifikationsmethoden fälschlicherweise häufig als Synonyme verwendet. Daher sind viele Forschungsarbeiten zu der Klassifikation energetischer Lastprofile zu finden, die jedoch mehrheitlich auf Clustering beruhen, sodass an dieser Stelle auf den vorangestellten Abschnitt verwiesen wird. Klassifikationstechniken folgen auf das Clustering zur Festlegung von Klassencharakteristika. Figueiredo et al. haben dahingehend ein Modell zur Klassifikation von Energieabnehmern basierend auf Entscheidungsbäumen entwickelt, das die kundenindividuelle Vertragsgestaltung in Abhängigkeit der Energiebedarfscharakteristika ermöglicht (Figueiredo et al., 2005).

Weiterhin dienen Klassifikationsmodelle auch der Prognose des gebäudespezifischen Energiebedarfs. Yu et al. haben ein Modell basierend auf einem Entscheidungsbaum vorgeschlagen, mit dem die energetische Performanz von Wohngebäuden beurteilt werden kann (Yu et al., 2010). In der Studie von Li wurde der Einfluss verschiedener Wettervariablen auf den täglichen Spitzenbedarf elektrischer Energie eines städtischen Versorgungssystems untersucht (Li, 2012). Im Sinne der Sicherstellung eines effizienten Gebäudebetriebs dienen

Klassifikationstechniken der Entwicklung von Entscheidungsregeln zur Kontrolle von Betriebsprozessen (Molina-Solana et al., 2017). Auch in diesem Bereich finden vorwiegend Entscheidungsbäume Anwendung (May-Ostendorp et al., 2013; Gao et al., 2010).

Im Zusammenhang mit der Analyse von Energiesystemen ist insgesamt festzuhalten, dass Entscheidungsbäume und SVM häufig gewählte Methoden zur Lösung von Klassifikationsproblemen darstellen (Molina-Solana et al., 2017). Unter Berücksichtigung der Studien zum Clustering energetischer Lastprofile ist zudem festzustellen, dass die Gruppierung von Energieabnehmern mit der Clusteranalyse abgeschlossen ist und eine Kombination mit Klassifikationstechniken ausbleibt.

4.2.3 Implikationen und Anforderungen an die Modellierung

Das Ziel der Integration einer Klassifikationsmethode in das zu entwickelnde Modell dieser Arbeit ist die Identifizierung jener Gebäude und Gebäudegruppen, die kurzfristig und über einen längeren Zeitraum das charakteristische Lastprofil mit der größten Ähnlichkeit zu jedem auf Quartiersebene aufweisen. Auf Grundlage dessen ist im Kontext dieser Arbeit eine Mehrklassen-Klassifikationsmethode zu wählen. LVQ weist dabei zahlreiche Vorteile gegenüber anderen Methoden auf. Weiterhin lässt sich dieser Algorithmus intuitiv mit SOM kombinieren, da die selben Modellprämissen zugrunde liegen. Aus diesen Gründen wird LVQ in das Gesamtmodell integriert.

4.3 Zusammenfassung

Im Rahmen der Kategorisierung von Clustering-Algorithmen lassen sich mit partionierenden, hierarchischen, dichtebasierten, gitterbasierten, topologiebasierten und probabilistisch modellbasierten Methoden sechs Gruppen mit unterschiedlichen Modellcharakteristika identifizieren. Während die verschiedenen Modelle Stärken und Schwächen hinsichtlich des Rechenaufwands und der Ergebnisgüte aufweisen, liegt der Fokus hinsichtlich der Wahl einer Methode vornehmlich auf der Frage, welche Modellannahmen vor dem Hintergrund zur Verfügung stehender Informationen getroffen werden können. Diesbezüglich bestimmt das *a priori* Wissen maßgeblich die Auswahl der Algorithmen.

Tabelle 4.1: Zusammenfassung der Analyseergebnisse zur Eignung der Methodenkategorien im Bereich des Clustering für die Modellbildung in dieser Arbeit

Kategorie	**Algorithmen**	***a priori* Wissen**	**Besonderheiten**	**Performanz**	**Eignung**
Partitionierend	k-Means, k-Medoids, k-Modes,...	Clusteranzahl	kugelförmige Cluster, für kleine und mittlere Datensätze geeignet	o	-
Hierarchisch	BIRCH, Chameleon, CURE, Follow-the-Leader,...	Clusteranzahl oder Distanzmaß	hohe Rechenzeiten, detaillierte Clusterabgrenzungen	o	-
Dichtebasiert	DBSCAN, OPTICS, CLARANS,...	minimale Anzahl an Objekten um Clusterzentrum	Cluster mit beliebigen Formen, Overfitting Effekte	o	-
Gitterbasiert	STING, GDILC, CLIQUE,...	Gitterzellenanzahl	geringe Rechenzeiten	-	o
Topologiebasiert	MANOVA, SOM,...	Netztopologie	für hochdimensionale Datensätze geeignet	+	+
Probabilistisch modellbasiert	Fuzzy-C-Means,...	Clusteranzahl	multiple Clusterzugehörigkeiten	-	-

Tabelle 4.2: Zusammenfassung der Analyseergebnisse zur Eignung der Methodenkategorien im Bereich der Klassifikation für die Modellbildung in dieser Arbeit

Kategorie	**Algorithmen**	**Besonderheiten**	**Performanz**	**Eignung**
Hierarchisch	CART, C4.5, ID3,...	hohe Modelltrainingsdauer, Probleme bei diagonalen Klassengrenzen	+	-
Perzeptron	Winnow, voted-perceptron,...	hohe Rechenzeiten, robust gegenüber Ausreißern	o	o
SVM	DAGSVM,...	hohe Modelltrainingsdauer, robust gegenüber Ausreißern, Klassenrepräsentation mit Extremvektoren	+	-
Statistisch	LDA,...	Wahrscheinlichkeiten für Klassenzugehörigkeiten müssen vorliegen	o	-
Learning Vector Quantization	LVQ1, LVQ2, LVQ3,...	repräsentative Vektoren für Klassenregionen, nicht auf binäre Klassifikation beschränkt, intuitive Kombinierbarkeit mit SOM	o	+
Ensembles	AdaBoost, Random Forests,...	geringer Trainingsaufwand	o	o

Weiterhin wurde herausgestellt, dass sich topologiebasierte Clustering-Methoden zur Gruppierung von thermischen Lastprofilen ohne *a priori* Wissen in Bezug auf die Clusteranzahl eignen, wobei in der Modellierung verschiedene Netzdimensionen berücksichtigt werden, um Unterschiede in der Clustergranularität analysieren zu können. Tabelle 4.1 fasst die analysierten Implikationen für die Modellbildung dieser Arbeit zusammen. Klassifikationsmethoden können in fünf Kategorien eingeteilt werden: hierarchische Methoden, Perzeptron-Methoden, Support Vector Machines, statistische Methoden und Ensemble-Methoden. Im Rahmen der Studien zur Nutzung von Klassifikationsalgorithmen für die Festlegung charakteristischer Gruppen energetischer Lastprofile, zur Prognose gebäudespezifischen Energiebedarfs oder für den effizienten Betrieb von Versorgungseinrichtungen, werden vornehmlich Entscheidungsbäume und SVM eingesetzt. Während erstere eine Modelltrainingsdauer bedingen, basieren letztere auf der Klassifikation anhand von Extremvektoren. Klassifikationen mit LVQ dagegen beruhen auf typischen, repräsentativen Vektoren. Der Rechenaufwand hängt von der vorzugebenden Anzahl letzterer ab und geht mit der gewählten Topologie des Kohonen-Netzes im Kontext der Kombination mit SOM einher. In Tabelle 4.2 sind abschließend die Analyseergebnisse für die Teilmodellentwicklung im Themenbereich der Klassifikation aufgelistet.

KAPITEL 5

Entwicklung eines autoadaptiven prädiktiven Modells

Das vorliegende Kapitel dokumentiert die Entwicklung des autoadaptiven prädiktiven Modells dieser Arbeit zur Quantifizierung von Gleichzeitigkeitseffekten in Lastverteilungen urbaner Energiesysteme. Eingangs werden in Abschnitt 5.1 die Methoden zur Quantifizierung von Gleichzeitigkeitseffekten als Zielvariablen vorgestellt und definiert. Abschnitt 5.2 beinhaltet die mathematischen Grundlagen der Datenvorverarbeitung zur Definition des ersten Teilmodells. Im Anschluss daran wird in Abschnitt 5.3 die modelltheoretische Entwicklung des Prognose-Ensembles als zweites Teilmodell vorgestellt und erläutert. Abschnitt 5.4 rückt die damit verknüpften theoretischen Grundlagen zur Abbildung prädiktiver Lastszenarien in den Fokus. Ferner beinhaltet Abschnitt 5.5 die Methodenkombination aus Clustering und Klassifikation zur automatisierten Analyse und Gruppierung thermischer Lastprofile. Abschließend sind in Abschnitt 5.6 die Kernaspekte des theoretischen Modellaufbaus zusammengefasst.

5.1 Methoden zur Quantifizierung von Gleichzeitigkeitseffekten

Nachfolgend wird zunächst die konventionelle Definition des Gleichzeitigkeitsfaktors erweitert, um diese für prädiktive Analysen zu nutzen. Die Einführung eines Ähnlichkeitsfaktors dient im Folgenden der numerischen Beschreibung einer Ähnlichkeitsstruktur innerhalb thermischer Lastverteilungen und stellt damit einen Zielwert des Clusterings dar. Abschließend dient die Methode zur Identifizierung von gruppenspezifischen Spitzenlastanteilen der Abschätzung kontinuierlichen Lastverschiebungspotentials einzelner Gebäude und Gebäudegruppen, mit dem Ziel Spitzenlasten auf Quartiersebene durch die Nutzung von Gleichzeitigkeitseffekten fortlaufend zu reduzieren.

5.1.1 Erweiterung der Definition des Gleichzeitigkeitsfaktors

Die bisherige empirische Ermittlung des Gleichzeitigkeitsfaktors beruht in der Regel auf Auswertung historischer Messdaten. Dabei wird die Nennleistung der Gebäude auf Grundlage einer linearen Regressionsanalyse bestimmt, indem die thermische Leistung der Außentemperatur gegenüber gestellt wird. Diese Methodik impliziert einen ausschließlichen direkten Zusammenhang zwischen der Außenlufttemperatur und der benötigten Heizleistung des Gebäudes. Diese Kausalität trifft im Wohngebäudebereich tendenziell eher

zu als im Nichtwohngebäudebereich, in dem witterungsunabhängiger, prozessbedingter Wärmebedarf besteht. Auch die Nutzungskomponenten, wie Werktags- oder Ferienzeitcharakteristiken in der Gebäudebelegung, sind somit nicht berücksichtigt. Das im Rahmen dieser Arbeit entwickelte Modell beinhaltet das Prognose-Ensemblemodell, dessen mathematische Grundlagen im weiteren Verlauf dieses Kapitels vorgestellt werden. Auf Basis der trainierten Modelle lassen sich Lastgänge mit verschiedenen Eingangsvariablen, wie Witterungs- oder Nutzungsparametern, simulieren. Für die Simulation und Ermittlung der maximalen Leistungsanforderungen im Kontext des jeweiligen gebäudespezifischen Wärme- und Kühlbedarfs werden Extremwetterdatensätze des Deutschen Wetterdienstes (DWD) verwendet, die in Kapitel 7 eingeführt werden. Für den maximalen Wärmebedarf dient der Extrem-Winterfall aufgrund des damit zu erwartenden hohen Heizbedarfs. In umgekehrter Analogie dazu beruht die Quantifizierung des maximalen Kühlbedarfs auf den Daten des Extrem-Sommerfalls. Die Maxima dieser Zeitreihen werden als Nennleistung angesetzt. Daher wird im Rahmen dieser Arbeit in Anlehnung an Guan et al. (Guan et al., 2016) Gleichung (5.1) verwendet. In diesem Kontext ist der Gleichzeitigkeitsfaktor ϕ als Verhältnis zwischen der maximalen überlagerten Gruppenleistung des Betrachtungsjahres $Q_G\left(\tau_{max}\right)$ und der Summe der maximalen individuellen Leistungsabnahmen $Q_{i,max}$ definiert.

$$\phi\left(\tau\right) = \frac{\dot{Q}_G\left(\tau_{max}\right)}{\sum\limits_{i=1}^{N} \dot{Q}_{i,max}} \tag{5.1}$$

Für jedes Jahr τ des Betrachtungszeitraumes T erfolgt die Berechnung eines Gleichzeitigkeitsfaktors ϕ_τ. Der anzusetzende Gleichzeitigkeitsfaktor ϕ_{max} ergibt sich auf Basis von Gleichung (5.2).

$$\phi_{max} = \max\left(\phi_1, \phi_2, ..., \phi_T\right) \tag{5.2}$$

Diese Definition berücksichtigt implizit den Aspekt der Versorgungssicherheit bei der Auslegung von Erzeugern und Verteilsystemen, indem das Risiko einer Unterdimensionierung infolge zu geringer Gleichzeitigkeitsfaktoren vermindert wird. Darüber hinaus erfolgt die Quantifizierung des Gleichzeitigkeitsfaktors, abgesehen von der simulierten Maximalleistung, anhand weiterer Leistungsspitzen, um Unsicherheiten zu berücksichtigen und einen Entscheidungskorridor für die Wahl eines Gleichzeitigkeitsfaktors anzugeben. Im Sinne des Vergleichs der mit dem Modell dieser Arbeit ermittelten Gleichzeitigkeitsfaktoren und jener auf Grundlage konventioneller statistischer Methoden, werden für die Modelldemonstration in Kapitel 7 auf Basis der gegebenen Datengrundlage Kennlinien für die Wärme- und Kälteversorgung auf Grundlage historischer Messdaten hergeleitet. Die mathematischen Grundlagen und Zusammenhänge dafür sind Abschnitt A.3 in Anhang A zu entnehmen.

5.1.2 Einführung eines Ähnlichkeitsfaktors

Im Rahmen dieser Arbeit wird zusätzlich ein Ähnlichkeitsfaktor ψ eingeführt und definiert. Dieser beschreibt die Ähnlichkeit der charakteristischen thermischen Lastverläufe innerhalb des Quartiers. Die Quantifizierung beruht auf den Ergebnissen des Clusterings der normierten Lastgänge. Dabei werden die folgenden Annahmen getroffen:

I. Können alle Lastprofile einem Cluster zugeordnet werden, so nimmt ψ den Wert 1 an, da alle Gebäude zu ähnlichen Zeitpunkten ihre maximale thermische Leistung abrufen.

II. Ist jedem Cluster jeweils ein Lastprofil zugeordnet, so nimmt ψ den Wert 0 an, da jedes Gebäude zu anderen Zeitpunkten seine maximale thermische Leistung abruft.

III. Für jedes Cluster wird angenommen, dass ψ den Wert 1 annimmt, da die Ähnlichkeit der Lastprofile sehr hoch ist.

Dementsprechend wird ψ nach Gleichung (5.3) definiert, wobei N_c die Clusteranzahl und N_{Geb} die Anzahl der Gebäude repräsentiert.

$$\psi = 1 - \frac{N_c}{N_{Geb}} \tag{5.3}$$

Es ergibt sich beispielsweise für acht Gebäude bei drei identifizierten Clustern ein Ähnlichkeitsfaktor von 0,625. Als Eingangsdaten für das Clustering der Wärmeleistungsprofile dienen Zeitreihen des Winterhalbjahres von Anfang Oktober bis Ende März. Analog dazu finden Kühlleistungsprofile im Sommerhalbjahr von Anfang April bis Ende September Eingang in die Ähnlichkeitsanalyse.

5.1.3 Identifizierung von gruppenspezifischen Spitzenlastanteilen

Um den Einfluss der Ähnlichkeitsanalyse zu demonstrieren, werden gruppenspezifische Spitzenlastanteile ermittelt. Der Anteil κ der thermischen Leistung einer Gebäudegruppe $\dot{Q}_G$ an der thermischen Spitzenleistung des Quartiers zum Zeitpunkt $\tau_{max,Quart}$ lässt sich in Anlehnung an Stern und Guan et al. (Stern, 2013; Guan et al., 2016) mittels Gleichung (5.4) definieren.

$$\kappa = \frac{\dot{Q}_G\left(\tau_{max,Quart}\right)}{\dot{Q}_{Quart}\left(\tau_{max,Quart}\right)} \cdot 100 \tag{5.4}$$

Im Rahmen dieser Arbeit wird κ in Prozent angegeben und für die größten Leistungsspitzen analog zum Gleichzeitigkeitsfaktor gruppenspezifisch ermittelt, sodass ein Wertebereich angegeben werden kann.

5.2 Teilmodell Datenvorverarbeitung

Reale Datensätze stammen häufig aus unterschiedlichen heterogenen Quellen, enthalten Rauschen und beinhalten Fehlstellen (Han et al., 2012). Diese Anomalien beeinflussen die Ergebnisse von Datenanalysen wesentlich. Um valide Zeitreihenanalysen vornehmen zu können, müssen entsprechende fehlerhafte Werte zunächst identifiziert werden. Im Anschluss daran ist festzulegen, wie mit diesen Werten umgegangen werden soll. Um die Datensätze für das Modelltraining vorzubereiten, ist darüber hinaus eine Normalisierung vorzunehmen, die im zweiten Unterabschnitt thematisiert wird.

5.2.1 Identifizierung von Ausreißern

In der Literatur werden globale Ausreißer von kontextuellen und kollektiven Ausreißern unterschieden (Han et al., 2012). Nach Han et al. werden folgende Kategorisierungen vorgenommen: Globale Ausreißer unterscheiden sich stark von den Werten der übrigen Datenpunkte der zugrunde liegenden Datenbasis. Die Klassifizierung als kontextueller Ausreißer wird jenen Datenpunkten zu Teil, welche in einem bestimmten Kontext, wie einer definierten Messperiode, stark von den Werten der übrigen Datenpunkte in der betreffenden Messperiode abweichen. Dementsprechend wird eine Gruppe von Datenpunkten, deren Werte stark von den Werten der verbleibenden Datenpunkte der Datenbasis differieren als kollektive Ausreißer deklariert. Anomalien in Datensätzen können auf zufällige oder systematische Fehler zurück geführt werden. (Han et al., 2012) Zu den zufälligen Fehlern zählen dabei Mess- und Übertragungsfehler. Systematische Anomalien hingegen können ursächlich auf Fehler bei der Berechnung von Merkmalen, auf fehlerhafte Skalierungen oder Drifteffekte zurück geführt werden. Ausreißer können als Fehlerart sowohl zufälliger als auch systematischer Natur sein. (Runkler, 2015) Im Hinblick auf die Identifizierung von Ausreißern in Datensätzen existieren vielfältige Ansätze. In einer detaillierten Studie haben Hodge et al. Methoden zur Identifikation von Ausreißern identifiziert: Statistische Methoden, Techniken des maschinellen Lernens, Künstliche Neuronale Netze und hybride Systeme (Hodge et al., 2004). Gupta et al. haben sich im Speziellen jenen Methoden für temporale Daten gewidmet (Gupta et al., 2014).

Im Rahmen der Modellentwicklung werden Methoden zur Identifizierung von Ausreißern der drei vorgestellten Kategorien von Han et al. entsprechend vorgesehen. Zur Deklaration von globalen Ausreißern wird der Ansatz von John W. Tuckey aus *Exploratory Data Analysis* verwendet, welcher sich an der Charakteristik von Boxplots orientiert. Abbildung 5.1 visualisiert diesen exemplarisch. Hierbei sind Zäune, sogenannte *fences*, für die Identifizierung möglicher Ausreißer definiert. Diese entsprechen den Quartilen des zu analysierenden Datensatzes. Auf Basis des Interquartilsabstands sind anschließend die äußeren Zäune (*outer fences*) festzulegen, welche Grenzen unterhalb des unteren bzw. oberhalb des oberen Quartils darstellen. Datenpunkte, die sich außerhalb der äußeren Zäune befinden, werden als Ausreißer deklariert. Für die Festlegung letzterer ist eine Konstante zu wählen, die, multipliziert mit dem Interquartilsabstand, die äußeren Zäune definiert. (David C. Hoaglin,

2003) Für die Konstante wird entweder der Wert 1,5 oder 3,0 gewählt, um Werte ober- bzw. unterhalb dieser Grenzen als Ausreißer zu klassifizieren. Im Rahmen dieser Arbeit wird die Konstante auf einen Wert von 1,5 festgelegt.

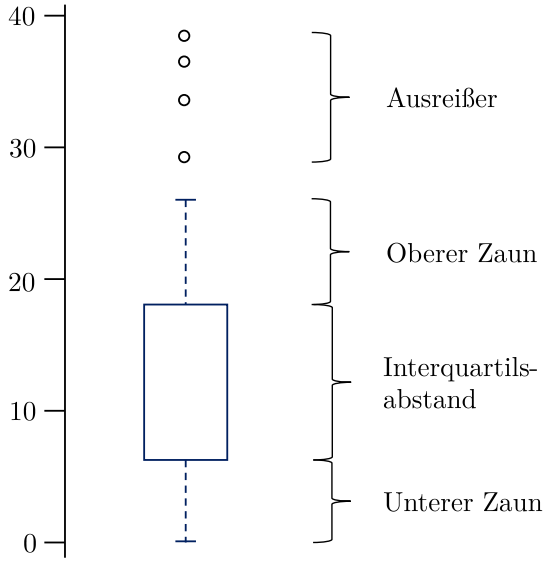


Abbildung 5.1: Datenanalyse mittels Boxplot in Anlehnung an David C. Hoaglin, 2003

Zur Deklaration von kontextuellen Ausreißern erfolgt die Integration der Methode des gleitenden Medians in das Modell. Um den Kontext zu definieren, wird ein Zeitfenster von fünf Zeitschritten festgelegt, in welchem Datenpunkte über oder unter der dreifachen absoluten Abweichung vom lokalen Median, Γ, als Ausreißer gemäß Gleichung (5.5) (in Anlehnung an Cleff, 2015) eingestuft werden.

$$\Gamma = Median\left(|X_i - Median\left(X\right)|\right) \tag{5.5}$$

Kollektive Ausreißer werden mittels Clustering identifiziert. Die Methodik wird im folgenden Abschnitt vorgestellt. Im Rahmen dieser Analyse werden Lastgänge von Gebäuden deklariert, welche ausschließlich Werte von 0 aufweisen.

Im Anschluss an die Identifizierung von Ausreißern muss eine Entscheidung bezüglich des Umgangs mit diesen Werten getroffen werden. Neben der Kennzeichnung entsprechender Daten als Fehlerwerte und dem Entfernen dieser Daten, kann auch eine Korrektur bei fehlerhaften oder Schätzungen fehlender Werte vorgenommen werden. Dazu finden sich verschiedene Ansätze in der Literatur wieder, welche von dem Ersetzen fehlerhafter Werte, beispielsweise durch den Median oder den Mittelwert, über die Interpolation bei Zeitrei-

hen, bis hin zu Filterungsmethoden oder modellbasierten Schätzungen mittels Regression reichen (Runkler, 2015). Um für Regressionsanalysen Zeitreihendatensätze gleicher Länge zu behalten, werden die globalen und kontextuellen Ausreißer ersetzt. Für äquidistante Zeitschritte können Werte mittels linearer Interpolation ermittelt werden (Runkler, 2015).

$$x_k^{(i)} = \frac{x_{k-1}^{(i)} + x_{k+1}^{(i)}}{2} \tag{5.6}$$

Kollektive Ausreißer im Sinne von 0-Lastgängen werden für die weiteren Berechnungsschritte nicht berücksichtigt und dementsprechend aus dem Datensatz entfernt.

5.2.2 Normalisierung der Daten

Die Skalierung von Daten stellt innerhalb des Prozesses der Datenvorverarbeitung einen entscheidenden Schritt dar. Mit der Normalisierung des Datensatzes, welcher als Input für Verfahren des maschinellen Lernens dient, kann die Trainingsgüte erhöht und die Problematik von Ausreißern reduziert werden, da bei nicht-normalisierten Daten jeder Datenpunkt, welcher nicht einen Wert nahe Null aufweist, den Lernprozess verlangsamt (LeCun et al., 1998). Es existieren verschiedene Methoden, wie *Minimum-Maximum Normalisierung* oder *Dezimalskalierungen* (Han et al., 2012). Im Rahmen dieser Arbeit findet die Z-Score-Normalisierung nach Gleichung (5.7) Anwendung, indem von dem Eingangswert x der Mittelwert μ subtrahiert und durch die Standardabweichung σ der Zeitreihe dividiert wird (Han et al., 2012).

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{5.7}$$

Diese Methode ist insbesondere für die Normalisierung von Zeitreihen geeignet, deren Minima und Maxima Ausreißer darstellen könnten (Han et al., 2012).

5.3 Teilmodell Prognose-Ensemble

Nachfolgend werden die Methoden zur Lastprognose im Rahmen der Entwicklung eines Ensemblemodells vorgestellt und die mathematischen Zusammenhänge erläutert. Neben letzteren liegt der Fokus in diesem Abschnitt explizit auch auf individuellen Konfigurationen der Modellparameter und -randbedingungen. Die Inhalte wurden vom Autor mehrheitlich im Rahmen einer Studie von Koschwitz et al., 2018 veröffentlicht. Sowohl NARX RNN als auch ε-SVM-R Modelle werden auf Grundlage von Eingangsdaten und Zielwerten trainiert, validiert, getestet und abschließend evaluiert.

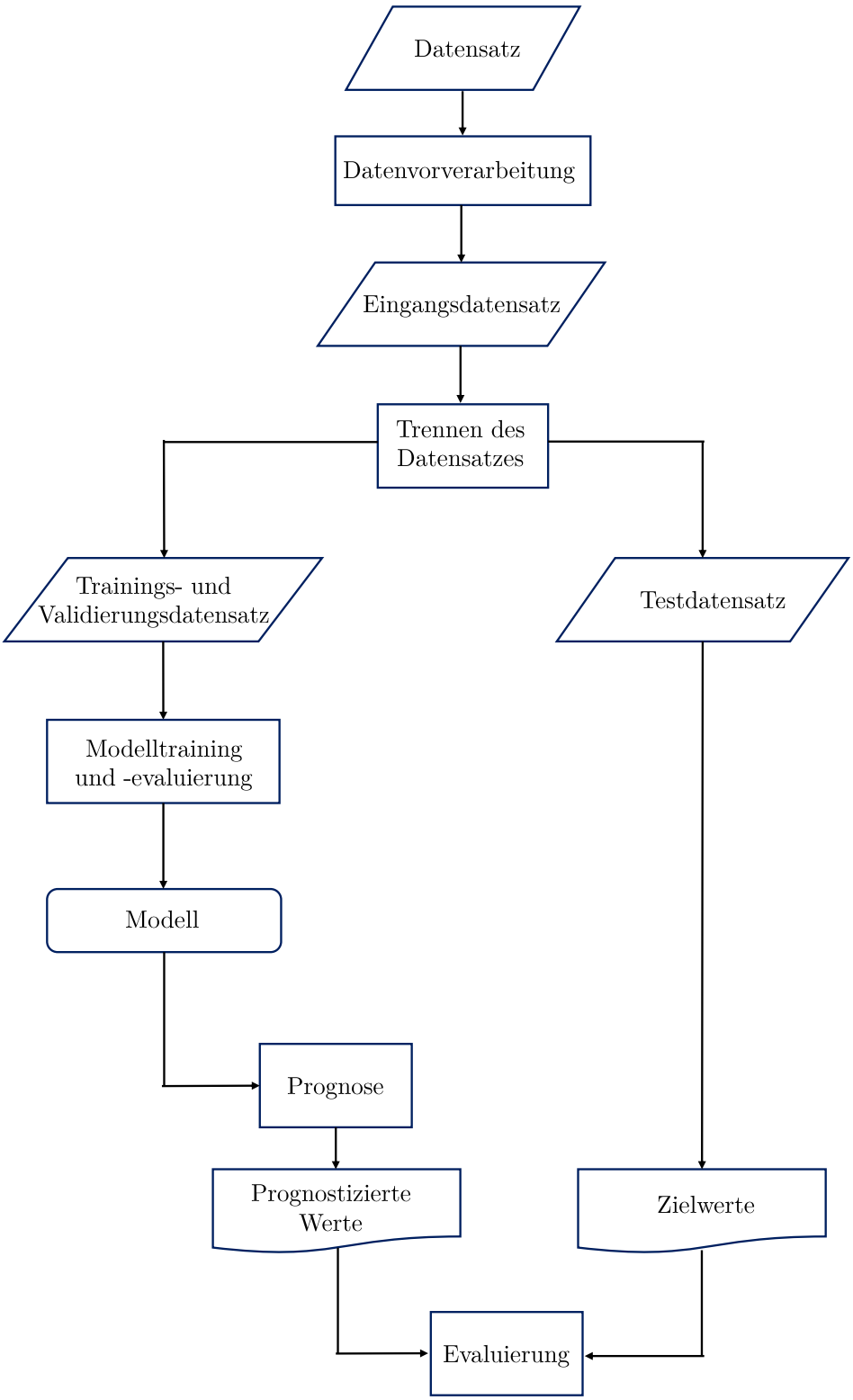


Abbildung 5.2: Flussdiagramm zum Lernprozess der Prognosemodelle in Anlehnung an Magoulès et al., 2016

Abbildung 5.2 zeigt diesen Lernprozess exemplarisch, wobei die einzelnen Teilprozesse des Modelltrainings sowie der Modellvalidierung und Modelltests methodenspezifisch in den folgenden Unterabschnitten detailliert erläutert werden. Der Eingangsdatensatz muss zunächst entsprechend der Ausführungen in Abschnitt 5.2 vorverarbeitet und skaliert werden. Anschließend erfolgt eine Aufteilung in einen Trainings- und Validierungs- sowie in einen Testdatensatz. Hierbei existieren verschiedene Ansätze in der Literatur, wie beispielsweise *Simple Random Sampling*, *Systematic Sampling* oder *Convenience Sampling*, wobei sich die genannten Methoden insbesondere für Zeitreihen eignen (Reitermanova, 2010). *Simple Random Sampling* beinhaltet die zufällige Auswahl einer Teilmenge eines Datensatzes, wobei jede mögliche Teilmenge die gleiche Wahrscheinlichkeit besitzt ausgewählt zu werden (Lohr, 1999). Im Sinne des *Systematic Sampling* erfolgt eine Ordnung des zugrunde liegenden Datensatzes nach einer vorzugebenden Regel. Anschließend wird eine zu definierende Anzahl an Teilmengen gebildet, welche sukzessive als Trainingsgrundlage Anwendung findet (Reitermanova, 2010). *Convenience Sampling* bedarf einer Unterteilung des Datensatzes in indiskrete Abschnitte. Die Teilmengen müssen zudem eine gleiche Verteilung ihrer Komponenten aufweisen. (Reitermanova, 2010) Die Vorgabe einer datensatzspezifischen Ordnungsregel im Kontext des *Systematic Sampling* widerspricht der Automatisierungsprämisse des Modells dieser Arbeit. Ferner kann keine Periodizität in den Zeitreihen der Eingangsdatensätze, insbesondere bei den thermischen Lastgängen und Wettervariablen, eingehalten werden, sodass sich auch *Convenience Sampling* nicht für die Modellentwicklung eignet. Aus diesen Gründen findet der häufig verwendete Ansatz auf Grundlage des *Simple Random Sampling* Eingang. Es ist an dieser Stelle darauf zu achten, dass der Anteil des Trainingsdatensatzes am Gesamtdatensatz nicht zu gering ist. (Reitermanova, 2010) Für die Modellentwicklung in dieser Arbeit werden randomisiert 70% der Daten für das Modelltraining und jeweils 15% für die Validierung und Modelltests verwendet.

5.3.1 Rekurrente Neuronale Netzwerke

Die Ausführungen dieses Unterabschnitts stellen die Aufbereitung von mathematischen Grundlagen sowie Konzepten für den Aufbau und die gewählten Parametrierungen der NARX RNN in den Fokus.

Netzwerkarchitekturen

Als Grundlage für die schematische Darstellung von Informationsvernetzungen dienen Netzwerkarchitekturen, deren Konfigurationen für die Modellentwicklung dieser Arbeit nachfolgend vorgestellt werden. Abbildung 5.3 veranschaulicht dazu die Topologie eines NARX RNN mit einer verdeckten Schicht. Die integrierten Parameter werden im weiteren Verlauf vorstellt und in einen formalen Kontext gesetzt. Ein RNN bietet die Möglichkeit, Rückkopplungen zu integrieren, welche dazu dienen, zeitlich vorgelagerte Zustände in die Berechnung aktueller Zustände mit einzubeziehen. Zur Modellierung dieser Eigenschaft werden häufig Zeitverzögerungen in Neuronen oder in Verbindungen zwischen Neuronen

integriert (Diaconescu, 2008).

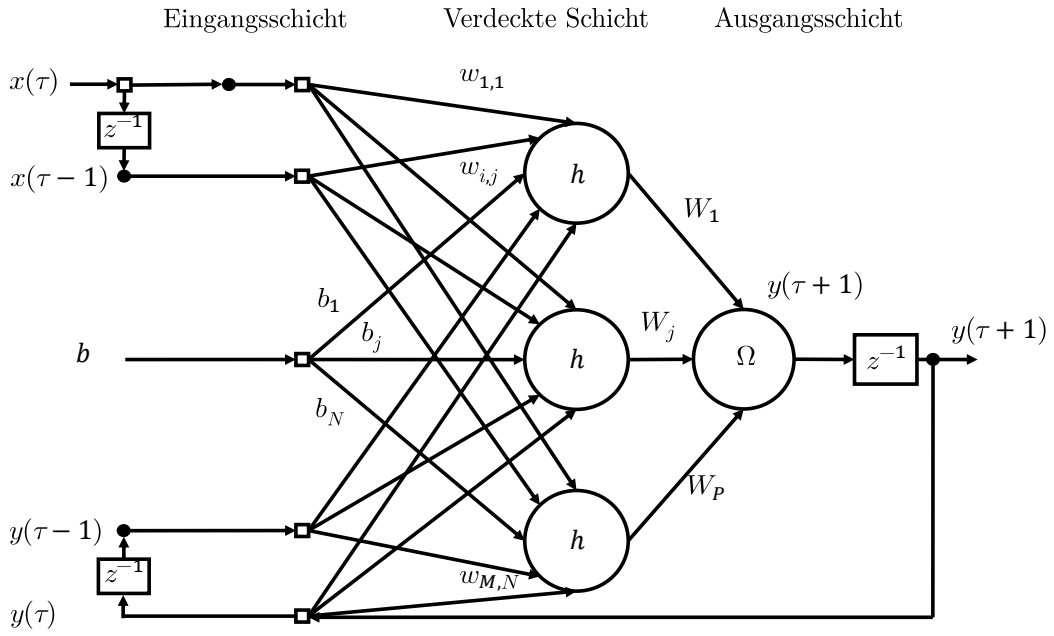


Abbildung 5.3: Architektur des NARX RNN mit einer verdeckten Schicht in Anlehnung an Haykin, op. 1999 und Koschwitz et al., 2018

Dazu ist zunächst die Dimension n zu definieren, welche die Anzahl zu berücksichtigender historischer Werte der Zeitreihe $y(\tau)$ festlegt, um anschließend den Rekonstruktionsvektor $\Lambda(\tau)$ mit dem Zeitverzögerungsparameter λ zu definieren (Bradley et al., 2015). $\Lambda(\tau)$ dient zusammen mit den weiteren exogenen Eingangsvektoren $x(\tau)$ als Trainingsparameter für das Netzwerk zur Prognose von $y(\tau+1)$. Der formale Zusammenhang ist Gleichung (5.8) zu entnehmen.

$$\Lambda(\tau) = [y(\tau), y(\tau-\lambda), ..., y(\tau-(n-1)\lambda)] \tag{5.8}$$

Die Dimensionierung des Rekonstruktionsvektors muss iterativ bestimmt werden und folgt keinen allgemein gültigen Vorgaben. Garland und Bradley haben gezeigt, dass hohe Rekonstruktionsdimensionen nicht zwangsläufig zu verbesserten Prognoseergebnissen führen, gerade wenn die Datengrundlage Rauschen aufweist. Dies liegt darin begründet, dass jeder verrauschte historische Wert, welcher mit in die Prognose eines zukünftigen Wertes einbezogen wird, das Rauschen überträgt (Garland et al., 2015). Für beide Netzwerke werden logistische Sigmoidfunktionen als Aktivierungsfunktionen verwendet, welche sich insbesondere aufgrund ihrer kontinuierlichen funktionalen Charakteristik für die Prognose von Zeitreihen eignen. Dabei ermöglichen die Sattelpunkte der Funktion die Modellierung

von minimalen und maximalen thermischen Leistungen (Powell et al., 2014). Abbildung 5.4 zeigt den funktionalen Verlauf der Aktivierungsfunktion.

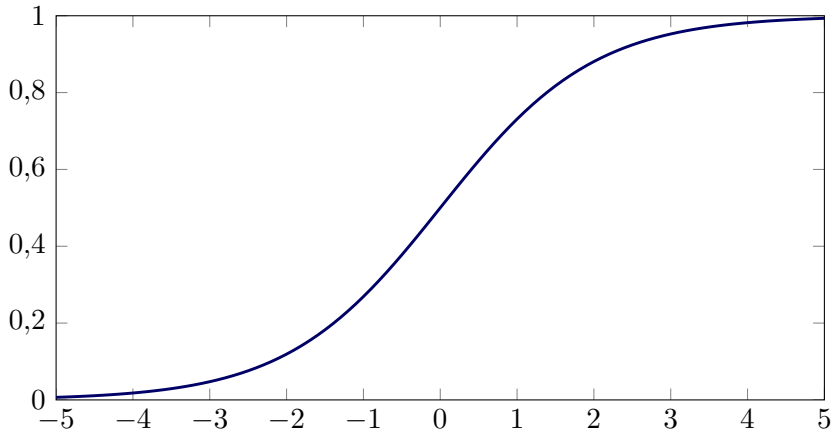


Abbildung 5.4: Logistische Sigmoidfunktion

Gleichung (5.9) definiert die Aktivierungsfunktion h für jedes Neuron j in der verdeckten Schicht, in Abhängigkeit von Eingangsgröße x des Netzwerkes (Bishop, 2006). Jede Verbindung zwischen Eingangsgröße und einem Neuron der verdeckten Schicht wird mit einem Gewichtungsfaktor w multipliziert und Bias b addiert.

$$h_j\left(x_i\right) = \tanh\left(\sum_{i=1}^{M} w_{i,j} x_i + b_i\right) \tag{5.9}$$

Gleichung 5.10 definiert die Summe der Ausgangsgrößen Ω der verdeckten Schicht, in Abhängigkeit von den dazugehörigen Gewichtungsfaktoren W und Bias B (Bishop, 2006).

$$\Omega\left(h_j\right) = \sum_{j=1}^{N} W_j h_j + B \tag{5.10}$$

Der resultierende Prognosewert $y\left(\tau\right)$ wird anschließend wieder zurück geführt und fungiert als Eingangswert für die Berechnung des nachfolgenden Prognosewertes. Die vorangestellten grundlegenden Gleichungen zur Parameterquantifizierung sind prinzipiell auf eine Netzwerkarchitektur mit zwei oder multiplen verdeckten Schichten übertragbar. Es ergeben sich jedoch komplexere Abhängigkeiten und Strukturen im Netzwerk. Abbildung 5.5 veranschaulicht die Architektur des NARX Rekurrentes Neuronales Netz (RNN) mit zwei verdeckten Schichten. Während des Trainingsprozesses erfolgt eine Modifikation der Netzwerkarchitekturen, dem Ansatz aus der Automatentheorie von Narendra and Parthasarathy

folgend.

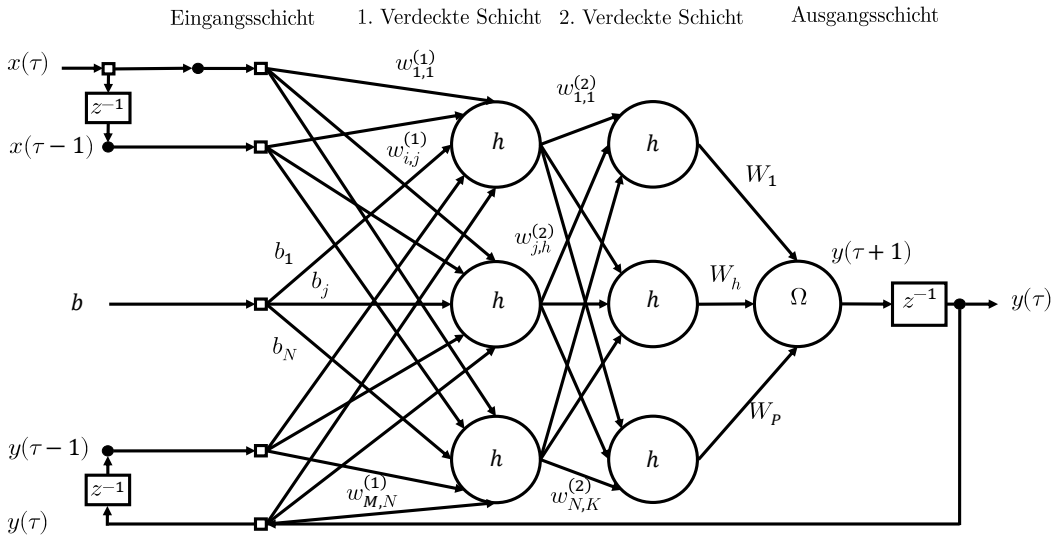


Abbildung 5.5: Architektur des NARX RNN mit zwei verdeckten Schichten in Anlehnung an Haykin, op. 1999

Die Autoren haben gezeigt, dass es für das Training des Netzwerkes sinnvoll ist, den rekursiven Charakter des Netzwerkes nicht mit einzubeziehen und die Ausgangsgrößen nicht zurückzuführen, sofern reale Zielwerte zur Verfügung stehen (Narendra et al., 1991). Dementsprechend wird die Architektur erst in der Testphase des Netzwerkes geschlossen, indem Λ Berücksichtigung findet.

Informationsfluss und Fehlerrückführung

Als Grundlage für die weiteren Ausführungen dieses Unterkapitels dient Abbildung 5.6, welche den Zusammenhang zwischen dem vorwärtsgerichteten Informationsfluss (Feedforward) und der Fehlerrückführung (Backpropagation) durch das Netzwerk exemplarisch visualisiert. Die folgenden mathematischen Grundlagen von Feedforward und Backpropagation sowie den Prinzipien des Netzwerktrainings im Folgeabschnitt können Bishop, 2006 entnommen werden. Für jedes Neuron j innerhalb der Netzwerkarchitektur wird jeder Eingangswert z_i mit einem Gewichtungsfaktor $w_{j,i}$ innerhalb des vorwärtsgerichteten Informationsflusses (*Feedforward*) nach Gleichung (5.11) multipliziert (Bishop, 2006). Die Summe der gewichteten Eingangswerte ν für ein beliebiges Neuron j ergibt sich somit aus Gleichung (5.11).

$$\nu_j = \sum_i w_{j,i} \cdot z_i \tag{5.11}$$

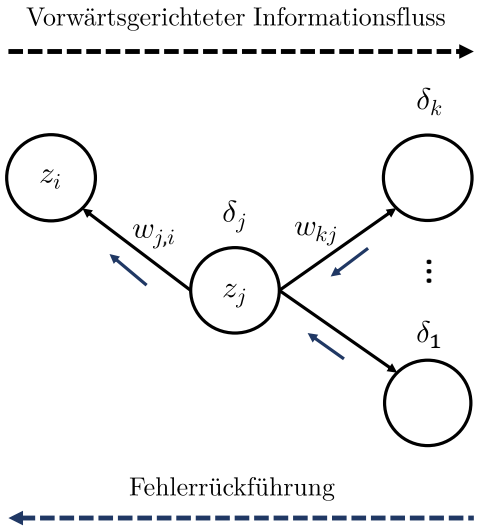


Abbildung 5.6: Prinzip der Fehlerrückführung (Backpropagation) in einem Feedforward-Netzwerk in Anlehnung an Bishop, 2006

Die Integration der bereits eingeführten nichtliniearen Aktivierungsfunktion h ergibt Gleichung (5.12) (Bishop, 2006).

$$z_j = h\left(\nu_j\right) \tag{5.12}$$

Infolgedessen lässt sich die Aktivierung des Neurons j festgelegen. Auf Grundlage der Gleichungen (5.11) und (5.12) erfolgt die vorwärtsgerichtete Informationsweitergabe durch die Netzwerkstruktur. Im Anschluss daran werden die Differenzen δ zwischen den Ausgangswerten y und den dazugehörigen Zielwerten t ermittelt. Für Neuron k ist somit die Differenz δ_k nach Gleichung (5.13) zu bestimmen (Bishop, 2006).

$$\delta_k = y_k - t_k \tag{5.13}$$

Auf Basis von δ_k lässt sich δ_j rekursiv mittels der ersten Ableitung der Aktivierungsfunktion von Neuron j, $h'\left(\nu_j\right)$ gemäß Gleichung (5.14) bestimmen.

$$\delta_j = h'\left(\nu_j\right) \sum_k w_{k,j} \cdot \delta_k \tag{5.14}$$

Mittels dieses formalen Zusammenhangs ist jedes δ einer verdeckten Schicht des Netzwerkes zu berechnen (Bishop, 2006) und die partielle Ableitung des Fehlers E_n nach $w_{i,j}$ entspricht

dem Produkt aus δ_j und z_j, entsprechend Gleichung (5.15).

$$\frac{\partial E_n}{\partial w_{i,j}} = \delta_j \cdot z_i \tag{5.15}$$

Netzwerktraining

Das Netzwerktraining geht mit der Minimierung der quadratischen Fehlerfunktion $E(w)$ nach Gleichung (5.16) einher, bei gegebenem Eingangswert x_i und Zielwert t_i (Bishop, 2006).

$$E(w) = \frac{1}{2}\sum_{i=1}^{N}\{y(x_i,w) - t_i\}^2 \tag{5.16}$$

Obige Gleichung quantifiziert somit die quadratische Abweichung zwischen Eingangs- und Zielwert. Eine Anpassung von w zu $w + \delta w$ bedingt ferner eine Adaption der Fehlerfunktion gemäß Gleichung (5.17) (Bishop, 2006).

$$\delta E \backsimeq \delta w^T \nabla E(w) \tag{5.17}$$

Das Ziel der sukzessiven Anpassung der Fehlerfunktion liegt darin, letztere zu minimieren, sodass Gleichung (5.18) gilt (Bishop, 2006).

$$\nabla E(w) = 0 \tag{5.18}$$

Die iterative Modifikation von w zum Zeitpunkt $\tau + 1$ nach Gleichung (5.19) basiert auf dem Wert von w zum vorherigen Zeitschritt τ und dem mit einer zu definierenden Lernrate η zu multiplizierenden dazugehörigen Gradienten der Fehlerfunktion $\nabla E\left(w^{(\tau)}\right)$.

$$w^{(\tau+1)} = w^{(\tau)} - \eta \nabla E\left(w^{(\tau)}\right) \tag{5.19}$$

Somit entscheidet η über die Geschwindigkeit, in der das Minimum von $E(w)$ approximiert wird. Dieses Prinzip wird in der Literatur unter *Gradient Descent* oder *Steepest Descent* zusammengefasst, wobei $\eta > 0$ gilt (Bishop, 2006). Die Untersuchung von Lernraten zur Optimierung des Trainingsprozesses stellt ein breites Forschungsfeld dar. In diesem Zusammenhang haben Wilson & Martinez (Wilson et al., 2001) eine Reihe verschiedener

konstanter Lernraten hinsichtlich ihres jeweiligen Zeitpunkts der besten Trainingsepoche, Genauigkeit und getesteter Trainingsepochen untersucht. Eine Lernrate von 0,1 bedingte die schnellste Konvergenz bei 1000 Trainingsepochen. Größere Werte von η führten zu geringerer Genauigkeit und wesentlich längeren Trainingszeiten zur Erreichung der maximalen Genauigkeit. Mit $\eta > 0{,}1$ ließ sich die Genauigkeit noch um ungefähr vier Prozent erhöhen, wodurch die Anzahl der Trainingsepochen zum Erreichen der höchsten Genauigkeit um da circa 34-fache erhöhte. (Wilson et al., 2001) Vor dem Hintergrund der Prognose thermischer Lastgänge von urbanen Regionen wird daher zur Begrenzung von Rechenzeiten im Rahmen Modellentwicklung dieser Arbeit eine Lernrate von 0,1 als Kompromiss verwendet. Zur Initialisierung von w werden Gewichtungsfaktoren von 1 gewählt, um allen exogenen Einflussparametern zu Beginn des Lernprozesses die gleiche Bedeutung zu verleihen.

Zur Lösung des numerischen Optimierungsproblems und der Aktualisierung der Parameter w und b im Adaptionsprozess findet der Levenberg-Marquardt (LM) Algorithmus Eingang in das Modell (Marquardt, 1963; Hagan et al., 1994). Die folgenden Ausführungen dazu basieren auf Lourakis, 2005. Mit LM lässt sich das Minimum multivariater Funktionen bestimmen (Levenberg, 1944; Marquardt, 1963). Der Algorithmus findet eine breite Anwendung im Bereich der Lösung nicht-linearer Probleme mittels der Methode der kleinsten Quadrate und vereint die Eigenschaften der Gauß-Newton-Methode und *Steepest Descent* (Lourakis, 2005). Die folgenden mathematischen Zusammenhänge basieren auf des Ausführungen von Lourakis (Lourakis, 2005).

Zunächst ist ein Parametervektor $p \in \mathbb{R}^m$ festzulegen, wobei $y = f(p)$ gilt. Wenn weiterhin f eine Funktion ist, die den Parametervektor auf den Zielvektor y projiziert, gilt der Zusammenhang gemäß Gleichung (5.20) (Lourakis, 2005).

$$y = f(p), y \in \mathbb{R}^n \tag{5.20}$$

Im Weiteren ist ein Parametervektor p^+ zu finden, welcher diesen funktionalen Zusammenhang am besten beschreibt und damit die quadrierte Distanz bzw. den quadrierten Fehler $\varepsilon^T \varepsilon$ zwischen x und y mit

$$\varepsilon = x - y \tag{5.21}$$

minimiert (Lourakis, 2005). Die Funktion f wird in der Umgebung von p beim LM Algorithmus linear approximiert. In diesem Zusammenhang gilt für ein kleines $\|\delta_x\|$ und auf Grundlage der Taylorreihenentwicklung Gleichung (5.22)

$$f(p + \delta_p) \approx f(p) + J\delta_p \tag{5.22}$$

mit der Jakobi-Matrix J (Lourakis, 2005):

$$J = \frac{\partial f\,(p)}{\partial p} \tag{5.23}$$

Basierend auf einer anfänglichen Parameterschätzung p_0, produziert die Methode eine Reihe von Parametervektoren, welche den lokalen Minimierer der Funktion f, repräsentiert durch p^+, approximieren. Folglich ist in jedem Iterationsschritt jenes δ_p zu finden, das

$$\|x - f\,(p + \delta_p)\,\| \approx \|x - f(p) - J\delta_p\| = \|\varepsilon - J\delta_p\| \tag{5.24}$$

minimiert. Damit stellt δ_p die Lösung eines linearen Optimierungsproblems kleinster Quadrate dar, wobei das Minimum gefunden ist, wenn $J\delta_p - \varepsilon$ orthogonal zu dem Spaltenraum von J ist (Lourakis, 2005). Dieser Zusammenhang resultiert in Gleichung (5.25).

$$J^T\,(J\delta_p - \varepsilon) = 0 \tag{5.25}$$

δ_p stellt folglich die Lösung der *Normalgleichungen* gemäß (5.26) dar, wobei J^TJ die Approximation der Hesse-Matrix definiert.

$$J^TJ\delta_x = J^T\varepsilon \tag{5.26}$$

Im Rahmen des LM Algorithmus wird die augmentierte Normalgleichung nach (5.27) gelöst, wobei die Elemente der Matrix M, welche nicht auf der Diagonalen liegen, identisch mit J^TJ sind.

$$M\delta_p = J^T\varepsilon \tag{5.27}$$

Demgegenüber sind die Elemente auf der Diagonalen durch

$$N_{ii} = \iota + [J^TJ]_{ii} \tag{5.28}$$

für $\iota < 0$ gegeben. Der Parameter ι wird auch als *Dämpfungsterm* bezeichnet, welcher in jedem Iterationsschritt eine Anpassung erfährt, um die Reduzierung des Fehlers ε zu gewährleisten. Der Dämpfungsterm wird reduziert, wenn der aktualisierte Parametervektor

aus Gleichung (5.27) zu einer Reduzierung von ε führt und somit die Aktualisierung akzeptiert ist. Andernfalls erfolgt die Erhöhung des Dämpfungsterms und die Gleichungen nach (5.27) werden daraufhin erneut gelöst. Dieser Iterationsprozess wird so lange wiederholt, bis ein δ_p gefunden wurde, welches ε reduziert. (Lourakis, 2005)

Alternativ zu LM-Algorithmus kann beispielsweise auch Bayesian Regularization verwendet werden, welche im Vergleich zum erstgenannten die Generalisierungseigenschaften des KNN erhöht (Dan Foresee et al., 1997). Mittels Bayesian Regularization wird die Hesse-Matrix berechnet. Demgegenüber erfolgt die Berechnung der Hesse-Matrix mittels des LM-Algorithmus nicht explizit, sondern approximiert. Daraus ergibt sich der Vorteil einer geringeren Rechenzeit, jedoch auch der Nachteil eines trainierten Netzwerkes mit geringeren Generalisierungseigenschaften. Im Rahmen der Evaluierung der Performanz beider Optimierungsalgorithmen konnte im Rahmen der Studie von Koschwitz et al. mit der Bayesian Regularization Optimierung eine Verringerung des Prognosefehlers um durchschnittlich ungefähr vier Prozent erreicht werden. Gleichzeitig verlängerte sich die Rechenzeit auf das 23-fache gegenüber der Anwendung des LM Algorithmus (Koschwitz et al., 2018). Hinsichtlich der Entwicklung eines performanten Modells für die Analyse urbaner Regionen ist es das Ziel, die Rechenzeit zu begrenzen bei geringen Einbußen im Performanzbereich. Nicht zuletzt wird Rechenzeit für wie die anderen Ensemble-Komponenten und weitere Teilmodelle benötigt.

Regularisierung

Für die Regularisierung des Trainingsprozesses gibt es verschiedene Ansätze, wie *Soft weight sharing* oder *Tangent propagation* (Bishop, 2006), wobei im Rahmen dieser Arbeit das Konzept *Early stopping* verfolgt wird. Letzteres ist vorteilhaft, da es das *Saturation* Phänomen umgeht. Dieses tritt auf, wenn keine Verbesserung im Trainingsprozess mehr erreicht werden kann, infolge des Erreichens eines bestimmten Regulierungsniveaus (Yao et al., 2007). Weiterhin lassen sich mit *Early stopping* in Kombination mit *Backpropagation* *Overfitting*-Effekte auch bei größeren Netzwerken eingrenzen (Caruana et al., 2001). Im Kontext von *Early stopping* erfolgt die Beendigung des Trainingsprozesses, sofern der geringste Fehler unter Verwendung der Validierungsdaten erreicht ist. Abbildung 5.7 zeigt dahingehend exemplarisch einen typischen Verlauf der Fehlerfunktion und den Punkt des minimalen Fehlers. Die Regularisierung im Rahmen der Modellentwicklung in dieser Arbeit beruht auf den folgenden Kriterien:

I. Die maximale Anzahl der vorgegebenen Trainingsepochen ist erreicht

II. Der Prognosewert entspricht dem Zielwert

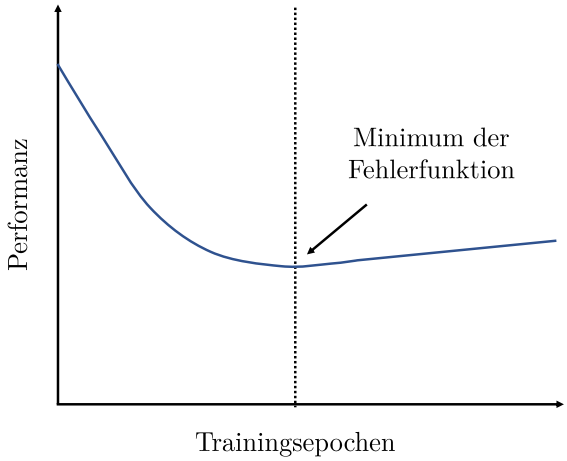


Abbildung 5.7: Darstellung relevanter Zusammenhänge zwischen Parametern zur Regularisierung

Initialisierung für Netzwerksimulationen

Für die Initialisierung der Netzwerksimulation auf Grundlage neuer Datensätze ergibt sich eine Besonderheit aufgrund der implizierten rekursiven Eigenschaften des verwendeten Netzwerktyps. Dies liegt darin begründet, dass historische Zielwerte zur Prognose aktueller Zielwerte Anwendung finden. Diese liegen jedoch außerhalb des Trainings-, Validierungs- und Testprozesses für Prognosen nicht vor, da beispielsweise der Kühl- oder Heizleistungsbedarf für den morgigen Tag unbekannt ist und erst mittels der Prognosemethoden abgeschätzt werden soll. Daher ist es notwendig, für den ersten Prognosezeitschritt einen Zielwert vorzugeben. Im Rahmen dieser Arbeit wird der arithmetische Mittelwert der historischen Zeitreihe zum gemessenen Wärme- bzw. Kühlleistungsbedarf als Zielwert angesetzt.

5.3.2 Support Vector Machine Regression

Im Folgenden werden die mathematischen Zusammenhänge zur ε-Support Vector Regression nach Vapnik vorgestellt, welche, auf Grundlage zweier im weiteren Verlauf dieses Abschnitts näher erläuterter Kernel-Funktionen, Eingang in das Prognose-Ensemble finden. Das Ziel hierbei ist wiederum die Minimierung der Abweichung zwischen Prognose- und Zielwerten. Die folgenden mathematischen Zusammenhänge basieren auf Drucker et al., 1997 und Bishop, 2006. Im Sinne der Modellierung des zu lösenden Optimierungsproblems erfolgt zunächst die Formulierung einer ε-insensitiven Verlustfunktion (*Loss function*) L (Vapnik, 1995), welche in Abbildung 5.8 dargestellt ist. Im Speziellen wird bei der ε-Support Vector Machine Regression ein Parameter ε definiert, welcher einen Abstand zu der Regressionsfunktion $y(x)$ festlegt, innerhalb dessen prognostizierte Werte keinen Verlust bedingen.

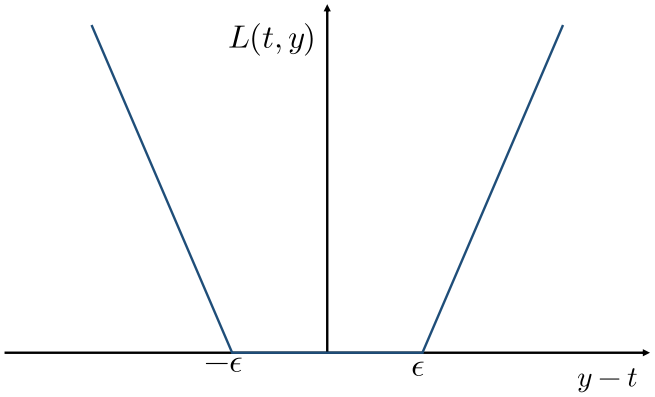


Abbildung 5.8: Darstellung der Fehlerfunktion für die ε-Support Vector Machine Regression in Anlehnung an Drucker et al., 1997 und Koschwitz et al., 2018

Befindet sich ein prognostizierter Wert $y(x_i)$ außerhalb dieses Bereiches, berechnet sich der Verlust aus der Differenz zwischen dem prognostizierten Wert und dem Radius von ε (Drucker et al., 1997). Dieser Zusammenhang ist in Gleichung (5.29) formuliert.

$$L\left(y\left(x\right)-t\right)=\begin{cases}0 & \text{if } \left|y\left(x\right)-t\right|<\varepsilon\\ \left|y\left(x\right)-t\right|-\varepsilon & \text{otherwise}\end{cases} \tag{5.29}$$

Eingangs erfolgt auf Grundlage dessen die Formulierung eines linearen Regressionsproblems, das im weiteren Verlauf für nicht-lineare Lösungen erweitert wird. Daran anknüpfend ist zu Beginn ein lineares Modell gemäß Gleichung (5.30) aufzustellen, mit Regressor x und einem dazugehörigen Funktionswert $y\left(x\right)$ (Haykin, 2009).

$$y(x)=w^{T}x+b \tag{5.30}$$

Die Parameter w und b sind zunächst unbekannt und werden auf Grundlage der weiteren mathematischen Zusammenhänge hergeleitet. Im Sinne der Lösung des Optimierungsproblems ist in der Folge die Fehlerfunktion nach Gleichung 5.31 zu minimieren, wobei w den Gewichtungsvektor für die Parametrierung der Eingangsvariablen repräsentiert (Bishop, 2006; Drucker et al., 1997).

$$\frac{1}{2}\|w\|^{2}+C\left(\sum_{i=1}^{l}\xi_{i}^{*}+\sum_{i=1}^{l}\xi_{i}\right) \tag{5.31}$$

Hierbei sind die nachfolgenden Nebenbedingungen zu berücksichtigen.

$$
\begin{aligned}
&t_i - y\left(x_i\right) \leq \varepsilon + \xi_i \\
&y\left(x_i\right) - t_i \leq \varepsilon + \xi_i^* \\
&\xi_i^* \geq 0 \\
&\xi_i \geq 0
\end{aligned}
$$

Für die Formulierung der Zielfunktion werden zwei Schlupfvariablen ξ_i und ξ_i^* eingeführt. Sofern der beobachtete Punkt über dem mittels ε definierten Bereich liegt, weist ξ_i die positive Differenz zwischen dem beobachteten Punkt und ε aus. Analog dazu ist ξ_i^* Null, wenn der beobachtete Punkt unterhalb des ε-Bereiches liegt. Mit dem konstanten Regulierungsparameter C kann die Gewichtung des aufsummierten Fehlers gewichtet werden (Drucker et al., 1997). Die Festlegung der Parameter ε und C ist als sensitiv zu betrachten, wobei insbesondere erstgenannter einen wesentlichen Einfluss auf die Regressionsgüte hat (Cherkassky et al., 2004). In der Literatur sind verschiedene Ansätze zur Abschätzung von ε zu finden, welche jedoch in den meisten Fällen auf Erfahrungswerten beruhen. Analytische Ansätze bedingen jedoch Informationen über die Datengrundlage und basieren daran anknüpfend auf der Standardabweichung der Rauschamplitude in t. Letztere ist auf Basis der zugrunde liegenden Modellart des Rauschens in t zu berechnen (Kwok, 2001). Vor dem Hintergrund des autoadaptiven Charakters des Modells dieser Arbeit lässt sich letztgenannter Ansatz jedoch nicht weiter verfolgen, weil die Datenbasis aufgrund der Automatisierungsprämisse als unbekannt vorauszusetzen ist. Daher wird ein alternativer Ansatz auf Grundlage der Standardabweichung von t, σ_t, verwendet. Die Standardabweichung ist das Maß für die Differenz einzelner Datenpunkte zu ihrem Mittelwert (Han et al., 2012). Im Rahmen der Modellentwicklung in dieser Arbeit wird ε zu einem Zehntel der Standardabweichung von t festgelegt. Für die Quantifizierung von C wird Gleichung (5.32) nach Cherkassky & Ma verwendet, welche sich in der praktischen Anwendung auf Grund ihrer Robustheit gegenüber Ausreißern in Datensätzen bewährt hat (Cherkassky et al., 2004).

$$
C = \max\left(|\mu_t + 3\sigma_t|, |\mu_t - 3\sigma_t|\right) \tag{5.32}
$$

Abbildung 5.9 visualisiert den Kontext zwischen den bisher eingeführten Parametern der SVM-R.

Mit der Einführung einer Lagrange-Funktion kann der Sattelpunkt von L bestimmt werden. Dazu sind die Lagrange Multiplikatoren α, α^*, γ und γ^* zu definieren (Drucker et al., 1997). Weiterhin stellt der integrierte Bias b der Schätzfunktion einen im Folgenden zu

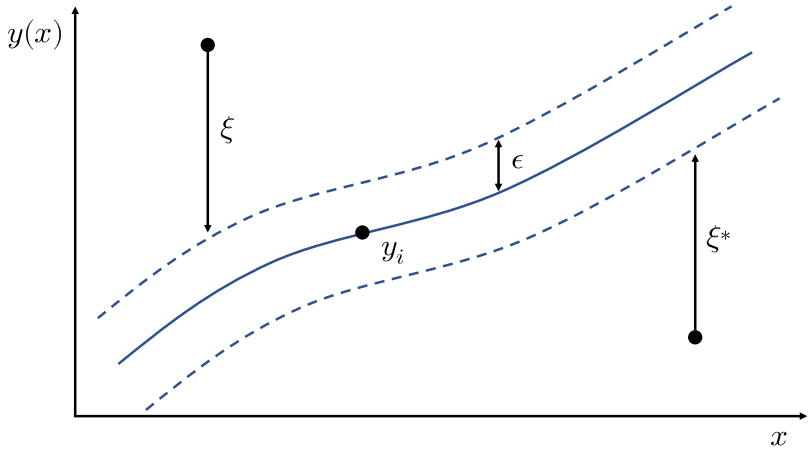


Abbildung 5.9: Darstellung der Kernparameter für die ε-Support Vector Machine Regression in Anlehnung an Bishop, 2006 und Koschwitz et al., 2018

definierenden Parameter dar.

$$
\begin{aligned}
L\left(\omega,b,\xi^*,\xi,\alpha^*,\alpha,\gamma^*,\gamma\right) = \\
&\frac{1}{2}\|\omega\|^2 + C\left(\sum_{i=1}^{l}\xi_i^* + \sum_{i=1}^{l}\xi_i\right) \\
&- \sum_{i=1}^{l}\alpha_i\left[y_i - (\omega x_i) - b + \varepsilon + \xi_i\right] \\
&- \alpha_i^*\left[y_i - (\omega x_i) - b + \varepsilon + \xi_i^*\right] \\
&- \sum_{i=1}^{l} l\left(\gamma_i^*\xi_i^* + \gamma\xi\right)
\end{aligned}
\tag{5.33}
$$

Die partielle Ableitung von L nach ω_i, b und ξ wird anschließend entsprechend Gleichung 5.34 zu Null gesetzt (Drucker et al., 1997).

$$
\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \xi^*} = \frac{\partial L}{\partial \xi} = 0 \tag{5.34}
$$

Es ergeben sich dadurch die Gleichungen 5.35 bis 5.38 (Clarke et al., 2005).

$$
\frac{\partial L}{\partial b} = \sum_{i=1}^{l}\left(\alpha_i^* - \alpha_i\right) = 0 \tag{5.35}
$$

$$\frac{\partial L}{\partial \omega} = w - \sum_{i=1}^{l} \left(\alpha_i^* - \alpha_i\right) x_i = 0 \tag{5.36}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \gamma_i = 0 \tag{5.37}$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \gamma_i^* = 0 \tag{5.38}$$

Aus Gleichung (5.36) ergibt sich unmittelbar der Schätzer für die bislang unbekannte Variable w. Die Integration obiger Gleichungen in Gleichung 5.34 resultiert in dem zu lösenden Maximierungsproblem gemäß Gleichung 5.39.

$$\begin{aligned} W\left(\alpha_i^*,\alpha_i\right) = \\ & \sum_{i=1}^{l} t_i \left(\alpha_i^* - \alpha_i\right) - \varepsilon \sum_{i=1}^{l} \left(\alpha_i^* + \alpha_i\right) \\ & - \frac{1}{2} \sum_{i,j=1}^{l} \left(\alpha_i^* - \alpha_i\right) \left(\alpha_j^* - \alpha_j\right) \left(x_i \cdot x_j\right) \end{aligned} \tag{5.39}$$

mit den Nebenbedingungen

$$0 \leq \alpha_i^* \leq C, 0 \leq \alpha_i \leq C$$
$$\sum_{i=1}^{l} \alpha_i^* = \sum_{i=1}^{l} \alpha_i$$

Somit lässt sich im Allgemeinen ein Prognosewert $y(x_i)$ mittels Gleichung 5.40 im Rahmen einer linearen Regression bestimmen (Drucker et al., 1997).

$$y\left(x\right) = \sum_{i=1}^{l} \left(\alpha_i^* - \alpha_i\right) \cdot \left(x_i \cdot x\right) + b \tag{5.40}$$

Regressionsprobleme lassen sich nicht zwangsläufig mit linearen Modellen sinnvoll lösen. Daher werden Kernel-Funktionen verwendet. Die duale Formulierung des Optimierungsproblems der SVM-R erlaubt die Integration nicht-linearer Funktionen. Das Konzept der Kernel-Funktionen wurde von Aizerman et al. (1964) eingeführt und später im Kontext von Large margin classifiers angewendet (Boser et al., 1992). Gleichung 5.40 lässt sich entsprechend der Gleichung 5.41 umformulieren, indem das Skalarprodukt der Eingangsvektoren durch eine Kernel-Funktion $K\left(x_i,x\right)$ in Abhängigkeit der Eingangsvektoren ersetzt wird.

$$y\left(x\right) = \sum_{i=1}^{l} \left(\alpha_i^* - \alpha_i\right) \cdot K\left(x_i,x\right) + b \tag{5.41}$$

Dabei sind ferner die Karush-Kuhn-Tucker(KKT)-Bedingungen gemäß der Gleichungen 5.42 bis 5.45 einzuhalten (Bishop, 2006).

$$\alpha_i \left(\varepsilon + \xi_i + y\left(x_i\right) - t_i\right) = 0 \tag{5.42}$$

$$\alpha_i^* \left(\varepsilon + \xi_i^* + y\left(x_i\right) - t_i\right) = 0 \tag{5.43}$$

$$\left(C - \alpha_i\right) \xi_i = 0 \tag{5.44}$$

$$\left(C - \alpha_i^*\right) \xi_i^* = 0 \tag{5.45}$$

Für die Lösung von Gleichung (5.41) bedarf es weiterhin der Schätzung von b. Diese lässt sich auf Grundlage der einzuhaltenden Karush-Kuhn-Tucker-Bedingungen bestimmen und resultiert in den Gleichungen (5.46) und (5.47). Die dazugehörige Herleitung kann Haykin, 2009 entnommen werden.

$$b = t_i - w^T x_i - \varepsilon \ f\ddot{u}r\ 0 < \alpha_i < C \tag{5.46}$$

$$b = t_i - w^T x_i + \varepsilon \ f\ddot{u}r\ 0 < \alpha_i^* < C \tag{5.47}$$

Im Rahmen der Studie von Koschwitz et al. (Koschwitz et al., 2018) wurden zur Entscheidungsunterstützung bei der Auswahl geeigneter Kernel-Funktionen verschiedene Kurvenanpassungen für einzelne Lastgangdaten durchgeführt.

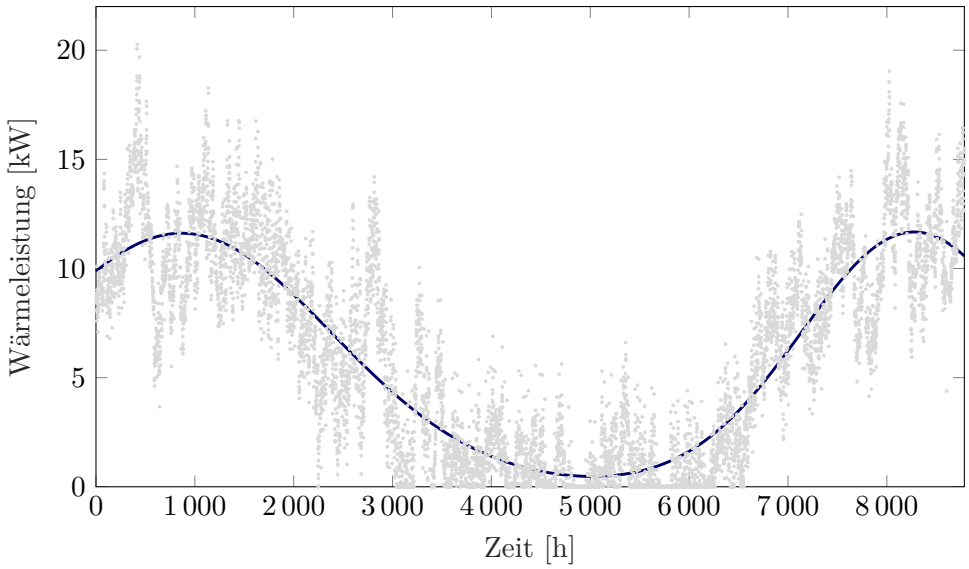


Abbildung 5.10: Kurvenanpassung mit einer Gauss-Funktion in Anlehnung an Koschwitz et al., 2018

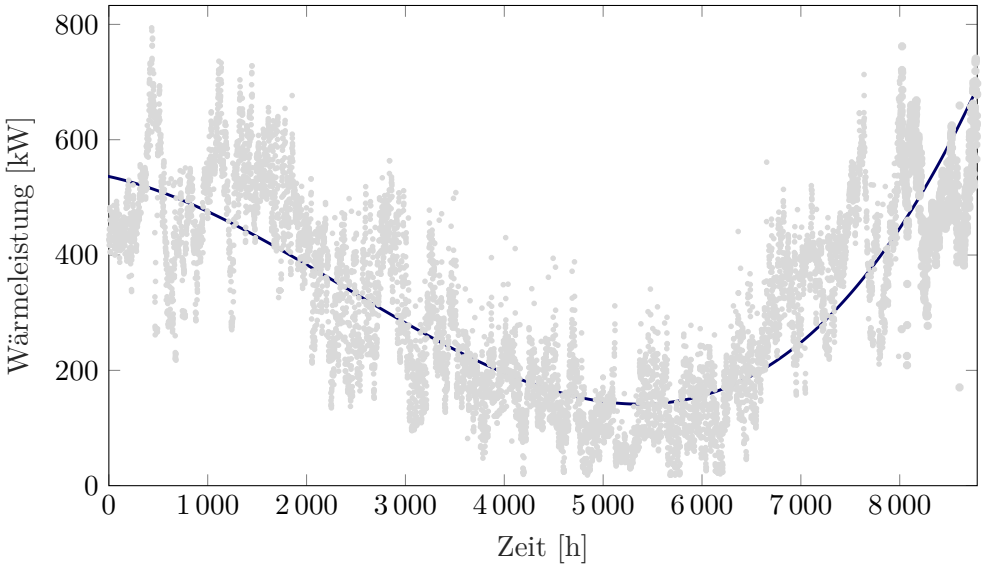


Abbildung 5.11: Kurvenanpassung mit einer polynomialen Funktion dritten Grades in Anlehnung an Koschwitz et al., 2018

Im Ergebnis zeigte sich eine explizite Eignung der Gaußschen Radialen Basisfunktion für das Anwendungsgebiet der thermischen Lastprognosen. Die dazugehörige Kurvenanpassung ist exemplarisch in Abbildung 5.10 dargestellt. Als Variation davon wurde zudem eine polynomiale Kernel-Funktion dritten Grades gewählt, um abweichende Lastcharakteristika von Gebäuden funktional besser abbilden zu können. Abbildung 5.11 veranschaulicht die dazugehörige exemplarischen Kurvenanpassung. Die Gaußsche Radiale Basisfunktion (RBF) ist nach Gleichung (5.48) und die polynomiale Kernel Funktion allgemein nach Gleichung (5.49) definiert, wobei x und x' zwei Eingangsvektoren und D den Grad der polynomialen Funktion repräsentieren (Clarke et al., 2005).

$$K\left(x,x'\right) = \exp\left(-\| \, x - x'\|^2\right) \tag{5.48}$$

$$K\left(x,x'\right) = \left(x \cdot x' + 1\right)^D \tag{5.49}$$

Im Kontext der modelltheoretischen Entwicklung des Prognose-Ensembles findet eine SVM-Konfiguration auf Grundlage von Gleichung 5.48 und eine beruhend auf Gleichung 5.49 Anwendung.

5.3.3 Entscheidungslogik

Die Auswahl des gebäudespezifisch geeigneten Prognosemodells aus dem vorgestellten Ensemble basiert auf der Beurteilung der Prognosegüte in der Modelltestphase. Letztere wird mittels eines Fehlerindices, welcher die Abweichungen zwischen y und t quantifiziert, bestimmt. In der Literatur werden verschiedene Indikatoren vorgeschlagen und verwendet (Amasyali et al., 2018). Für die Modellentwicklung dieser Arbeit eignet sich insbesondere der Mean Squared Error (MSE) gemäß Gleichung (5.50).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - t_i\right)^2 \tag{5.50}$$

Dies begründet sich in der Quadrierung der Abweichungen zwischen y und t und der damit implizierten verstärkten Gewichtung von großen Differenzen. Somit lässt sich bei der Methodenauswahl eine Fokussierung auf Stärken in der Spitzenlastprognose erreichen. Diese Prämisse ist entscheidend für die auf Leistungsspitzen beruhende Analyse von Gleichzeitigkeitseffekten. Abbildung 5.12 stellt zusammenfassend den Grundgedanken der Entscheidungslogik für das Prognose-Ensemble dar. Die Güte der gebäudespezifisch trainierten vier Prognosemethoden wird im Anschluss an den Validierungsprozess mit Hilfe der Testdaten anhand des MSE evaluiert. Sofern der MSE-Wert für den Test einer Methode

im Vergleich zu den MSE-Werten auf Grundlage der Anwendung der anderen Methoden den geringsten Wert annimmt, erfolgt die Verwendung der erstgenannten Methode für die gebäudespezifische Prognoseaufgabe.

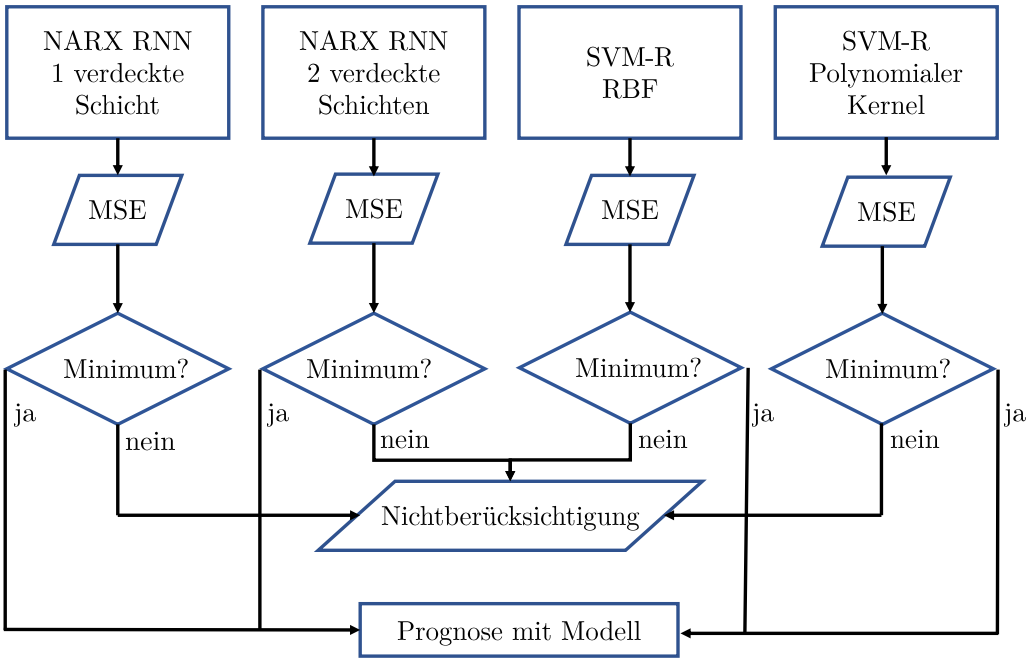


Abbildung 5.12: Schematische Darstellung der Entscheidungslogik für die gebäudespezifisch ausgewählte Prognosemethode auf Grundlage der MSE-Werte in der jeweiligen Modelltestphase

5.4 Teilmodell prädiktive Lastszenarien

Im Hinblick auf die Modellierung prädiktiver Szenarien für langfristige Planungszeiträume werden im Rahmen dieses Abschnitts die modelltheoretischen Grundlagen zur Abbildung und Quantifizierung von Lastzuständen infolge von Wetteränderungen und Gebäudesanierungen vorgestellt. Abbildung 5.13 veranschaulicht dazu die Szenarien, welche im Rahmen dieser Arbeit zur Demonstration zukünftiger Einflüsse auf Gleichzeitigkeiten in urbanen thermischen Lastverteilungen dienen. Den Ausgangspunkt bilden drei unterschiedliche Entwicklungspfade für die zugrunde liegenden Wetterbedingungen. Hierbei erfolgt eine Unterscheidung zwischen einem Wettertrend und zwei Extremwetterszenarien, welche durch einen außergewöhnlich kalten Winter bzw. einen außergewöhnlich warmen Sommer repräsentiert werden. Darauf aufbauend dient die Möglichkeit der Integration von Sanierungsfahrplänen auf Quartiersebene der Abbildung von Lastgangmodifikationen infolge

von Strukturänderungen in der betrachteten Region.

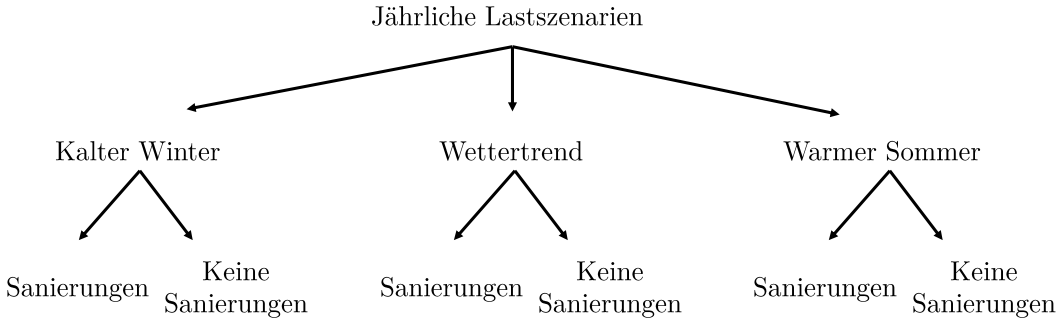


Abbildung 5.13: Schematische Darstellung der Modellszenarien

5.4.1 Zukünftige Wetterszenarien

Die folgenden Ausführungen dieses Unterabschnitts basieren auf Koschwitz et al., 2019. Für die Generierung unterschiedlicher Wetterszenarien finden die Testreferenzjahr (TRJ) Datensätze des Deutschen Wetterdienstes (DWD) Anwendung. Neben einem Datensatz zur Abbildung der gegenwärtigen ortsspezifischen Wetterbedingungen, welcher auf Beobachtungsdaten im Zeitraum von 1995 bis 2012 beruht, steht auch ein Zukunftswetterdatensatz für den Zeitraum 2031 bis 2060 zur Verfügung. Ferner sind für beide Kategorien sowohl Trend- als auch extreme Winter- und Sommerdaten vorhanden, sodass für den gegenwärtigen Zeitpunkt bis 2060 Wetterentwicklungen mit unterschiedlichen Ausprägungen begrenzt quantifizierbar sind. Modelltheoretisch erfolgt die Generierung verschiedener Wetterdatensätze für den Zeitraum zwischen der Gegenwart und 2031 für die Trend- und Extremszenarien mittels linearer Interpolation nach der allgemeinen Gleichung (5.51), wobei f der Funktionswert von x ist und x aus dem Zusammenhang zwischen $f(x)$ an der Stelle 0 und 1 bestimmt wird.

$$f(x) = f_0 + \frac{f_1 - f_0}{x_1 - x_0} \cdot (x - x_0) \tag{5.51}$$

Infolge der Übertragung des Grundgedankens der linearen Interpolation auf die Modellierung von Wetterszenarien, ergibt sich für jeden Datenpunkt eines Wetterdatensatzes der schematische Zusammenhang nach Abbildung 5.14, welcher sich am Beispiel der zukünftigen Außentemperaturentwicklungen formal verallgemeinert mittels Gleichung (5.52)

beschreiben lässt, wobei θ_A die Außentemperatur und τ den Zeitpunkt beschreibt.

$$\theta_A(\tau) = \theta_{A,0} + \frac{\theta_{A,1} - \theta_{A,0}}{\tau_1 - \tau_0} \cdot (\tau - \tau_0) \tag{5.52}$$

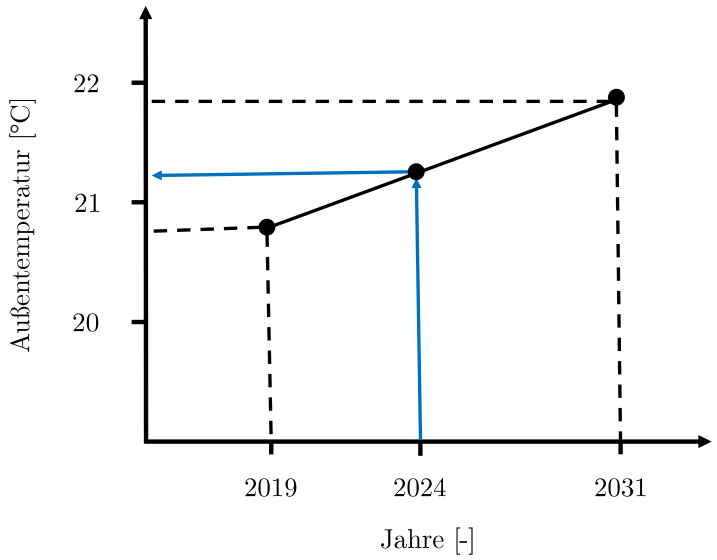


Abbildung 5.14: Schematische Darstellung der Interpolationsmethode zur Generierung zukünftiger Wetterdatensätze auf Basis der TRJ

Mit dieser Methode lässt sich der zeitliche Bezug des Betrachtungsjahres und des Referenzjahres in den Kontext einer numerischen Nähe des Wetterprofils im Betrachtungsjahr zu dem Wetterprofil des Referenzjahres setzen.

5.4.2 Lastgangbeeinflussungen durch Gebäudesanierungen

Die folgenden theoretischen Grundlagen können Koschwitz et al., 2019 entnommen werden. Infolge der Sanierung von Gebäuden ergeben sich neue thermische Eigenschaften der Gebäudehülle und ein angepasstes Nutzerverhalten. Auf urbaner Ebene steht jedoch in der Regel keine ausreichende Datenbasis als Informationsgrundlage zur Verfügung, um einzelne gebäudespezifische Sanierungsmaßnahmen modellieren und anschließend simulieren zu können. Dennoch sind Veränderungen von Lastgängen einzelner Gebäude maßgebend für die Abbildung zukünftiger Gleichzeitigkeitsfaktoren. Daher erfolgt die Generierung von Szenarien sanierungsbedingter Anpassungen zukünftiger Lastprofile auf Grundlage eines vereinfachten Prinzips. In diesem Zusammenhang dienen langfristige Lastprognosen auf Grundlage der vorgestellten Prognosemodelle als Ausgangspunkt für zukünftige Lastgangmodifikationen. Der gebäudespezifisch gemessene thermische Energieverbrauch

wird in einen witterungsabhängigen und einen witterungsunabhängigen Anteil $\dot{Q}_{wa,san}$ beziehungsweise $\dot{Q}_{wu}$ aufgeteilt. Dabei erfolgt die Ermittlung beider Parameter mittels linearer Regression nach der vorangestellten Methodik. Weiterhin werden die folgenden Annahmen für die Modellbildung getroffen:

I. Der zukünftige witterungsunabhängige thermische Leistungsbedarf $\dot{Q}_{wu}$ wird nicht von Sanierungsmaßnahmen beeinflusst und lässt sich anhand historischer Messdaten bestimmen.

II. Der zukünftige witterungsabhängige thermische Leistungsbedarf $\dot{Q}_{wa,san}$ eines Gebäudes nach der Sanierung wird im Vergleich zum thermischen Leistungsbedarf des selbigen vor der Sanierung um $\Delta\dot{Q}$ reduziert.

III. $\Delta\dot{Q}$ wird über das Verhältnis zwischen dem spezifischen Verbrauchswert $q_{wa,san}$ nach VDI 3807 und dem historischen spezifischen Verbrauchswert $q_{wa,hist}$ auf Grundlage von Messdaten approximiert.

Abbildung 5.15 visualisiert dahingehend die Methodik.

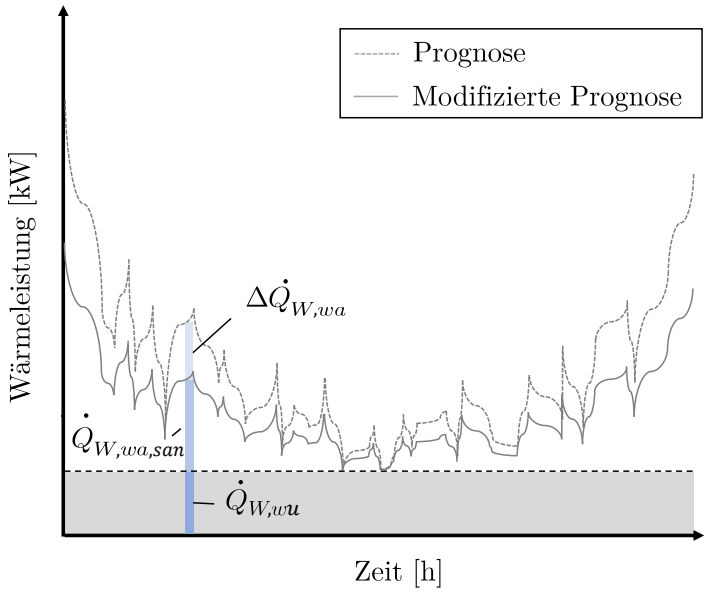


Abbildung 5.15: Schematische Darstellung der vereinfachten Lastprofilanpassung infolge von Gebäudesanierungen in Anlehnung an Koschwitz et al., 2019

Basierend auf der Annahme, dass die thermische Grundlast eines Gebäudes annähernd gleich bleibt, wird der witterungsabhängige Leistungsanteil Gleichung (5.53) entsprechend um $\Delta\dot{Q}$ verringert.

$$\Delta\dot{Q} = \left(1 - \frac{q_{wa,san}}{q_{wa,hist}}\right) \cdot \sum \dot{Q}_{wa,hist} \tag{5.53}$$

Als Grundlage hierfür dient das Verhältnis zwischen dem spezifischen Energieverbrauchswert eines Gebäudes auf Grundlage historischer Messdaten $q_{wa,hist}$ und einem spezifischen Referenzverbrauchskennwert als Richtwert für ein saniertes Gebäude $q_{wa,san}$ mit vergleichbarer Nutzung, jeweils in Abhängigkeit der Nettogrundfläche (NGF). Die prozentuale Abweichung p_i eines einzelnen witterungsunabhängigen thermischen Leistungswertes von der Summe aller witterungsabhängiger Leistungswerte lässt sich mit Gleichung (5.54) formal beschreiben.

$$p_i = \frac{\dot{Q}_{wa,hist,i}}{\sum_{j=1}^{n} \dot{Q}_{wa,hist,j}} \tag{5.54}$$

In der Folge lässt sich der thermische Leistungswert nach Gebäudesanierung $\dot{Q}_{wa,m,i}$ gemäß Gleichung (5.55) bestimmen.

$$\dot{Q}_{wa,san,i} = \dot{Q}_{wa,hist,i} - \Delta\dot{Q} \cdot p_i \tag{5.55}$$

Der Algorithmus zur Quantifizierung langfristiger Lastprognosen basiert auf dem folgenden zweistufigen Prozess:

I. Mit den gebäudespezifisch trainierten Prognosemodellen werden auf Basis der Wetterszenarien und prognostizierter Anwesenheitsprofile zukünftige Lastgänge simuliert.

II. Für jedes Gebäude wird überprüft, ob es saniert werden soll. Wenn dies zutrifft, erfolgt die Modifikation der prognostizierten Lastgänge nach dem obigen Prinzip ab dem ersten Jahr nach Wiederaufnahme des Betriebs.

Weiterhin wird die Annahme getroffen, dass das Gebäude während der Sanierungsphase nicht genutzt wird und kein thermischer Leistungsbedarf besteht.

5.5 Teilmodell Clustering und Klassifikation

Im Rahmen dieses Abschnitts werden die mathematischen Grundlagen des SOM- und LVQ-Algorithmus nach Kohonen vorgestellt, welche auf dem Konzept des kompetitiven Lernens basieren (Kohonen, 1990b). Der Grundgedanke hierbei beruht auf dem Vergleich zwischen einem zeitabhängigen Eingangsvektor $x(\tau) \in \mathbb{R}^n$ und einer Menge an k Referenzvektoren $\{m_i(\tau), \in \mathbb{R}^n, i = 1,2,...,k\}$, wobei τ der Zeitinstanz entspricht. Mittels eines Distanzmaßes lässt sich ferner der Abstand zwischen $x(t)$ und Referenzvektor $m_i(\tau)$ bestimmen. Auf Grundlage dessen kann jener Referenzvektor c mit dem geringsten Abstand zu $x(\tau)$ identifiziert werden. In der Folge ist die Distanz zu reduzieren, indem m_c in Richtung x verschoben wird, während die verbleibenden Referenzvektoren keine Anpassung erfahren.

(Kohonen, 1990b) Mit diesem sukzessiven Vorgehen lässt sich mittels der Referenzvektoren ein numerisches Abbild einer Menge an Eingangsvektoren im Sinne des Clusterings erzeugen. Die Referenzvektoren lassen sich zudem für die Klassifikation neuer Eingangsvektoren verwenden. In Abbildung 5.16 ist dazu die Methodik des Clusterings und der Klassifikation von Zeitreihen dargestellt.

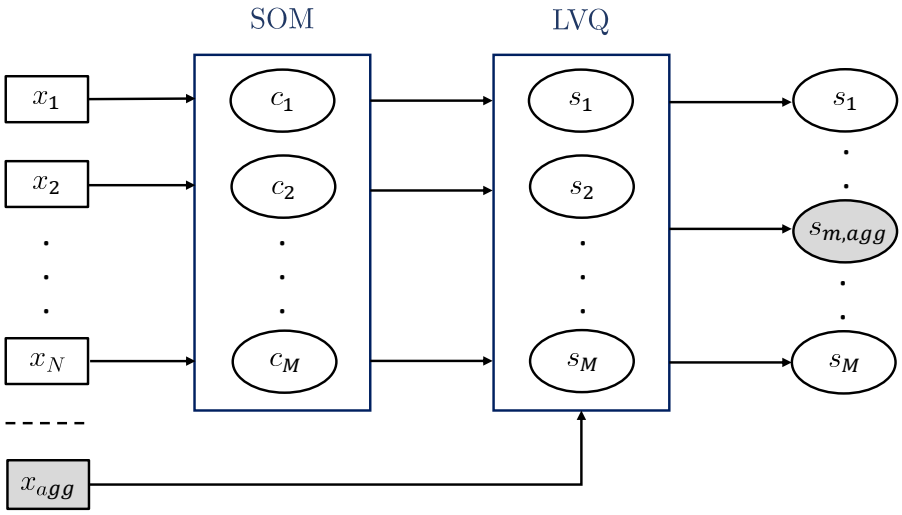


Abbildung 5.16: Ablaufschema zur automatisierten Identifizierung und Quantifizierung von charakteristischen gebäudespezifischen Lastprofilen mittels SOM und LVQ

Gegeben ist eine Menge an energetischen Lastgängen als Eingangsvektoren $\{x_j(\tau), \in \mathbb{R}^n, j = 1,2,...,p\}$. Mittels SOM-Clustering erfolgt die Zuordnung der Referenzvektoren $\{m_i(\tau), \in \mathbb{R}^n, i = 1,2,...,k\}$ nach dem Prinzip des kompetitiven Lernens, wobei ein Referenzvektor als Clusterzentrum mehrere Eingangsvektoren repräsentieren kann und auch leere Cluster möglich sind. So wird eine Menge an Clustern $l = 1,2,...,m$ ermittelt, denen entsprechende Label als eindeutige Zuordnungsschlüssel zugewiesen werden. Infolge der Zuordnung von Eingangsvektoren zu Labeln als Zielvariablen ergibt sich eine Datengrundlage für die Klassifikation neuer Eingangsvektoren. Im Rahmen der langfristigen Planung zukünftiger Energiesysteme ist es erforderlich, nicht nur Lastgänge eines Jahres zu clustern, sondern sowohl alle verfügbaren historischen Messdaten als auch prognostizierte zukünftige Lastgänge in den Prozess der Informationsextraktion zu integrieren. Um diese Anforderung methodisch abzubilden, werden die jährlichen Clusterzuordnungen in der Matrix $M^{m \times n}$ zusammengefasst. Die Zeilen repräsentieren dabei die Gebäudenummern innerhalb der Gebäudeanzahl n und die Spalten den Clusterzeitpunkt im Zeitraum T, welcher mehrere Jahre umfasst. Somit ergibt sich für eine beliebiges Gebäude m, dessen Lastprofil im Zeitpunkt τ geclustert wurde, die Notation $c_{m,\tau}$. Aus M werden im Anschluss gleiche Zeilen

zur Definition neuer zeitinvarianter Cluster identifiziert und in $\vec{v}$ übertragen. Gleichung (5.56) fasst letzteres zusammen.

$$M := \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,T} \\ c_{2,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ c_{n,1} & \cdots & \cdots & c_{n,T} \end{bmatrix} \rightarrow \vec{v} := \begin{pmatrix} c_1 \\ \vdots \\ \vdots \\ \vdots \\ c_n \end{pmatrix} \tag{5.56}$$

Gleichung (5.57) beinhaltet ferner beispielhaft eine Zuordnung von fünf Gebäuden zu zeitinvarianten Clustern.

$$M = \begin{bmatrix} 14 & 3 & 8 \\ 3 & 7 & 1 \\ 14 & 3 & 8 \\ 3 & 7 & 1 \\ 5 & 3 & 9 \end{bmatrix} \rightarrow \vec{v} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 3 \end{pmatrix} \tag{5.57}$$

Im Sinne der Identifizierung charakteristisch ähnlicher Lastgänge, welche damit eine Gleichzeitigkeit in ihren Verläufen aufweisen, erfolgt an dieser Stelle die Zuordnung des normierten Lastprofils der gesamten Gebäudegruppe zu einer Lastprofilklasse einer Subgebäudegruppe mittels LVQ.

5.5.1 Self-Organizing Map

Auf Grundlage gegebener Eingangs- und Referenzvektoren ist zunächst die Anordnung der Referenzvektoren als Neuronen festzulegen (Vesanto et al., 2000). Dahingehend sind verschiedene Topologien möglich, wobei hexagonale Strukturen zu bevorzugen sind, da sie nicht ausschließlich horizontale oder vertikale Modifikationen des elastischen Netzwerkes zur Abbildung der Eingangsvektoren zulassen (Kohonen, 2001). In Abbildung 5.17 wird eine hexagonale Anordnung der Neuronen exemplarisch dargestellt.

Ferner ist die Initialisierung der Referenzvektoren und damit die Startbedingung für den SOM-Algorithmus festzulegen (Kohonen, 2001). Es wird an dieser Stelle im Wesentlichen zwischen einer randomisierten und einer linearen Initialisierung unterschieden. Während bei ersterer eine zufällige Definition der Referenzvektoren gewählt wird, folgt zweitere einer näherungsweisen Vordefinition, sodass die Verteilung und Dichte der Referenzvektoren im Raum jener der Eingangsvektoren im Groben entspricht. In diesem Zusammenhang erfolgt zunächst die Ermittlung der Autokorrelationsmatrix von x, deren zwei Eigenvektoren mit den höchsten Eigenwerten einen zweidimensionalen Raum definieren, in welchem die

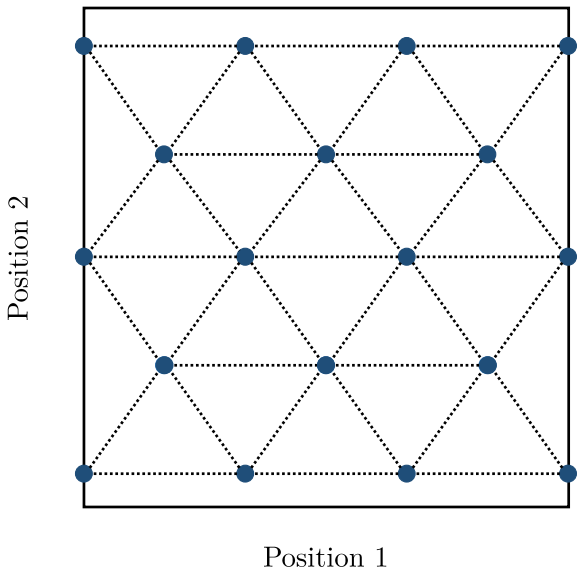


Abbildung 5.17: Schematische Darstellung einer hexagonalen Anordnung der Neuronen im Referenzvektorraum

Referenzvektoren in der zuvor festgelegten Struktur angeordnet werden. Das Zentrum des dazugehörigen Arrays fällt dabei mit dem Mittelwert von $x(\tau)$ zusammen (Kohonen, 2001). Nachfolgend wird die Anzahl der Referenzvektoren k festgelegt. In jedem Trainingsschritt sind für einen zufällig aus dem Eingangsdatensatz ausgewählten Vektor x_j die euklidischen Distanzen zu den Referenzvektoren nach Gleichung (5.58) zu berechnen, wobei m_c den Referenzvektor mit der geringsten Distanz zu x_j im stationären Zustand darstellt (Kohonen, 2001).

$$\|x_j - m_c\| = \min_i\{\|x_j - m_i\|\} \tag{5.58}$$

Für das Gewinnerneuron gilt somit nach Kohonen (Kohonen, 2001)

$$c = \operatorname*{argmin}_i\{\|x_j - m_i\|\} \tag{5.59}$$

Es besteht jedoch auch die Möglichkeit alle Neuronen innerhalb einer zu definierenden räumlichen Umgebung um das Gewinnerneuron herum N_c in den sequentiellen Adaptionsprozess zu integrieren. Letzteres wird in dem Konzept *The Batch Map* zusammengefasst (Kohonen, 2001) und in Abbildung 5.18 exemplarisch dargestellt.

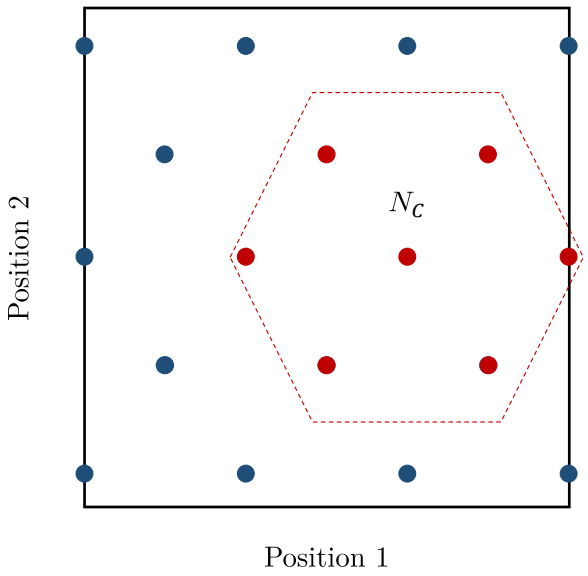


Abbildung 5.18: Exemplarische Darstellung von N_c innerhalb der hexagonalen Struktur des Referenzvektorraums

Darin erfolgt die Modifizierung der Neuronen innerhalb der ersten Nachbarschaftslinie von N_c. Im Anordnungsprozess wird die topologische Nachbarschaftsumgebung sukzessive verkleinert. In diesem Zusammenhang ist die initiale Anzahl an Nachbarschaftslinien ausreichend groß zu wählen, um kleinteilige Parzellierungen zu vermeiden. (Kohonen, 2001) Für die Modellentwicklung wird ein Radius von drei Nachbarschaftslinien um das Gewinnerneuron herum gewählt, welcher im Trainingsprozess sukzessive verringert wird. Für jeden Referenzvektor m_i innerhalb der definierten Nachbarschaftsumgebung gilt überdies die Modifikationsregel nach Gleichung (5.60) mit $\rho(\tau)$ als zeitabhängigen Faktor für die Lernrate (Kohonen, 2001).

$$m_i(\tau+1) = \begin{cases} m_i(\tau) + \rho(\tau) \cdot [x(\tau) - m_i(\tau)], & \text{für } i \in N_c(\tau) \\ m_i(\tau), & \text{für } i \notin N_c(\tau) \end{cases} \tag{5.60}$$

$\rho(\tau)$ ist ferner zwischen $(0 < \rho(\tau) < 1)$ definiert und monoton fallend über τ (Kohonen, 2001). Dabei kann $\rho(\tau)$ sowohl einer linearen als auch einer exponentiellen Funktion folgen, wobei keine wesentlichen Unterschiede im Ergebnis bestehen (Kohonen, 2001). Für die Modellentwicklung wird eine konstante Lernrate von 0,01 angesetzt, welche der Empfehlung entspricht, $\rho(\tau) < 0{,}02$ fest zu legen (Kohonen, 2001).

Um verschiedene Clustergranularitäten für die Gleichzeitigkeitsanalyse berücksichtigen zu können, erfolgt die Definition der folgenden drei Detailgrade:

I. *fein*: Anzahl der Neuronen entspricht ungefähr der Anzahl an gebäudebezogenen Lastgängen

II. *mittel*: Anzahl der Neuronen entspricht ungefähr der Hälfte der Anzahl an gebäudebezogenen Lastgängen

III. *gering*: Anzahl der Neuronen ergibt sich ungefähr zu $\sqrt{\frac{n}{2}}$ (Schätzer für unbekannte Clusteranzahl nach Han et al., 2012), wobei n der Anzahl an gebäudebezogenen Lastgängen entspricht.

Die Bildung feiner Cluster entspricht der Überlegung, dass (fast) jeder Referenzvektor einem Eingangsvektor zugeordnet werden kann und somit Referenzvektoren, die mehr als einem Eingangsvektor zugeordnet sind ein Cluster festlegen, das aus sehr ähnlichen Lastverläufen besteht. Der mittlere Detailgrad erlaubt eine etwas gröbere Gruppierung von Lastgängen, während der geringe Detailgrad tendenziell die Bildung großer Gruppen erlaubt.

5.5.2 Learning Vector Quantization

Mittels SOM lassen sich die Referenzvektoren definieren (Kohonen, 1990a), sodass der Abstand dieser zu einem beliebigen neuen Eingangsvektor bestimmt und dieser in der Folge klassifiziert werden kann. Es gibt dabei mit LVQ1, LVQ2 oder LVQ3 unterschiedliche Ausprägungen des lernenden Algorithmus, wobei die Ergebnisse von ähnlicher Genauigkeit geprägt sind (Kohonen, 2001). Im Rahmen der Modellentwicklung findet LVQ1 Anwendung. In einem iterativen Prozess wird dabei der Referenzvektor m_c mit dem geringsten Abstand zu Eingangsvektor x gesucht, wobei nachfolgend für die Aktualisierung von m_c Gleichung (5.61) gilt, sofern x und m_c zur selben Klasse gehören (Kohonen, 1990a).

$$m_c(\tau+1) = m_c(\tau) + \rho(\tau) \cdot (x(\tau) - m_c(\tau)) \tag{5.61}$$

Im gegenteiligen Fall ist Gleichung (5.62) anzusetzen (Kohonen, 1990a).

$$m_c(\tau+1) = m_c(\tau) - \rho(\tau) \cdot (x(\tau) - m_c(\tau)) \tag{5.62}$$

Jeder Referenzvektor m_i, der nicht m_c entspricht, erfährt gemäß Gleichung (5.63) keine Anpassung (Kohonen, 1990a).

$$m_i(\tau+1) = m_i(\tau) \quad \text{für } i \neq c \tag{5.63}$$

Mittels der Klassifizierung der Referenzvektoren besteht die notwendige Informationsgrundlage, um einen neuen Vektor x nach dem Prinzip der geringsten Distanz zuzuordnen.

5.6 Zusammenfassung

Das autoadaptive Modell dieser Arbeit basiert auf vier Teilmodellen zur Quantifizierung von Gleichzeitigkeitseffekten. Dazu dienen die eingeführten Zielvariablen in Form der erweiterten Definition des Gleichzeitigkeitsfaktors, eines Ähnlichkeitsfaktors und des Spitzenlastanteils von Gebäudegruppen. Der erste Teil des autoadaptiven Modells beinhaltet Methoden zur Datenvorbereitung. In diesem Kontext finden jene zur Identifizierung und Eliminierung von Anomalien in den Eingangsdaten sowie zur Nachverdichtung resultierender Fehlstellen in den bereinigten Datensätzen Eingang. Um einheitliche Bezugsgrößen herzustellen und den Modelltrainingsprozess zu beschleunigen, ist zudem die Datennormalisierung eine Komponente dieses Teilmodells. Die Erweiterung der Informationsbasis mittels Lastprognosen stellt den zweiten Teil des autoadaptiven Modells dar und besteht aus zwei NARX RNN unterschiedlicher Tiefe und zwei ε-SVM-R Konfigurationen. Die Auswahl der gebäudeindividuell geeigneten Prognosemethode aus dem Ensemble erfolgt mittels einer Entscheidungslogik auf Grundlage eines ermittelten Fehlerindices in der Methodentestphase. Zur numerischen Abbildung der Unsicherheit des zukünftigen thermischen Energiebedarfs werden prädiktive Lastszenarien benötigt, deren Modellierung am Beispiel von zukünftigen Wetterentwicklungen und Lastgangbeeinflussungen infolge von Gebäudesanierungen umgesetzt und im Rahmen des dritten Teilmodells integriert ist. Das vierte Teilmodell zur automatisierten Mustererkennung innerhalb einer Menge an Lastprofilen basiert auf einer Kombination aus SOM und LVQ. Mittels SOM werden zunächst Gebäudelastgänge in Cluster zusammengefasst. Die daraus resultierenden Standardlastprofile der Cluster dienen unter Anwendung von LVQ als Grundlage für die Zuordnung des charakteristischen Lastprofils auf Quartiersebene. Dadurch sind die Gebäude und Gebäudegruppen identifizierbar, deren Lastprofile jenem auf Quartiersebene am ähnlichsten sind. Die Zusammenführung der entwickelten Teilmodelle in Kombination mit den Methoden zur Quantifizierung von Gleichzeitigkeitseffekten ergibt ein Gesamtmodell, das auf Grundlage variierender Eingangsdaten aus heterogenen Quellen eine komplexe szenarienbasierte Informationsbasis generiert und diese auf Kennwerte zurückführt, welche für die prädiktive Planung und den Betrieb von zukünftigen urbanen Energiesystemen notwendig sind.

KAPITEL 6

Modellimplementierung

In diesem Kapitel ist die Implementierung des entwickelten autoadaptiven prädiktiven Modells dokumentiert, mit dem Namen Auto-adaptive Model for Simultaneity Analysis (AMSA). In Abschnitt 6.1 erfolgt eingangs die Formulierung softwaretechnischer Anforderungen an *AMSA*. Nachfolgend wird mit Abschnitt 6.2 die Umsetzung der Modellimplementierung im Sinne der Softwarearchitektur, Schnittstellen und Benutzeroberflächen als zentrale Steuerungsmodule erläutert. Abschnitt 6.3 dokumentiert die Eingliederung von *AMSA* in ein bestehendes und stetig weiterentwickeltes webbasiertes Energie-Informationssystem auf urbaner Ebene. Abschließend werden in Abschnitt 6.4 die Inhalte dieses Kapitels zusammengefasst.

6.1 Anforderungen

An dieser Stelle erfolgt zunächst die Zusammenfassung der hergeleiteten inhaltlichen Anforderungen an die Modellentwicklung auf Grundlage der Kapitel 2 bis 5.

I. Inhaltliche Abbildung und Verknüpfung der Komponenten des autoadaptiven Modells

II. Strukturierte Aufbereitung der Gleichzeitigkeitsanalysen

III. Szenariobasierte Ergebnisausgabe unter Berücksichtigung von Unsicherheit

IV. Gut interpretierbare Ergebnisse auf Grundlage komplexer und dynamischer Datengrundlagen

Darüber hinaus dienen die folgenden softwaretechnischen Anforderungen der Sicherstellung von Weiterentwicklungsmöglichkeiten und der Integrationsfähigkeit von *AMSA* innerhalb entwickelter Energie-Informationssysteme.

I. Modularer Aufbau von *AMSA* mittels Klassenstrukturen

II. Intuitiver Programmablauf

III. Automatisierter Funktionsablauf

IV. Übersichtliche grafische Benutzeroberflächen

V. Aufbau von Datenstrukturen

VI. Sicherstellung der Integrierbarkeit in bestehende Energie-Informationssysteme

Vor diesem Hintergrund wird im weiteren Verlauf dieses Abschnitts die Umsetzung der formulierten Anforderungen vorgestellt und erläutert.

6.2 Softwarearchitektur

Die Implementierung des Modells wird mittels der Entwicklungsumgebung *MATLAB* realisiert. Abbildung 6.1 stellt die Softwarearchitektur dar. Letztere dient im Rahmen dieses Kapitels als Übersicht, wobei Anhang B die detaillierte Klassenhierarchie (Abschnitt B.1) inklusive Funktionsbeschreibungen (Abschnitt B.2) zu entnehmen ist.

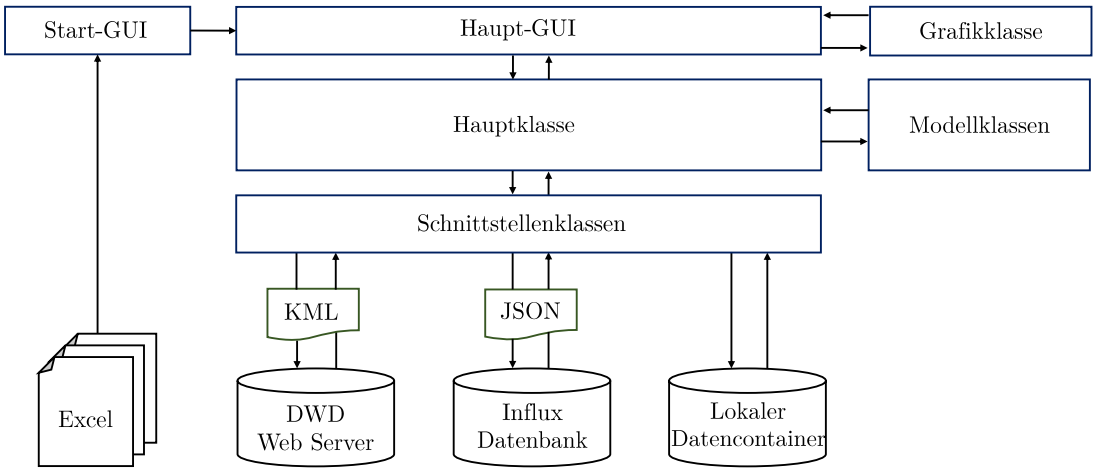


Abbildung 6.1: Schematische Darstellung der entwickelten Softwarearchitektur von *AMSA*

Die Architektur besteht aus verschiedenen Klassen, die grafisch in funktionalen Teilbereichen zusammengefasst sind. Mittels der Start-GUI werden verschiedene Einstellungen hinsichtlich der zur Verfügung stehenden Datengrundlage vom Benutzer abgefragt und ergänzende Informationen in Form von Excel-Listen in das Programm geladen. Die Haupt-GUI stellt die zentrale Interaktionsplattform von *AMSA* dar. Für die grafische Aufbereitung von Daten und Ergebnissen erfolgt ein Zugriff auf die Grafikklasse, in der die dafür vorgesehenen Funktionen enthalten sind. Über die Hauptklasse werden sequentielle Funktionsabläufe innerhalb von Modell- und Schnittstellenklassen gesteuert sowie Informationen für die Haupt-GUI zur Visualisierung und Weiterverarbeitung aufbereitet. Die Schnittstellenfunktionen von *AMSA* definieren und regeln die Kommunikation mit externen Datenquellen sowie Datenspeichersystemen. Kurzfristige Wetterprognosen werden von dem Web Server des DWD abgerufen und im Keyhole Markup Language (KML)-Format übertragen. Mittels Dokumenten im Format JavaScript Object Notation (JSON) findet der Zeitreihenaustausch

zwischen *AMSA* und der Influx Datenbank statt. Überdies dienen lokale Datencontainer der (temporären) Speicherung von trainierten Modellen des Prognose-Ensembles.

6.2.1 Grafische Benutzeroberflächen

Im Folgenden werden die beiden grafischen Benutzeroberflächen von *AMSA* vorgestellt. Nach Programmstart sind zunächst initiale Einstellungen zu Datenquellen vorzunehmen. Dazu öffnet sich ein Dialogfenster zur Nutzerinteraktion, das in Abbildung 6.2 dargestellt ist. Im Bereich *Connection to Influx DB* ist neben der URL, unter welcher die Datenbank zum Import und optionalen Export von Zeitreihen zu erreichen ist, der Name der Datenbank sowie der Nutzername mit dem dazugehörigen Passwort anzugeben. In dem darunter angeordneten Abschnitt *Sensor Designations* ist sowohl eine Liste mit den gebäudebezogenen und jene mit den wetterspezifischen Sensorbezeichnungen für die Datenbankabfragen jeweils als Excel-Liste zu importieren. Abschließend ist im Bereich *General Building Information* eine Excel-Liste mit den Gebäudenamen und -koordinaten in das Programm zu laden.

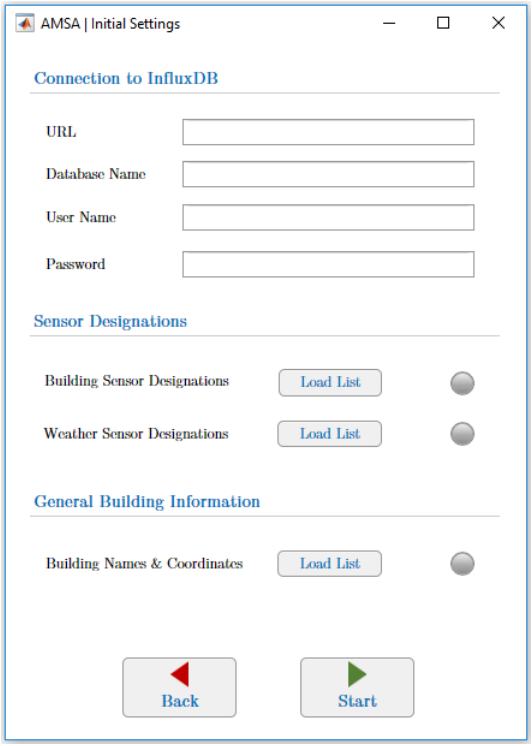


Abbildung 6.2: Darstellung der Benutzeroberfläche für initiale Einstellungen bezüglich der Datenquellen nach Aufruf von *AMSA*

In den beiden letztgenannten Bereichen der Benutzeroberfläche sind zudem Kontrollleuchten integriert, welche die Farbe grün annehmen, sofern der Datenimport erfolgreich war und andernfalls die Farbe rot. Nur wenn die erforderlichen Datenimporte erfolgreich waren,

lässt sich mittels des *Start*-Buttons das Hauptfenster von *AMSA* öffnen. Gleichzeitig erfolgt die Sperrung der Interaktionsfelder der ersten Benutzeroberfläche. Diese kann jedoch wieder für Änderungen über den *Back*-Button reaktiviert werden. Die Auswahloptionen im Hauptfenster von *AMSA* sind in Tabelle 6.1 aufgelistet.

Tabelle 6.1: Übersicht zu den Auswahloptionen im Hauptfenster von *AMSA*

Bezeichnung	**Auswahloptionen**
Energy Type	Heat
	Cold
Forecast Horizon	Long-term
	Medium-term
	Short-term
Prediction Period	24h, 48h, 72h, 96h
	120h, 144h, 168h
Climate Scenario	Trend
	Extreme Winter
	Extreme Summer
Retrofit Scenario	Include Retrofit Order
	None
First Year of Training Period	-
Last Year of Training Period	-
Station ID	-
First Year of Prediction Period	-
Last Year of Prediction Period	-
Model Setting	Train & Save Forecast Models
	Load Trained Forecast Models
Cluster Granularity	Low
	Medium
	High
Analysis Period	Include Historical Data
	Exclude Historical Data
Number of Historical Peaks	-
Number of Predicted Peaks	-
Write Prediction to Time Series Database	-
Save Predictions in a File	-
Save Diversity Factors in a File	-

Das Hauptfenster, welches in Abbildung 6.3 dargestellt ist, weist eine Zweiteilung auf. Der in vier Kategorien gegliederte Eingabebereich im linken Teil der Benutzeroberfläche dient der Spezifikation von Modelleingangsdaten, Modellkonfigurationen sowie der Festlegung von Ergebnissen, die exportiert werden sollen.

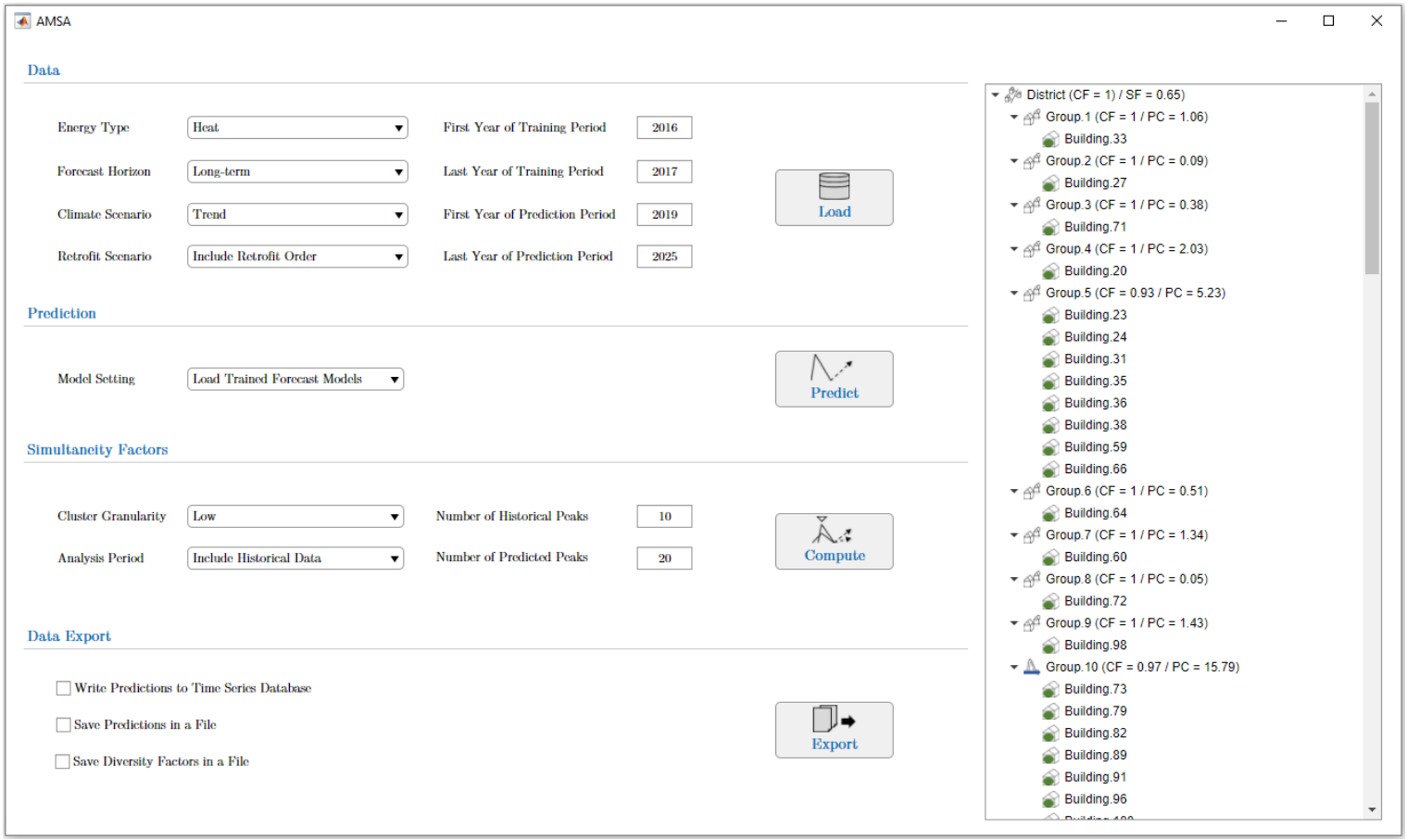


Abbildung 6.3: Benutzeroberfläche von *AMSA*

Der rechte Teil der Benutzeroberfläche ermöglicht die strukturierte Darstellung von numerischen Ergebnissen. Darüber hinaus werden zusätzliche grafische Ergebnisse infolge der Auswahl einzelner Elemente angezeigt. Dazu werden die Struktur und Funktionsweisen einzelner Komponenten näher erläutert. In der Kategorie *Data* ist zunächst mittels der Dropdown-Liste *Energy Type* zwischen der Analyse von Wärme- und Kühlleistung (*Heat* und *Cold*) zu unterscheiden. Nach der Auswahl eines Listenelements geht es in der nachfolgenden Dropdown-Liste *Forecast Horizon* um die Spezifikation des Zeithorizonts für die Lastprognosen, auf denen später die Quantifizierung der Gleichzeitigkeitseffekte beruht. Es besteht hier die Möglichkeit zwischen einem kurz-, mittel- und langfristigen Zeitraum, auf Grundlage der dazugehörigen Listenelemente *Short-term*, *Medium-term* sowie *Long-term*, zu wählen. An dieser Stelle verändern sich die nachfolgenden Dropdown-Listen in Abhängigkeit von der Auswahl der letztgenannten Listenelemente, wie im Folgenden erläutert: Neben den vorgestellten Dropdown-Listen sind zwei Eingabefelder mit den Bezeichnungen *First Year of Training Period* und *Last Year of Training Period* für die Festlegung des Trainingsdatenumfang angeordnet. Im erstgenannten Feld ist das Jahr zu Beginn und im zweitgenannten jenes zum Ende des Zeitraumes einzutragen. Hinsichtlich der dynamischen Anpassungen der Elemente im Bereich *Data* ist in Abbildung 6.4 das Erscheinungsbild dieser für die Auswahl eines kurzfristigen Prognosezeitraumes dargestellt.

Abbildung 6.4: Darstellung des dynamisch veränderlichen Teils der Benutzeroberfläche in der Kategorie *Data* bei der Auswahl des kurzfristigen Prognosezeitraumes

Unter *Forecast Horizon* ist die veränderliche Dropdown-Liste *Prediction Period* angeordnet. Hier ist der kurzfristige Prognosezeitraum näher zu definieren, indem zwischen den Einträgen

I. 24 h
II. 48 h
III. 72 h
IV. 96 h
V. 120 h
VI. 144 h
VII. 168 h

zu wählen ist. Es besteht somit die Möglichkeit Gleichzeitigkeitseffekte sowohl für die kommenden Stunden als auch für die folgende Woche zu analysieren. Ergänzend dazu erfordert dieser Pfad die Eingabe der Kennnummer der zu den Untersuchungsobjekten

nächstgelegenen Wetterstation des DWD. Nach Auswahl des mittelfristigen Prognosezeitraumes, wird die Dropdown-Liste *Prediction Period* durch jene mit der Bezeichnung *Climate Scenario* ersetzt. Abbildung 6.5 visualisiert diese Modifikation.

Abbildung 6.5: Darstellung des dynamisch veränderlichen Teils der Benutzeroberfläche in der Kategorie *Data* bei der Auswahl des mittelfristigen Prognosezeitraumes

Hier ist zwischen einer im Mittel zu erwartenden Wetterentwicklung (Element *Trend*), einer mit extremen Winter- (Element *Extreme Winter*) und extremen Sommermonaten (Element *Extreme Summer*) zu wählen. Der mittelfristige Prognosezeitraum ist im Rahmen dieser Arbeit für ein Jahr definiert, sodass dieses in dem Eingabefeld *First Year of Training Period* eingegeben werden muss. Letzteres wird automatisch auch als Wert dem korrespondierenden Eingabefeld *Last Year of Training Period*, das nicht editierbar ist, zugewiesen. Somit ist sichergestellt, dass das festgelegte Kriterium für mittelfristige Prognosezeiträume eingehalten wird. Für langfristige Zeiträume erfolgt eine Entsperrung des Eingabefeldes.

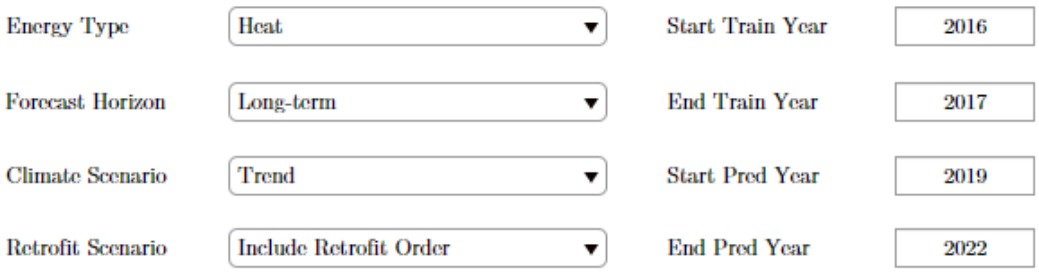


Abbildung 6.6: Darstellung des dynamisch veränderlichen Teils der Benutzeroberfläche in der Kategorie *Data* bei der Auswahl des langfristigen Prognosezeitraumes

In diesem Kontext erscheint zudem eine vierte Dropdown-Liste mit der Bezeichnung *Retrofit Scenario*, welche in Abbildung 6.6 dargestellt ist. Hier besteht mit der Auswahl der Option *Include Retrofit Order* die Möglichkeit, eine Sanierungsreihenfolge für die Gebäude vorzugeben und entsprechende Lastgangmodifikationen in die Gleichzeitigkeitsanalyse mit einzubeziehen. Der Aufbau der importierbaren .mat-Datei ist Anhang B, Abschnitt B.3 zu entnehmen. Damit verknüpft ist der Aufruf eines Dialogfensters, in dem die dazugehörige Informationsquelle als Datei ausgewählt werden muss. Die Auswahloption *None* kennzeichnet, dass kein Sanierungsszenario integriert wird. Unter *Data* ist die Kategorie *Prediction* angeordnet. Hier muss mittels der Dropdown-Liste *Model Setting* angegeben

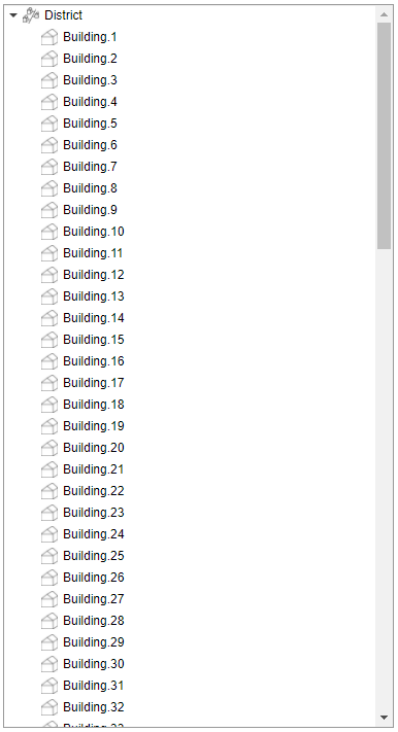
werden, ob die Prognosemodelle trainiert und gespeichert werden sollen (Element *Train & Save Forecast Models*) oder ob entsprechend trainierte Modelle vorliegen (Element *Load Trained Forecast Models*), die geladen werden können. Die letztgenannte Option verkürzt die Rechendauer dabei erheblich. In der Kategorie *Simultaneity Factors* sind Einstellungen für die Durchführung der Quantifizierung von Gleichzeitigkeitseffekten in der thermischen Lastverteilung des Quartiers erforderlich. Mittels der Dropdown-Liste *Cluster Granularity* ist der Detaillierungsgrad des Clusterings entsprechend Kapitel 5 vorzugeben. Es sind dafür die dazugehörigen Listenelemente *Low*, *Medium* und *High* zur Auswahl implementiert. Mit der darunter angeordneten Dropdown-Liste *Analysis Period* wird anschließend der Zeithorizont der Ähnlichkeitsanalyse auf Grundlage des Clusterings und der Klassifikation festgelegt. Es gibt hier mit den Elementen *Include Historical Data* und *Exclude Historical Data* die Möglichkeit die Trainingsdaten mit in die Analyse zu integrieren bzw. sie von dieser auszuschließen. Weiterhin sind die Anzahl der historischen und prognostizierten Leistungsspitzen auf Quartiersebene, welche für die Berechnung der Gleichzeitigkeitsfaktoren sowie gebäude- bzw. gebäudegruppenspezifische Anteile an den Leistungsspitzen auf Quartiersebene benötigt werden, anzugeben. Für den Fall, dass die Trainingsdaten von der Analyse ausgeschlossen werden, wird das Eingabefeld *Number of Historical Peaks* gesperrt. In der letzten Kategorie auf der linken Seite der Benutzeroberfläche von *AMSA* mit der Bezeichnung *Data Export* können die generierten numerischen Ergebnisse exportiert werden. Infolge der Markierung der Checkbox *Write Predictions to Time Series Database* wird die Wahl getroffen, die Zeitreihen der Lastprognosen in die Influx Datenbank zu schreiben. Die weiteren zwei Checkboxen ermöglichen es Ergebnisse als .mat-Datei zu exportieren.

Die rechte Seite der grafischen Benutzeroberfläche weist eine dynamisch-adaptive Baumstruktur zur Ergebnisdarstellung auf, welche in Abbildung 6.7 dargestellt ist. Die einzelnen Interaktionselemente sind auswählbar und öffnen Fenster mit grafischen Aufbereitungen der Ergebnisse in Abhängigkeit von dem Dateneingabe- und Analysefortschritt auf der linken Seite der Oberfläche. So werden die Gebäude einer Gebäudegruppe namentlich im Anschluss an den Ladevorgang der Datenbasis in *AMSA* aufgelistet (siehe Abbildung (a)). Die Auswahl eines Gebäudes aus der Liste bedingt einen Aufruf eines Grafikfensters mit der Darstellung des dazugehörigen thermischen Lastgangs. Analog dazu wird bei Auswahl der gesamten Gebäudegruppe *District* die aggregierte Zeitreihe angezeigt. Im Anschluss an die Prognose zukünftiger Lastgänge erfolgt eine Kennzeichnung der Gebäude nach dem folgenden Prinzip (siehe Abbildung (b)):

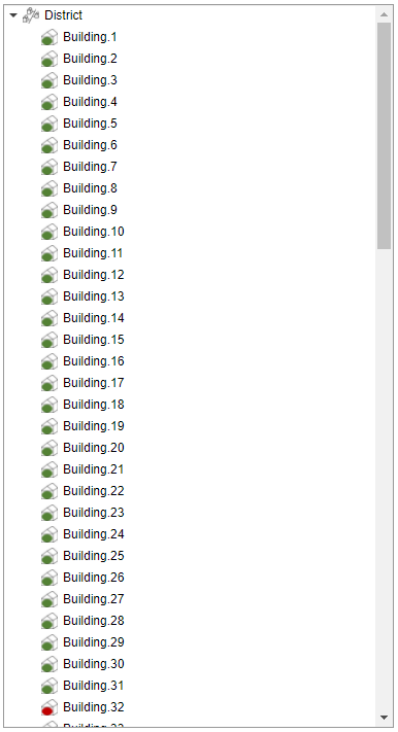
I. grüne Markierung: Lastprognose war erfolgreich

II. rote Markierung: Lastprognose ist fehlgeschlagen

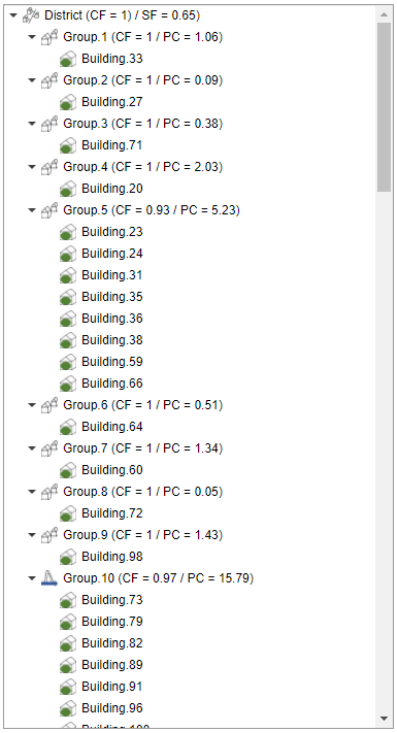
Die Auswahl eines Gebäudes oder der gesamten Gebäudegruppe ermöglicht in diesem Zustand der Baumstruktur die Darstellung der prognostizierten thermischen Lastgänge. Nach Abschluss der Gleichzeitigkeitsanalyse erfolgt eine Modifikation der Baumstruktur, indem eine weitere Hierarchieebene in Form von Subgebäudegruppen integriert wird.



(a) Zustand der Baumstruktur nachdem die Datengrundlage geladen wurde



(b) Zustand der Baumstruktur nachdem die Lastprognosen abgeschlossen wurde



(c) Zustand der Baumstruktur nachdem die Gleichzeitigkeitsanalyse abgeschlossen wurde

Abbildung 6.7: Darstellung der verschiedenen Zustände der Baumstruktur auf der Haupt-GUI von *AMSA*

Die Gruppierung der Gebäude stellt das Ergebnis des Clustering- und Klassifikationsprozesses dar. Jene Subgruppe mit dem zum aggregierten Lastprofil der Gesamtgruppe ähnlichsten Lastprofil wird mittels des folgenden Symbols gekennzeichnet:

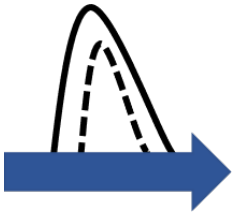


Abbildung 6.8: Darstellung des Symbols zur Kennzeichnung der Subgebäudegruppe mit dem ähnlichsten Lastprofil zu jenem der Gesamtgebäudegruppe

Auf den verschiedenen Gruppenebenen sind zudem die maximalen Gleichzeitigkeitsfaktoren (CF) und Spitzenlastanteile (PC) sowie auf der Ebene *District* der Ähnlichkeitsfaktor (SF) angegeben.

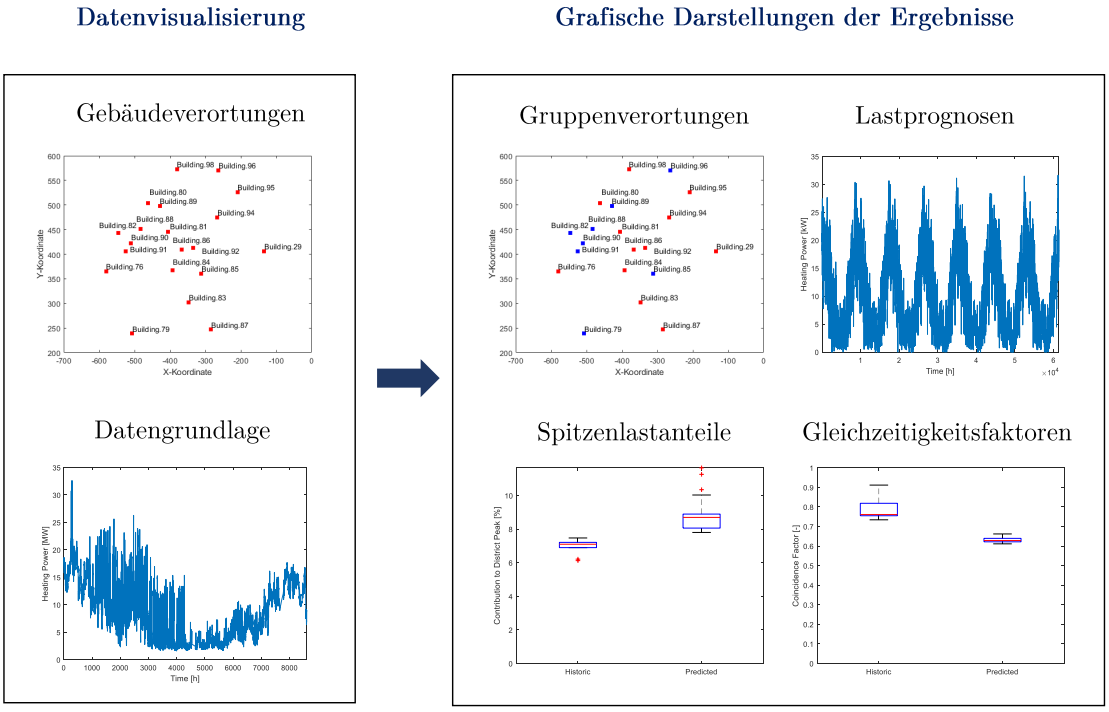


Abbildung 6.9: Exemplarische Darstellung der Datenvisualisierung und grafischen Ergebnisausgabe in *AMSA*

Nach Auswahl eines Gruppen-Interaktionselements werden in einem Grafikfenster neben dem prognostizierten Lastgang die geometrische Verortung des Gebäudes innerhalb

der Gesamtgebäudegruppe sowie Bandbreiten für den Gleichzeitigkeitsfaktor und den Spitzenlastanteil mittels Boxplots dargestellt. Letztere geben die Ergebnisse der Unsicherheitsanalyse auf Grundlage der auf der linken Seite der Benutzeroberfläche angegeben Anzahl zu untersuchender Leistungsspitzen wieder. Auf Gesamtgruppenebene können sowohl die Verortungen der einzelnen Gruppen im Quartier und die aggregierte Lastprognose als auch die Bandbreite des Gleichzeitigkeitsfaktors abgerufen werden. Abbildung 6.9 zeigt die grafische Aufbereitung der Daten und Ergebnisse in *AMSA* exemplarisch.

6.2.2 Klassenhierarchie

Die Klassenstruktur von *AMSA* weist verschiedene Hierarchiestufen auf, die im weiteren Verlauf dieses Abschnitts zusammen mit den dazugehörigen Kernaufgaben jeder Klassenebene erläutert werden. An dieser Stelle sei für detailliertere Informationsaufbereitungen im Sinne des vollständigen UML-Diagramms und Funktionsbeschreibungen auf Anhang B (Abschnitte B.1 und B.2) dieser Arbeit verwiesen.

Klassen der grafischen Benutzeroberflächen

Mit dem Aufbau von grafischen Interaktionsstrukturen für die Steuerung von Funktionsabläufen durch den Benutzer von *AMSA* geht die Verknüpfung von vorgesehenen Schaltflächen mit Modellfunktionen einher. Ferner sind die Anordnung von *Buttons* und anderen Interaktionskomponenten sowie verschiedene Zustände der grafischen Oberfläche festzulegen. Vor diesem Hintergrund lassen sich die Kernaufgaben der dazugehörigen Klassen wie folgt zusammenfassen.

I. Aufbau der GUI bei Programmstart
II. Ausblenden der GUI nach Schließen des Programms
III. Initiale Definition der GUI-Komponenten
IV. Veränderung der Eigenschaften einzelner GUI-Komponenten bei Benutzerinteraktion
V. Herstellung von Verbindungen zwischen der Hauptklasse und den GUI-Komponenten

Hauptklasse

Die Hauptklasse *main* fasst verschiedene Berechnungsabläufe von *AMSA* mittels verschiedener Funktionen zusammen und fungiert damit als zentrale Instanz für den Datentransfer und die Modellsimulation. Sie erfüllt die folgenden Kernaufgaben:

I. Datenimport
II. Modelltraining
III. Modellsimulation

IV. Quantifizierung der Gleichzeitigkeitsparameter

V. Datenexport

Als Grundlage hierfür dienen Modell- und Schnittstellenklassen, die im Folgenden vorgestellt werden.

Modellklassen

Die im Sinne der implementierten Hierarchie unter der Hauptklasse verorteten Modellklassen definieren die Methoden des autoadaptiven prädiktiven Modells. Die dazugehörigen Funktionen sind thematisch in einzelnen Klassen zusammengefasst, welche die folgenden Aufgaben ausführen:

I. Datenvorverarbeitung

II. Training und Simulation des Prognose-Ensembles

III. Quantifizierung von Wetterszenarien und Lastgangmodifikationen infolge von Gebäudesanierungen

IV. Clustering der thermischen Lastprofile

V. Training und Simulation des Klassifikationsmodells

VI. Quantifizierung von Gleichzeitigkeits- und Ähnlichkeitsfaktoren sowie gruppenspezifischen Spitzenlastanteilen

Schnittstellenklassen

Der Datentransfer wird mittels drei Schnittstellenklassen realisiert. Während die Daten vom DWD Webserver im KML-Format abgerufen werden, findet der Datenaustausch zwischen *AMSA* und der Influx Datenbank über JSON Dokumente statt. Nachfolgend sind die Aufgaben der Schnittstellenklassen zusammengefasst:

I. Regelung des Zeitreihendatentransfers zwischen *AMSA* und der Influx Datenbank

II. Abruf von Zeitreihen vom DWD-Server

III. Erstellen und Laden von lokalen Datencontainern

6.2.3 Datenstrukturen

Lokale, hierarchisch aufgebaute Datencontainer dienen der (temporären) Speicherung der gebäudespezifisch trainierten Modelle aus dem Prognose-Ensemble. Infolge der Integration dieses Datenhaltungssystems, müssen die Nonlinear Autoregressive Network with Exogenous Inputs (NARX) RNN und SVM-R-Modelle nicht bei jeder durchzuführenden Analyse neu trainiert, sondern können in *AMSA* importiert und anschließend simuliert werden.

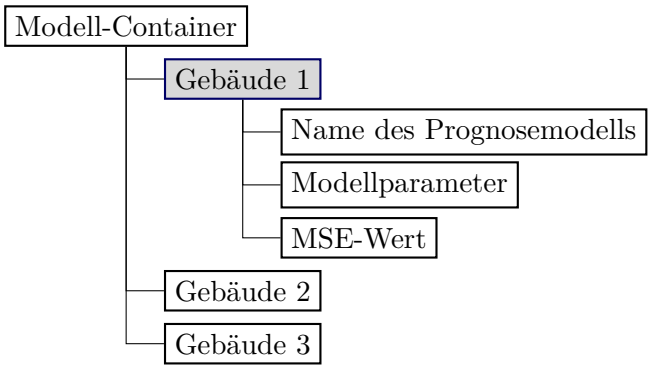


Abbildung 6.10: Exemplarische Darstellung des Aufbaus von lokalen Datencontainern zur Speicherung von gebäudespezifisch trainierten Prognosemodellen

Abbildung 6.10 zeigt die prinzipielle Struktur der Modellcontainer.

Mittels der in Matlab hinterlegten Baumstruktur eines *Cell Array* lässt sich automatisiert für jedes Gebäude das Modell mit dem dazugehörigen MSE-Wert hinterlegen, welches sich aufgrund der implementierten Entscheidungslogik im Rahmen des Ensembles am Besten für die Prognoseaufgabe eignet. Die Bezeichnung der im *.mat*-Format speicherbaren Datencontainer folgt der Konvention: Die als *mdlContainer* bezeichneten Dateien erhalten als Namenszusatz den Buchstaben *H* bzw. *C*, als Kennzeichnungsschlüssel für Modelle zur Prognose von Wärme- bzw. Kühllasten. Als weitere Namensergänzung folgen je nach gewähltem Zeithorizont die Schlüsselbegriffe *LTP* für einen langfristigen, *MTP* für einen mittelfristigen und *STP* für einen kurzfristigen Zeitraum. Die vollständige Dateibezeichnung für einen Datencontainer, der Prognosemodelle für kurzfristige Vorhersagen gebäudescharfer Wärmelasten beinhaltet, lautet beispielsweise *mdlContainer-H-STP*.

Für die strukturierte Speicherung von Zeitreihen dient die Open-Source-Datenbank *Influx Data*, welche als zentrale Austauschplattform fungiert. Zum einen lassen sich beispielsweise die zugrundeliegende Datenbasis zum thermischen Energieverbrauch oder Zeitreihen von Wettervariablen strukturiert ablegen, zum anderen können Lastprognosen zur Weiterverarbeitung innerhalb eines Informationssystems gespeichert werden. Die Kennzeichnungsschlüssel für die Zeitreihen sind Anhang B, Abschnitt B.3 zu entnehmen.

6.3 Eingliederung in ein Energie-Informationssystem

Das im Rahmen dieser Arbeit entwickelte Software-Tool *AMSA* gliedert sich in ein Energie-Informationssystem ein, das aus verschiedenen Software-Tools, Datenbanken und einer webbasierten Visualisierungsplattform besteht. Die Informationsbereitstellung und -aufbereitung von energetischen Zuständen urbaner Regionen nimmt eine wichtige Rolle im Rahmen des Informationsaustausches zwischen den Stakeholdern ein und sorgt nicht zuletzt

für Transparenz in der Planung und Entwicklung von Energieversorgungsgebieten. Dadurch lassen sich Szenarien auf Grundlage von Gebäudesanierungen oder Neubauten bildlich darstellen, wodurch Informationen auf die Kernaussagen reduzierbar sind (Koschwitz et al., 2017). Vor diesem Hintergrund werden Energie-Informationssysteme benötigt, die einerseits webbasierte Interaktions- und Visualisierungsfunktionen erfüllen und andererseits angegliederte Datenanalyse- und Optimierungswerkzeuge umfassen.

Auf urbaner Ebene stehen detaillierte gebäudebezogene Informationen in der Regel nicht zur Verfügung. Dies gilt im Besonderen für den Wohngebäudebestand, welcher beispielsweise nicht flächendeckend mit Sensortechnik ausgestattet ist, sodass detaillierte Monitoringdaten fehlen. Daran anknüpfend sind in der Regel auch keine ausreichenden Informationen über die thermischen Eigenschaften, Materialien oder das Nutzerverhalten vorhanden (Al-Homoud, 2001). Vor dem Hintergrund dieser Problemstellungen wurde das Software-Tool SCoTED (Simplified Calculation of Thermal Energy Demand) entwickelt (Koschwitz et al., 2016), welches die Ermittlung des thermischen Energiebedarfs von Gebäuden und urbanen Regionen bei Informationsknappheit ermöglicht. Die implementierten Berechnungsgrundlagen entsprechen jenen der DIN EN 12831. Durch die Verwendung von charakteristischen nutzungs- und baujahrabhängigen Verbrauchswerten auf Grundlage der VDI 3807 lassen sich auch mit minimalem Wissen über das Gebäude Lastgänge erstellen. Auf Grundlage des Hüllflächenverfahrens oder der Verbrauchsverfahren werden im Ergebnis mit SCoTED Lastprofile für Raumwärme und Trinkwarmwasser generiert und ausgegeben. Letztere können wiederum optional in die vorgestellte Influx Datenbank geschrieben werden. Somit bietet die Schnittstelle zu SCoTED die Möglichkeit bei fehlenden Sensor- und Gebäudedaten sowie variierender Informationsdichte den thermischen Energiebedarf von Gebäuden abzuschätzen und als Eingangsgrößen für *AMSA* zu nutzen.

Eine weitere Schnittstelle dient der Integration von Informationen in Bezug auf den Zeitpunkt und die Dauer von zukünftigen Gebäudesanierungen. Mit dem von Spinnräker entwickelten Software-Tool zur Ermittlung optimierter Sanierungsreihenfolgen für Stadtquartiere lassen sich unter anderem gebäudeindividuelle Sanierungszeiträume generieren und exportieren (Spinnräker et al., 2017). Diese sind unmittelbar in *AMSA* importierbar und können zur Abbildung von Lastgangmodifikationen entsprechend der in Kapitel 5 erläuterten Methodik für die szenarienbasierte Quantifizierung von Gleichzeitigkeitseffekten verwendet werden.

Um Datengrundlagen zu Gebäuden und Stadtquartieren grafisch aufzubereiten, wurde ferner eine webbasierte Visualisierungsplattform innerhalb des Energie-Informationssystems entwickelt. Dieses besteht aus zwei Komponenten, wobei *EnergyData-API* als Backend und *EnergyData-Web* als Frontend fungieren (Spinnräker et al., 2018). *EnergyData-API* stellt die Abstraktionsschicht für den Zugriff auf die PostgreSQL und Influx Datenbank dar. *EnergyData-Web* realisiert dagegen die webbasierte Visualisierung der Daten. Um den Aufruf auf mobilen Endgeräten zu ermöglichen, wobei möglichst flüssige Bildbewegungen zu gewährleisten sind, findet der Douglas-Peucker-Algorithmus Anwendung. Letzterer

ermöglicht die Approximation beliebiger Kantenzüge, wodurch sich die Gebäudegeometrien vereinfachen und die Performance der Visualisierungen erhöhen lassen (Spinnräker et al., 2018). Die zentrale Schnittstelle für das Energie-Informationssystem stellt die Datenbankinfrastruktur dar, welche den Informationsaustausch zwischen Software-Tools ermöglicht. Vor diesem Hintergrund wurden in *AMSA* die Voraussetzungen für eine Integration in das Energie-Informationssystem mittels der Schnittstellenklasse *influxInterface* geschaffen. Durch die Modularität des Softwarekonzepts können überdies weitere Schnittstellenklassen zukünftig vorgesehen werden, wodurch auch an dieser Stelle Weiterentwicklungsmöglichkeiten offen stehen.

6.4 Zusammenfassung

Das implementierte autoadaptive prädiktive Modell mit dem Namen *AMSA* basiert auf einer mehrteiligen, hierarchisch aufgebauten Klassenstruktur. Letztere umfasst zwei grafische Benutzeroberflächen zur Nutzerinteraktion, eine Hauptklasse als zentrale Steuerungsinstanz des Programms, eine Grafikklasse für Visualisierungen sowie Modell- und Schnittstellenklassen. Für den Im- und Export von Daten und Prognosemodellen werden eine Influx Datenbank sowie lokale Datencontainer verwendet. Die Integration kurzfristiger Wettervorhersagen in die Informationsgrundlage erfolgt über einen Datenabruf vom Webserver des DWD. Der modulare Aufbau von *AMSA* ermöglicht überdies eine Integration in bestehende Energie-Informationssysteme für die energetische Planung und Abbildung von Quartieren und urbanen Regionen.

KAPITEL 7

Modelldemonstration

Im Rahmen dieses Kapitels erfolgt die Demonstration von *AMSA* am Beispiel eines realen, räumlich abgeschlossenen Quartiers, das sowohl mit einem bestehendem Wärme- als auch Kältenetz versorgt wird. Dadurch kann das entwickelte Modell für den Wärme- und Kältesektor gleichermaßen angewendet und verifiziert werden. Das Kapitel ist in vier Sektionen unterteilt, beginnend bei der Vorstellung der Datengrundlage in Abschnitt 7.1. Abschnitt 7.2 beinhaltet die Demonstration der Modellkomponenten von *AMSA*. Im Rahmen von Abschnitt 7.3 werden die Ergebnisse der Quantifizierung von Gleichzeitigkeitseffekten mit *AMSA* vorgestellt. Dabei dient die konventionelle Herleitung von Kennlinien für den Gleichzeitigkeitsfaktor als Referenz. Abschließend werden in Abschnitt 7.4 die Resultate des vorliegenden Kapitels zusammengefasst.

7.1 Datengrundlage

Das entwickelte Modell wird auf Grundlage von Daten des Forschungszentrums Jülich, welche im Rahmen der Bearbeitung des Projektes *Entwicklung eines Energieversorgungskonzeptes mit modellbasierter prädiktiver Regelung anhand einer Living Roadmap am Beispiel des Forschungszentrums Jülich* des Förderschwerpunktes EnEff:Stadt - Campus-Projekte (durch das BMWi gefördertes sechstes Energieforschungsprogramm FKZ: 03ET1352A) zur Verfügung standen, demonstriert. Im Zuge dieses Projektes sollte ein ganzheitliches Energiekonzept für die betreffende Liegenschaft entwickelt werden, da auslaufende Verträge mit dem regionalen Fernwärmeversorger Weisweiler und strukturelle Veränderungen im Gebäudebestand eine energetische Neuausrichtung bedingen. Die Aufgabenstellung des Projektes steht somit exemplarisch für ein Anwendungsfeld von *AMSA*.

Projektdaten

Das Forschungszentrum Jülich stellt ein räumlich abgegrenztes Quartier mit ungefähr 200 Gebäuden und bestehender Energieinfrastruktur dar. Die thermische Energieversorgung wird mit Fernwärme aus dem nahegelegenen Braunkohlekraftwerk Weisweiler und mittels Kompressionskältemaschinen sicher gestellt. Das Gebäudealter des heterogenen Nichtwohngebäudebestands variiert zwischen 1918 und 2015. Im Zuge geplanter Modernisierungs- und

Neubaumaßnahmen sind jedoch strukturelle Veränderungen in der Zusammensetzung des Quartiers zu erwarten, die zukünftige Energieversorgungsvarianten maßgeblich beeinflussen. Abbildung 7.1 veranschaulicht weiterhin die unterschiedlichen Gebäudetopologien der Liegenschaft basierend auf den projektbezogenen Daten aus einem Geoinformationssystem (GIS).

Abbildung 7.1: GIS-basierte Darstellung der Gebäude des Forschungszentrums Jülich und der Baualtersklassen in Anlehnung an Spinnräker et al., 2018

Für die einzelnen Gebäudekomplexe des Forschungszentrums Jülich stehen detaillierte thermische Lastgänge ab dem Jahr 2013 bis Mitte 2018 zur Verfügung, welche über lokal installierte Wärme und Kältezähler an den Gebäudekomplexen generiert werden. Die Zählersysteme sind an einzelnen Gebäudeteilen verortet, sodass die Zeitreihen für die Betrachtung gesamter Gebäudekomplexe aggregiert werden müssen. Eine detaillierte Auflistung ist Anhang C zu entnehmen. Für letztere wird der Messdatensatz reduziert und es werden nur die Zeitreihen von Anfang 2016 bis Anfang 2018 einbezogen, da zum einen nur für sehr wenige Gebäudekomplexe thermische Leistungsdatensätze ab 2013 vorhanden sind und zum anderen die Zeitreihen von 2018 bis 2019 zum Analysezeitpunkt noch unvollständig vorliegen. Nach Datenaggregation und Ausschluss von Gebäuden, deren Zeitreihen fortlaufend den Wert Null aufweisen, ergibt sich an dieser Stelle die in Tabelle 7.1 aufgeführte Datenzusammenstellung, welche die Gebäudeanzahl in den einzelnen Gebäudenutzungsklassen im Bereich der Wärme- und Kälteversorgung beinhaltet. Ferner liegen auch die Koordinaten der Gebäudegeometriepunkte im Gauß-Krüger-Format vor, die in *AMSA* zur Visualisierung der metrischen Gebäudeverortungen genutzt werden und Teil der Ergebnisdarstellung im Rahmen der Modelldemonstration sind (Abschnitt 7.4).

Zur Verbesserung der Abbildungen wird der betreffende Datensatz um den dazugehörigen Mittelwert korrigiert.

Tabelle 7.1: Anzahl der betrachteten Abnehmer im Wärme- und Kältenetz in Abhängigkeit von Gebäudenutzungsklassen in Anlehnung an Koschwitz et al., 2018

Gebäudenutzung	Wärme (Anzahl der Abnehmer)	Kälte (Anzahl der Abnehmer)
Büro	26	6
Büro & Labor	26	19
Büro & Werkstatt	7	-
Büro & Betrieb	2	1
Büro & Experimentierhalle	1	1
Büro & Lehre	1	-
Labor	15	11
Experimentierhalle	8	7
Lager	5	-
Werkstatt & Lager	1	-
Soziales	4	1
Betrieb	1	1
Betrieb & Experimentierhalle	1	-
	98	47

Insgesamt finden 98 historische Lastgänge der Wärmeversorgung und 47 der Kälteversorgung Eingang in die Vorstudien und Modelldemonstration. Zu Modelltestzwecken werden dennoch Lastgänge deren Wert fortlaufend Null aufweist integriert, um zu überprüfen, ob der Algorithmus zur Datenvorverarbeitung diese identifiziert und aus dem weiteren Analyseprozess ausschließt. Den größten nutzungsspezifischen Gebäudeanteil stellen in der Wärmeversorgung Büro- sowie kombinierte Büro-Labor-Gebäudekomplexe dar. In der Kälteversorgung sind vorwiegend Gebäudekomplexe mit der Mischnutzung Büro und Labor vorhanden. Im Rahmen des Monitorings am Forschungszentrum Jülich werden zudem detaillierte Wetterparameter in Form von zur Verfügung gestellten Zeitreihen erfasst, welche in Tabelle 7.2 im Zusammenhang mit dem Messzeitraum aufgeführt sind. Für das Training des entwickelten Modells ergibt sich aufgrund verschiedener Zeiträume, in welchen Zeitreihen zu den einzelnen Wetterparametern vorliegen, die Problematik einer variierenden Informationsdichte. Um eine einheitliche Datenbasis für die Jahre 2016 und 2017 herzustellen, werden die Variablen *Absoluter Luftdruck* und *Relativer Luftdruck* nicht weiter in die Analyse miteinbezogen.

Tabelle 7.2: Auflistung der projektbezogenen Wetterparameter mit dem dazugehörigen individuellen Messzeitraum

Parameter	**gemessener Zeitraum**
Außenlufttemperatur	ab 01.01.2009
Taupunkttemperatur	ab 24.07.2015
Relative Außenluftfeuchtigkeit	ab 24.07.2015
Absoluter Luftdruck	24.07.2015 - 19.10.2017
Relativer Luftdruck (auf Meereshöhe bezogen)	24.07.2015 - 19.10.2017
Niederschlagsmenge	ab 24.07.2015
Niederschlagsintensität	ab 24.07.2015
Mittlere Windgeschwindigkeit	ab 24.07.2015
Mittlere Windrichtung	ab 24.07.2015

Ergänzend zu den in Tabelle 7.2 aufgelisteten Variablen ist die Globalstrahlung eine wichtige Einflussgröße in Bezug auf den gebäudespezifischen solaren Wärmeeintrag, insbesondere vor dem Hintergrund der Gebäudekühlung in den Sommermonaten. Informationen zur Globalstrahlung liegen für den betrachteten Zeitraum jedoch nicht als historische lokale Monitoringdaten vor, sodass sich die Globalstrahlung nicht als Parameter integrieren lässt.

Ergänzende Daten

Zur Prognose des kurzfristig zu erwartenden thermischen Leistungsbedarfs der Gebäudekomplexe werden Wettervorhersagen der kommenden Stunden und Tage benötigt, welche im Climate Data Center (CDC) des DWD zur Verfügung stehen. Als Eingangsvariablen für die Prognosen fungieren in diesem Fall

I. Außenlufttemperatur

II. Taupunkttemperatur

III. Windgeschwindigkeit

IV. Windrichtung

Die verbleibenden, in Tabelle 7.2 gelisteten Wettervariablen, finden keine Anwendung, da sie nicht im Datensatz für kurzfristige Wettervorhersagen des DWD enthalten sind. Für die Generierung unterschiedlicher zukünftiger Wetterszenarien mit längerfristigem Zeithorizont dienen die Testreferenzjahre des Deutschen Wetterdienstes, welche auf detaillierten Mess- sowie Beobachtungsdaten beruhen und typische ortsbezogene Wettercharakteristika repräsentieren. Dabei wird zwischen dem mittleren Testreferenzjahr und den extremen Winter- und Sommer-Testreferenzjahren unterschieden. Die Datensätze beinhalten verschiedene Wetterparameter, wie die Lufttemperatur, Windgeschwindigkeit oder relative Außenluftfeuchtigkeit. Abbildung 7.2a veranschaulicht die geordneten Jahresdauerlinien für die Außenlufttemperatur des Testreferenzjahr 2015 (Deutscher Wetterdienst et al., 2017).

Der Verlauf der Jahresdauerlinien zeigt den Unterschied zwischen dem mittleren Testreferenzjahr mit einer geringeren Temperaturamplitude und den extremen Testreferenzjahren mit vergleichsweise höheren Temperaturspektren. Für die datenbasierte Abbildung zukünftiger Klimaentwicklungen stehen ferner Datensätze für den Zeitraum von 2031 bis 2060, basierend auf Abschätzungen der Auswirkungen des Klimawandels auf die Wetterverhältnisse, zur Verfügung. Diese werden mittels der Testreferenzjahre 2045 repräsentiert (Deutscher Wetterdienst et al., 2017). Abbildung 7.2b visualisiert diese. Abbildung 7.2 zeigt darüber hinaus den Unterschied zwischen der Lufttemperatur des gegenwärtigen im Vergleich zu dem zukünftigen Testreferenzjahr im Trend. Die Lufttemperatur des zukünftigen Testreferenzjahres liegt dabei ganzjährig über der des gegenwärtigen Testreferenzjahres, sodass sich mit diesen Datensätzen der Einfluss zukünftig zu erwartender steigender Lufttemperaturen abbilden lässt. Als weitere Variablen zur Generierung zukünftiger Wetterszenarien dienen

I. Außenlufttemperatur
II. Relative Außenluftfeuchtigkeit
III. Windgeschwindigkeit
IV. Windrichtung

Dagegen werden die Taupunkttemperatur sowie Niederschlagsmenge und -intensität nicht in das Modelltraining für die mittel- und langfristigen Wetterszenarien miteinbezogen, da sie nicht in den TRJ vorhanden sind. Ergänzend zu den Wettervariablen werden Anwesenheitsprofile der Gebäudenutzer erstellt, welche auf der Annahme basieren, dass die Arbeitszeit im Mittel zwischen 8 und 18 Uhr liegt und am Wochenende nicht gearbeitet wird. Neben der Arbeitszeit erfolgt auch die Berücksichtigung der Ferienzeit mittels eines weiteren Anwesenheitsprofils. Die Variablen sind als Boolean kodiert, wodurch beiden Variablen der Wert 1 zugewiesen wird, sofern ein Werktag bzw. kein Ferientag vorliegt. Dementsprechend nehmen die Variablen im umgekehrten Fall den Wert 0 an.

Zur Evaluierung des Einflusses einzelner Variablen auf die Heiz- und Kühlleistungen lassen sich lineare Korrelationsanalysen anwenden. Die theoretischen Grundlagen sind Anhang A.1, Abschnitt A.2 zu entnehmen. Im Rahmen einer Studie wurden diese Einflüsse auf Basis eines nahezu identischen Datensatzes im Vergleich zu jenem dieser Arbeit untersucht (Koschwitz et al., 2018). Im Ergebnis zeigt sich, dass nicht alle Variablen einen großen Einfluss haben. Dennoch wurde eine Vorteilhaftigkeit der Integration für maschinelle Lernverfahren nachgewiesen (Guyon et al., 2003). Zur Abschätzung des Wärmebedarfs von Gebäuden nach Sanierung dienen die spezifischen Wärmeverbrauchswerte in Abhängigkeit von Gebäudenutzungsklassen nach VDI 3807 (VDI 3807 Blatt 2, 2014). In Tabelle 7.3 sind die entsprechenden Werte zu Nutzungsklassen der Gebäudekomplexe aufgeführt. Um die Referenzwerte für den Wärmeverbrauch von Gebäudekomplexen mit multipler Nutzung nach Sanierung auf Basis der vorgestellten Methodik ermitteln zu können, bedarf es einer anteiligen Verrechnung der Referenzwerte einzelner Gebäudeteile.

Tabelle 7.3: Spezifische Wärmeverbrauchskennwerte in Abhängigkeit von Gebäudenutzungsklassen nach VDI 3807 in Anlehnung an Koschwitz et al., 2019

Gebäudenutzung	Spezifischer Wärmeverbrauch $[kWh/(m^2 \cdot a)]$
Büro	69
Werkstatt	59
Bibliothek	56
Lager	58
Betrieb	70
Laborinstitut (Kategorie 2)	85
Laborinstitut (Kategorie 3)	81
Labor für Lehre und Forschung	63
Labor für Forschung und Versuche	136
Kantine	138
Garage	57
Experimentierhalle	136
Soziales	101

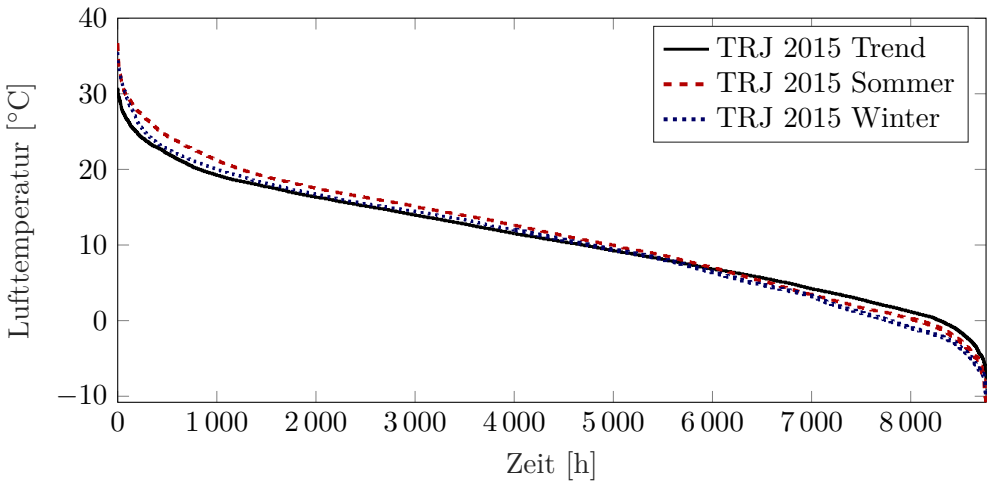
Beginnend mit der Informationsausgangssituation gemäß Tabelle 7.4 wird anhand eines Beispielgebäudes das Vorgehen nachfolgend näher erläutert.

Tabelle 7.4: Eingangsparameter für die exemplarische Berechnung des Referenzwertes für den Wärmeverbrauch eines Gebäudekomplexes mit multipler Nutzung nach Sanierung in Anlehnung an Koschwitz et al., 2019

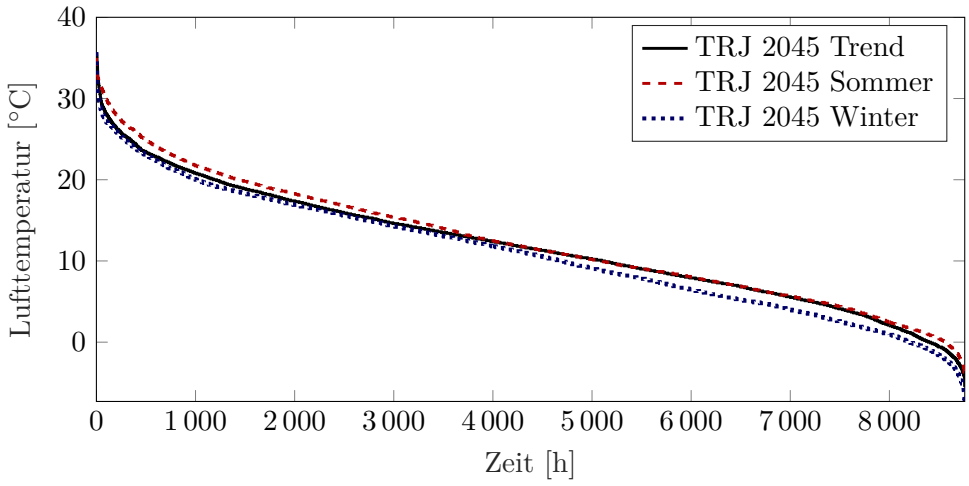
	Nettogrundfläche $\left[m^2\right]$	Verbrauchswert $\left[\frac{kWh}{m^2 \cdot a}\right]$	Referenzwert $\left[\frac{kWh}{m^2 \cdot a}\right]$
Beispielgebäude	9333	106	?
Laborteil 1	2334		81
Laborteil 2	2291		81
Experimentierhallenteil	1005		109
Büroteil	3703		69

Der Wärmeverbrauchskennwert q_{Ex} für das Beispielgebäude ergibt sich dann wie folgt (Koschwitz et al., 2019):

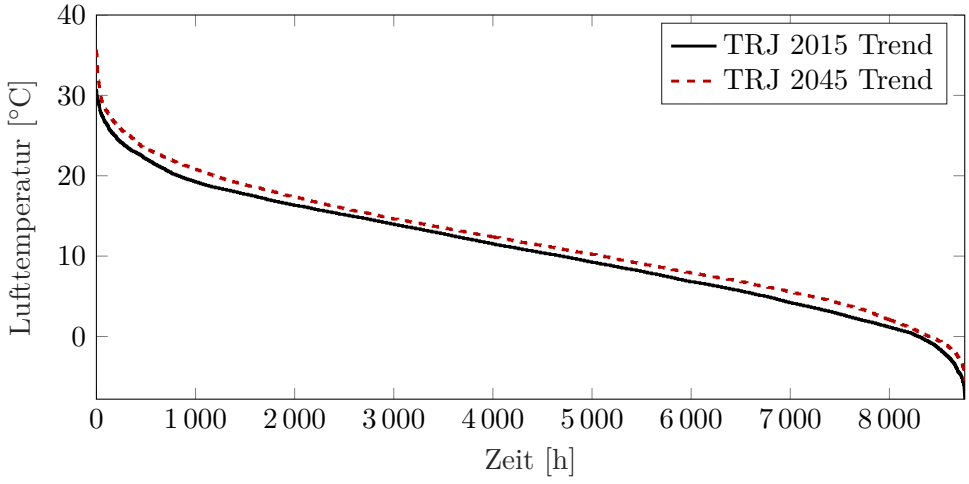
$$\begin{aligned} q_{Ex} &= \frac{(2334 + 2291) \cdot 81 + 1005 \cdot 109 + 3703 \cdot 69}{9333} \\ &\approx 79 \left[\frac{kWh}{m^2 \cdot a}\right] \end{aligned} \tag{7.1}$$



(a) TRJ 2015



(b) TRJ 2045



(c) Vergleich TRJ 2015 und 2045

Abbildung 7.2: Darstellung der Jahresdauerlinien zu den TRJ hinsichtlich der Außenlufttemperatur

7.2 Demonstration der Modellkomponenten

Im Folgenden werden die entwickelten Modellkomponenten auf Basis der zuvor vorgestellten Datengrundlage in verschiedenen Anwendungsfällen getestet und analysiert. Dadurch wird die Eignung von *AMSA* für die Lösung der in dieser Arbeit formulierten Problemstellung nachgewiesen und die Grundlage für die Quantifizierung von Gleichzeitigkeitseffekten demonstriert.

7.2.1 Ergebnisse der Datenvorverarbeitung

Dieser Unterabschnitt beinhaltet die Demonstration der in *AMSA* implementierten Modellkomponente zur Datenvorverarbeitung. Diese umfasst die Bereinigung der Datensätze um Anomalien in den Zeitreihen des historischen thermischen Energiebedarfs, der Klimavariablen sowie die Datennormalisierung.

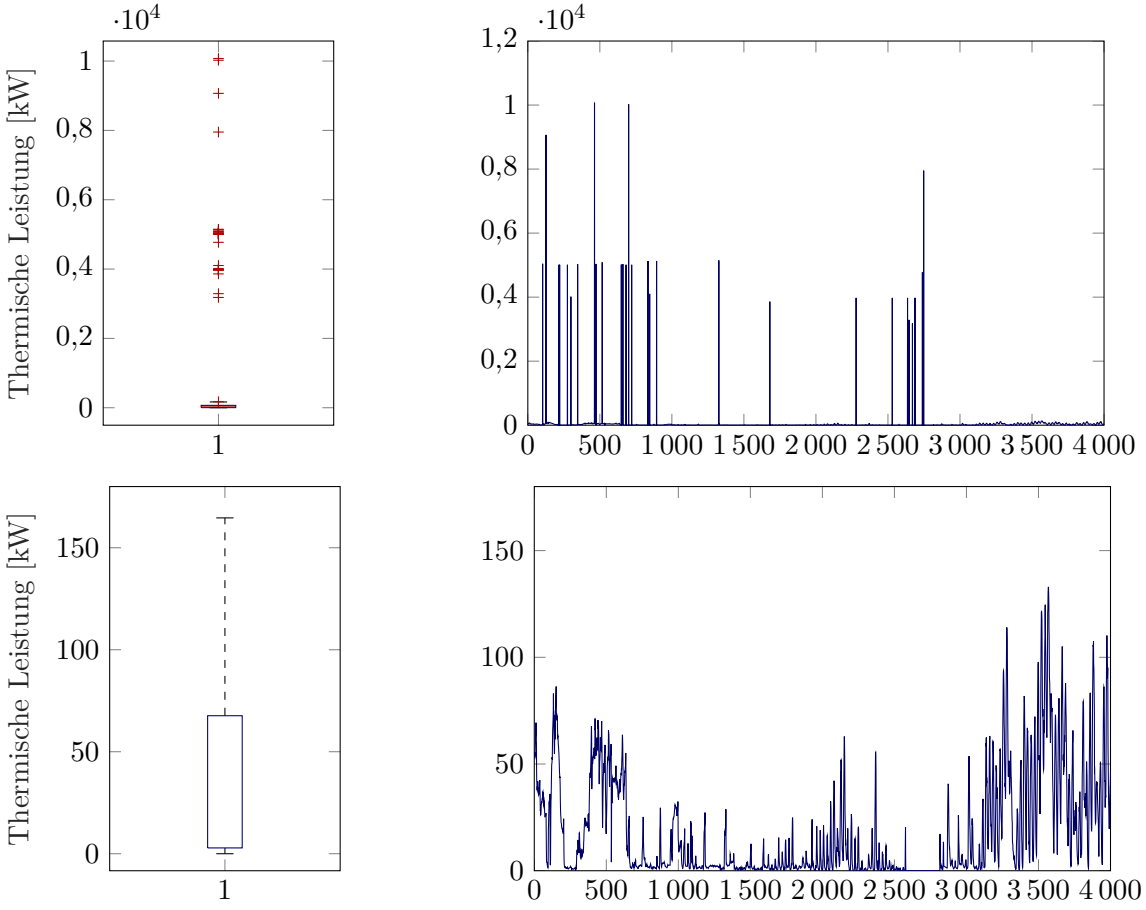
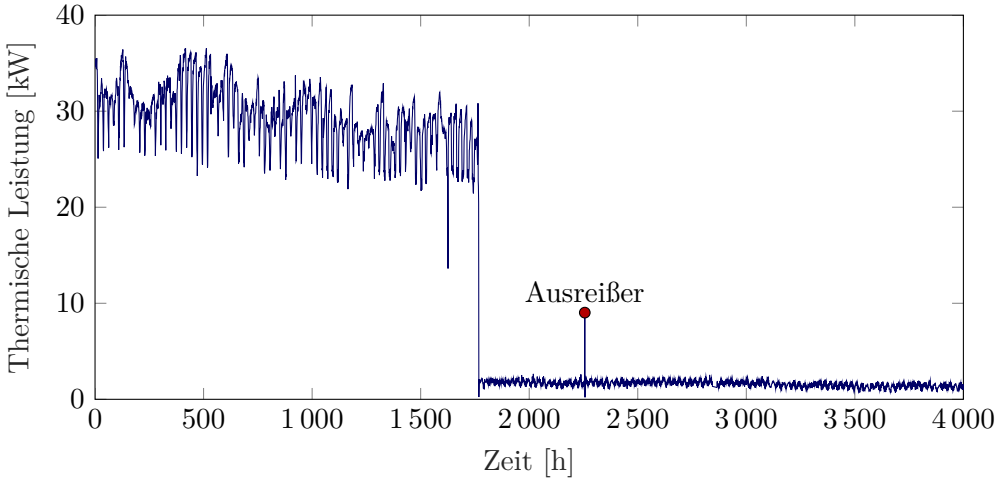
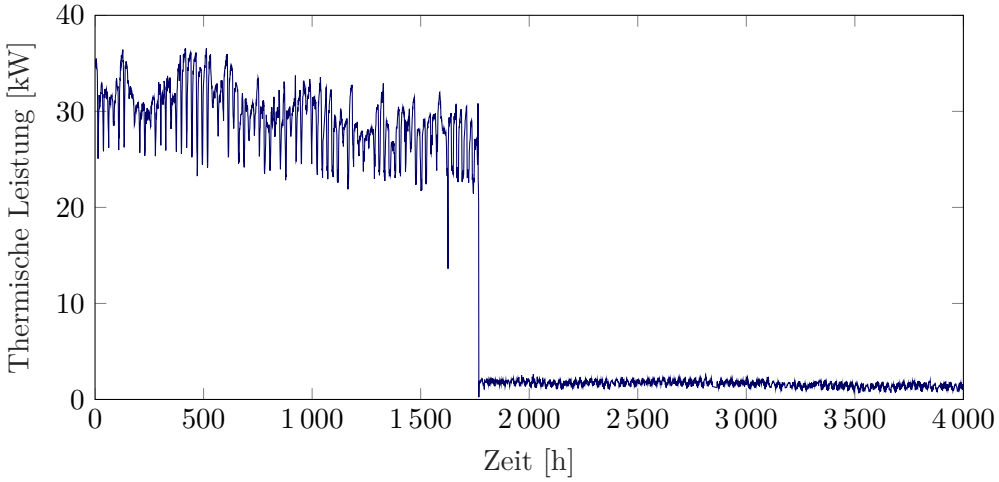


Abbildung 7.3: Vergleich zwischen Datensätzen zur thermischen Leistung eines Gebäudes inklusive und exklusive der Eliminierung von globalen Ausreißern



(a) Zeitreihe mit lokalem Ausreißer



(b) Um den lokalen Ausreißer bereinigte Zeitreihe

Abbildung 7.4: Vergleich zwischen Datensätzen zur thermischen Leistung eines Gebäudes inklusive und exklusive der Eliminierung von lokalen Ausreißern

Abbildung 7.3 zeigt exemplarisch das Ergebnis der Ausreißereliminierung in einer historischen Zeitreihe der thermischen Leistungsabnahme eines Gebäudes im Rahmen der Datenvorverarbeitung in *AMSA*. Im linken Teil der Abbildung sind die Boxplots des unbereinigten sowie des bereinigten Datensatzes für ein Gebäude und im rechten Teil sind die Auswirkungen auf die jeweiligen Zeitreihen dargestellt. Abbildung 7.4 visualisiert überdies die Eliminierung eines lokalen Ausreißers in einer Zeitreihe mit der Methode des gleitenden Medians, welcher unter Anwendung der Boxplot-Methode nicht identifiziert wird. Die Zeitreihen finden, wie in Kapitel 5 erläutert, in normierter Form Eingang in die Prognose- und Clustering-Modelle. Für die Lastgangprognosen werden die Zeitreihen der Wetterparameter und jene der Anwesenheitsprofile auf einen Mittelwert von 0 und eine

Standardabweichung von 1 normiert. In Vorbereitung auf die Ähnlichkeitsanalysen mittels Clustering und Klassifikation, sind in Analogie dazu auch die thermischen Lastprofile zu normieren. Abbildung 7.5 veranschaulicht exemplarisch das Ergebnis der Lastgangnormierung in *AMSA*.

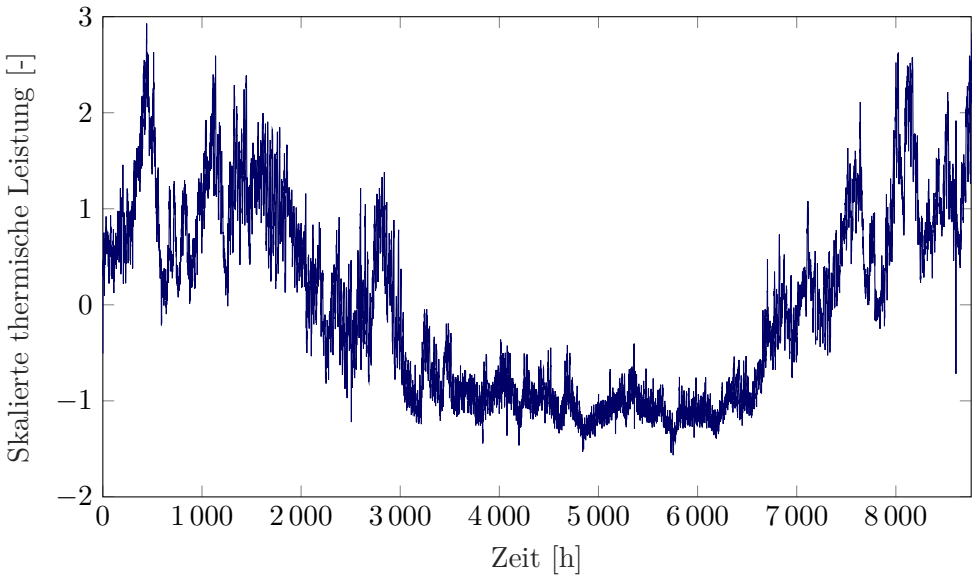


Abbildung 7.5: Beispiel für ein auf einen Mittelwert von 0 normiertes und auf eine Standardabweichung von 1 skaliertes Lastprofil eines Gebäudes

7.2.2 Ergebnisse zu thermischen Lastprognosen

Im Rahmen dieses Abschnitts werden die Ergebnisse des Prognose-Teilmodells vorgestellt. In diesem Zusammenhang erfolgt die Unterscheidung zwischen kurz- sowie mittel- und langfristigen Prognosen. Erstere basieren auf den Wettervorhersagen des DWD für die folgenden 200 Stunden ab dem aktuellen Zeitpunkt und letztere beruhen auf dem aktuellen und dem zukünftigen TRJ sowie den aus der Interpolation generierten Wetterzeitreihen zwischen beiden TRJ entsprechend der Methodik aus Kapitel 5. Zudem dienen für beide Varianten die erwarteten Anwesenheitsprofile in Abhängigkeit von Arbeits-, Wochenend-, und Ferienzeiten als ein Eingangsparameter. Infolge der jährlichen Fluktuationen der Wärmeleistung ergibt sich der Bedarf an adaptiven Strategien der Betriebsführung vorhandener Wärmeerzeuger, um beispielsweise Teillastbetriebszustände zu vermeiden.

Mit dieser Arbeit wird einerseits eine Methodik zur Generierung quantifizierbarer langfristiger Wärmelastszenarien auf Einzelgebäude- und Quartiersebene unter Berücksichtigung verschiedener Einflussfaktoren, wie zukünftiger lokaler Klimaveränderungen und Modifikationen im Gebäudebereich infolge Sanierungsmaßnahmen, aufgezeigt. Andererseits implizieren die Ergebnisse, dass in der prädiktiv ausgerichteten Planung zukünftiger Energiesysteme

temporäre Lastverschiebungen und Variationen in der Höhe des Spitzenleistungsbedarfs zu berücksichtigen sind. Vor diesem Hintergrund werden nachfolgend die Ergebnisse von Analysen der Anwendung der autoadaptiven prädiktiven Modellkomponenten vorgestellt.

Vergleichsstudie zu den Modellen des Prognose-Ensembles

Dieser Unterabschnitt beruht auf der Zusammenfassung von Ergebnissen aus der Studie *Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale* von Koschwitz et al. zum Vergleich der in Kapitel 5 vorgestellten Prognosemodelle (Koschwitz et al., 2018). Im Rahmen dieser Arbeit wurde die Performanz der NARX RNN und ε-SVM-R Modelle für die Prognose von saisonalem gebäudespezifischem Wärme- und Kühlleistungsbedarf auf Quartiersebene untersucht. Als Güteindikatoren dienten dazu die Fehlermaße MSE gemäß Kapitel 5 und Mean Average Error (MAE) gemäß Anhang A, Abschnitt A.1 sowie die modellindividuelle Trainingsdauer. Um in Hinblick auf erstere zusammenfassende Aussagen auf Gebäudegruppenebene treffen zu können, wurde quartalsweise der Median der Fehlermaße zu den gebäudespezifischen Lastprognosen gebildet. Diesbezüglich sind in Tabelle 7.5 die modellspezifischen MAE und MSE Medianwerte für die Vorhersage der gebäudespezifisch benötigten Wärmeleistung aufgeführt.

Auf Grundlage dieser lässt sich mit den NARX RNN eine höhere Prognosegenauigkeit in Bezug auf den gebäudespezifischen Wärmeleistungsbedarf auf Quartiersebene ableiten. Im quartalsübergreifenden Vergleich der MAE Medianwerte bezüglich der NARX RNN Konfigurationen zeigt sich das NARX RNN mit einer verdeckten Schicht jenem mit zwei verdeckten Schichten als überlegen. Unter Hinzuziehung der MSE Medianwerte lässt sich dieses Ergebnis für das Frühlings- und Sommerquartal bestätigen. Im Gegensatz dazu weisen die geringeren MSE Medianwerte für Winter und Herbst auf eine Überlegenheit der NARX RNN mit zwei verdeckten Schichten hin. Zusammenfassend lässt sich jedoch auf Grundlage der Ergebnisse unter Hinzuziehung beider Fehlermaße eine vergleichbare Performanz beider NARX RNN feststellen. Hinsichtlich der ε-SVM-R Modelle ist hervorzuheben, dass mit dem ε-SVM-R Modell auf Basis eines polynomialem Kernel dritter Ordnung in Bezug auf die Quartale Winter und Frühling im Vergleich zu dem Modell mit radialer Basisfunktion geringere MAE und MSE Medianwerte erzielt werden konnten. Ein gegenteiliges Ergebnis bleibt für das Herbstquartal festzuhalten. Hinsichtlich der auf das Sommerquartal bezogenen Prognosen des Wärmeleistungsbedarfs ist keine klare Tendenz zur Vorteilhaftigkeit eines ε-SVM-R Modells auf Grundlage der MAE und MSE erkennbar. Die unterschiedliche Performanz der Prognosemodelle für die saisonale Wärmelastvorhersage in dieser Studie ist exemplarisch in Abbildung E.1 des Anhangs E dargestellt. In Analogie zu Tabelle 7.5 sind in Tabelle 7.6 die modellspezifischen MSE und MAE Medianwerte für die Prognose der gebäudespezifisch saisonal benötigten Kühlleistung aufgeführt.

Auf Grundlage der MAE Medianwerte lässt sich in Bezug auf die Kühllastprognosen eine Überlegenheit des NARX RNN mit einer verdeckten Schicht im Vergleich zu der Konfi-

Tabelle 7.5: Auflistung der modellspezifischen MSE und MAE Medianwerte zu Prognosen des saisonalen gebäudeindividuellen Wärmeleistungsbedarfs in Anlehnung an Koschwitz et al., 2018

Jahreszeit	**NARX RNN 1 verdeckte Schicht**	**NARX RNN 2 verdeckte Schichten**
Winter	MAE: 3,82	MAE: 4,09
	MSE: 52,45	MSE: 49,44
Frühling	MAE: 3,73	MAE: 3,79
	MSE: 50,33	MSE: 52,22
Sommer	MAE: 3,92	MAE: 4,12
	MSE: 49,00	MSE: 50,10
Herbst	MAE: 3,89	MAE: 4,01
	MSE: 52,20	MSE: 51,52
	ε-SVM-R RBF Kernel	**ε-SVM-R polynomialer Kernel**
Winter	MAE: 16,38	MAE: 13,03
	MSE: 419,87	MSE: 263,45
Frühling	MAE: 9,46	MAE: 8,86
	MSE: 159,68	MSE: 151,86
Sommer	MAE: 7,35	MAE: 9,33
	MSE: 158,90	MSE: 120,00
Herbst	MAE: 9,66	MAE: 10,30
	MSE: 163,55	MSE: 181,58

guration mit zwei verdeckten Schichten feststellen. Dieses Ergebnis wird durch die MSE Medianwerte mit Ausnahme des Sommerquartals bestätigt. In Analogie zu den Wärmelastprognosen ist die Performanz der beiden NARX RNN hinsichtlich der betrachteten Fehlermaße auf einem ähnlichen Niveau einzuordnen. Im Hinblick auf die ε-SVM-R Modelle weisen geringere MAE und MSE Medianwerte bezüglich der Kühllastprognosen für die Quartale Winter und Frühling auf eine Dominanz des Modells mit polynomialem Kernel im Vergleich zu jenem mit radialer Basisfunktion hin. Demgegenüber lässt sich für das Sommer- und Herbstquartal an dieser Stelle keine eindeutige Tendenz diesbezüglich identifizieren. Abbildung E.2 zeigt daran anknüpfend beispielhaft einen 40-Stunden-Auszug der mit den unterschiedlichen Modellen prognostizierten Kühllast im Juli. Bezugnehmend auf die MAE Werte zeigen beide NARX RNN Konfigurationen vergleichbare Ergebnisse, wobei das NARX RNN mit einer verdeckten Schicht im Durchschnitt graduell genauere Ergebnisse liefert. Im Hinblick auf die MSE Werte ist festzuhalten, dass in einigen Fällen mit dem NARX RNN auf Grundlage von zwei verdeckten Schichten akuratere Prognosen erzielt werden konnten. Überdies bleiben die Prognosefehler über die verschiedenen Jahreszeiten hinweg konstant. Dagegen lässt sich eine jahreszeitlich inkonsistente Performanz der ε-SVM-R Modelle feststellen. So zeigte sich bei den Wärmelastprognosen mit dem ε-SVM-R Modell

Tabelle 7.6: Auflistung der modellspezifischen MSE und MAE Medianwerte zu Prognosen des saisonalen gebäudeindividuellen Kühlleistungsbedarfs in Anlehnung an Koschwitz et al., 2018

Jahreszeit	**NARX RNN 1 verdeckte Schicht**	**NARX RNN 2 verdeckte Schichten**
Winter	MAE: 2,44 MSE: 27,46	MAE: 2,57 MSE: 28,54
Frühling	MAE: 2,39 MSE: 26,95	MAE: 2,41 MSE: 33,20
Sommer	MAE: 2,39 MSE: 41,72	MAE: 2,54 MSE: 30,28
Herbst	MAE: 2,35 MSE: 30,16	MAE: 2,85 MSE: 40,40
	ε-SVM-R RBF Kernel	**ε-SVM-R polynomialer Kernel**
Winter	MAE: 7,77 MSE: 80,70	MAE: 6,60 MSE: 71,80
Frühling	MAE: 7,97 MSE: 90,55	MAE: 5,82 MSE: 69,57
Sommer	MAE: 14,66 MSE: 246,23	MAE: 13,94 MSE: 259,96
Herbst	MAE: 9,12 MSE: 140,90	MAE: 9,37 MSE: 131,98

in Kombination mit RBF im Vergleich zu dem Modell mit einem polynomialen Kernel eine geringere Güte der Ergebnisse für die Lastvorhersagen im Sommer und Herbst. Ein Grund hierfür könnte in dem real variierenden Nutzerverhalten aufgrund des individuellen Komforts und der damit einhergehenden Lüftungsgewohnheiten begründet liegen, wobei das modellierte Nutzerverhalten aufgrund der fehlenden Informationen durch die Features nicht genügend repräsentiert ist. Diesem Erklärungsansatz folgend, kann die bessere Performanz der NARX RNN mit den rekursiven Eigenschaften des Netzwerks infolge der Integration des realen Nutzerverhaltens in Form der historischen Messwerte begründet werden (Koschwitz et al., 2018). Im Hinblick auf die Ergebnisse der Kühllastprognosen kann dieser Effekt in umgekehrter Analogie zum Heizlastfall im Winter festgestellt werden. Dies begründet sich durch die im jahreszeitlichen Vergleich höchsten Fehlerwerte der Sommerlastprognosen. Weiterhin implizieren die Ergebnisse für die Kühllasten eine Vorteilhaftigkeit des ε-SVM-R Modells auf Basis des polynomialen Kernels in den Jahreszeiten Winter, Frühling und Sommer. Die Lastprognosen für den Herbst konnten mit beiden ε-SVM-R Modellen auf ähnlichem Genauigkeitsniveau erzielt werden. Um eine differenzierte Analyse vornehmen zu können, dienen die MAE Werte für jede prognostizierte Zeitreihe als Grundlage für die weitere Analyse. Für jede Prognoseaufgabe wurden die MAE Werte für die vier Modelle

berechnet und auf Grundlage dessen für jede Prognoseaufgabe der minimale MAE Wert bestimmt. In der Folge lässt sich das Modell mit der besten Performanz für die Prognoseaufgabe identifizieren. Dieses Vorgehen wurde für jede Prognoseaufgabe wiederholt, sodass die Anzahl der besten Performanzwerte für jedes Modell bestimmt werden konnte. In den Tabellen 7.7 und 7.8 ist dahingehend die Anzahl an Minimum MAE Werten im Rahmen des Modellvergleichs aufgelistet.

Tabelle 7.7: Anzahl der minimalen MAE Werte saisonaler Wärmelastprognosen in Anlehnung an Koschwitz et al., 2018

Jahreszeit	**NARX RNN 1 verdeckte Schicht**	**NARX RNN 2 verdeckte Schichten**
Winter	54	41
Frühling	63	31
Sommer	56	31
Herbst	54	36
	ε-SVM-R RBF Kernel	**ε-SVM-R polynomialer Kernel**
Winter	0	2
Frühling	2	1
Sommer	6	4
Herbst	3	4

Tabelle 7.8: Anzahl der minimalen MAE-Werte saisonaler Kühllastprognosen in Anlehnung an Koschwitz et al., 2018

Jahreszeit	**NARX RNN 1 verdeckte Schicht**	**NARX RNN 2 verdeckte Schichten**
Winter	19	13
Frühling	25	10
Sommer	24	11
Herbst	22	15
	ε-SVM-R RBF Kernel	**ε-SVM-R polynomialer Kernel**
Winter	12	5
Frühling	5	9
Sommer	5	9
Herbst	4	8

Die Ergebnisse zeigen, dass die NARX RNN den ε-SVM-R Modellen in den meisten Fällen

im Bereich der Wärmelastprognosen überlegen sind. Demgegenüber kann diese Tendenz für die Kühllastprognosen nicht in einem vergleichbaren Umfang festgestellt werden. Dennoch zeigen sich die NARX RNN im Hinblick auf die Vorhersagegenauigkeit den ε-SVM-R in den meisten Fällen überlegen, wobei einige Lastprognosen mit den ε-SVM-R Modellen eine vergleichbare Genauigkeit der Vorhersage ermöglicht haben. Näheres dazu veranschaulicht Abbildung E.1. Ein Grund für das niedrigere Performanzniveau der RNN im Bereich der Kühllastprognosen liegt in dem witterungsunabhängigen prozessbedingten Kühllastanteil und dem speziellen Nutzerverhalten in Bezug auf das An- und Abschalten von Kälteanlagen. In diesem Zusammenhang visualisiert Abbildung 7.6 beispielhaft die Charakteristik eines Kühllastprofils.

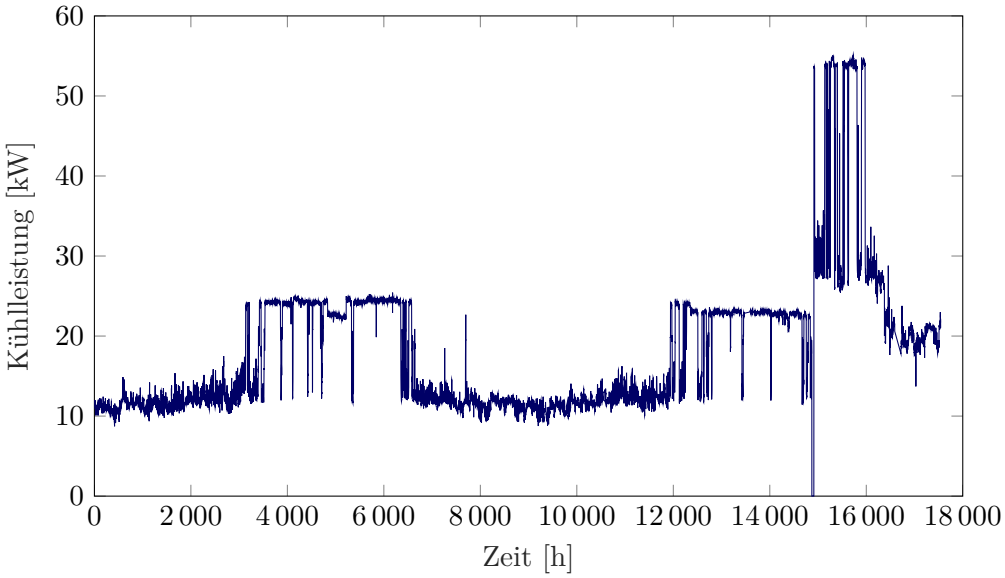


Abbildung 7.6: Exemplarische Darstellung eines gebäudebezogenen Kühlleistungsprofils in Anlehnung an Koschwitz et al., 2018

Die Eigenschaft der NARX RNN, die Ausgangsgrößen wieder als Eingangsparameter zurückzugeben, führt bei einer Nutzerintervention zum aktuellen Zeitschritt zu einem hohen Fehler in der Lastprognose des folgenden Zeitschritts. In diesem Fall tendiert das Netzwerk zum Overfitting. Die Wahl eines Modells wird nicht zuletzt auch durch den Rechenaufwand vor dem Hintergrund der zur Verfügung stehenden Hardware beeinflusst. Daher wurden die Rechenzeiten je gebäudebezogener quartalsweiser Lastprognosen untersucht. Im Durchschnitt benötigt das ε-SVM-R Modell auf Basis des RBF Kernels 7 Sekunden für die Trainings- und Testphase, wohingegen sich die Rechenzeit des Modells mit einem polynomialen Kernel dritten Grades durchschnittlich auf 358 Sekunden erhöht hat. Im Vergleich dazu benötigte ein NARX RNN mit einer verdeckten Schicht in dieser Fallstudie im Mittel 13 Sekunden und mit zwei verdeckten Schichten durchschnittlich 57 Sekunden für jede gebäudebezogene Lastprognose.

Einfluss variierender Datengrundlagen auf die Modellauswahl

Die folgenden Untersuchungen dienen der Analyse von Auswirkungen variierender Datengrundlagen auf die gebäudespezifische Modellauswahl innerhalb des Prognose-Ensembles. Beginnend mit der Auswertung der Trainingsergebnisse auf Basis der Datengrundlage für kurzfristige Lastprognosen, sind in Tabelle 7.9 die prozentualen Nutzungen der einzelnen Prognosemodelle innerhalb des Ensembles aufgeführt. Hinsichtlich der Ergebnisse ist festzustellen, dass sowohl für die gebäudebezogenen Wärme- als auch Kühllastprognosen ausschließlich die NARX RNN Anwendung finden. Ferner sind in Tabelle 7.10 in Analogie die Ergebnisse des Modelltrainings für mittel- und langfristige Prognosen von Wärme- und Kühllasten aufgelistet.

Tabelle 7.9: Prozentuale Nutzung der einzelnen Prognosemodelle innerhalb des Ensembles auf Basis der Datengrundlage für kurzfristige Lastvorhersagen

Modell	**Wärme**	**Kälte**
NARX RNN 1 verdeckte Schicht	15	34
NARX RNN 2 verdeckte Schichten	85	66
ε-SVM-R RBF Kernel	0	0
ε-SVM-R polynomialer Kernel	0	0

Tabelle 7.10: Prozentuale Nutzung der einzelnen Prognosemodelle innerhalb des Ensembles auf Basis der Datengrundlage für mittel- und langfristige thermische Lastvorhersagen

Modell	**Wärme**	**Kälte**
NARX RNN 1 verdeckte Schicht	12	36
NARX RNN 2 verdeckte Schichten	88	64
ε-SVM-R RBF Kernel	0	0
ε-SVM-R polynomialer Kernel	0	0

Auch in diesem Fall können sowohl bezüglich der Wärmelast- als auch hinsichtlich der Kühllastprognosen die höchste Genauigkeit mit den NARX RNN erreicht werden. Beide ε-SVM-R Modelle finden in diesem Fall keine Anwendung für Prognoseaufgaben. Dieses Ergebnis bestätigt zum einen tendenziell jene der Vergleichsstudie zu den vier verschiedenen Prognosemodellen auf Grundlage von Koschwitz et al., 2018 hinsichtlich der Vorteile von NARX RNN im Vergleich zu ε-SVM-R Modellen. Gleichzeitig besteht eine Diskrepanz in den Resultaten, da in der Vorstudie in Einzelfällen auch die ε-SVM-R Modelle genauere Prognoseergebnisse lieferten. Eine Ursache für diesen Effekt liegt einerseits in dem unterschiedlichen Umfang des Trainingsdatensatzes und andererseits in der unterschiedlichen Beschaffenheit des Trainingsdatensatzes begründet, da in der Vorstudie ausschließlich die Methode des gleitenden Medians für die Eliminierung von Datenausreißern verwendet

wurde. Demgegenüber ist in dem Modell dieser Arbeit zusätzlich die Boxplot-Methode integriert, wodurch sich der Trainingsdatensatz von jenem aus der Vorstudie unterscheidet. Zur Demonstration dieses Einflusses auf die Ergebnisse des Modelltrainings dienen folgende drei Varianten der Datenvorverarbeitung am Beispiel des Kühllastfalls.

Variante I: Eliminierung von Ausreißern mit der Methode des gleitenden Medians

Variante II: Eliminierung von Ausreißern mit der Boxplot-Methode

Variante III: keine Eliminierung von Ausreißern

In Tabelle 7.11 sind die Ergebnisse des Variantenvergleichs aufgeführt.

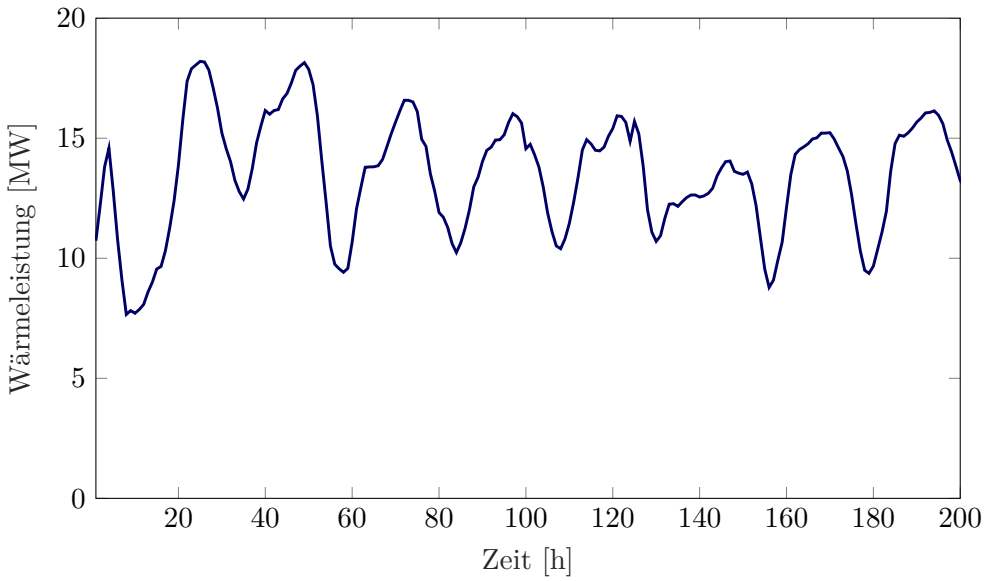
Tabelle 7.11: Prozentuale Nutzung der Prognosemodelle auf Grundlage eines variierenden Grades der Eliminierung von Ausreißern im Trainingsdatensatz

Modell	**Variante I**	**Variante II**	**Variante III**
NARX RNN 1 Hidden Layer	32	17	30
NARX RNN 2 Hidden Layer	60	83	42
ε-SVM-R RBF Kernel	4	0	15
ε-SVM-R polynomialer Kernel	4	0	13

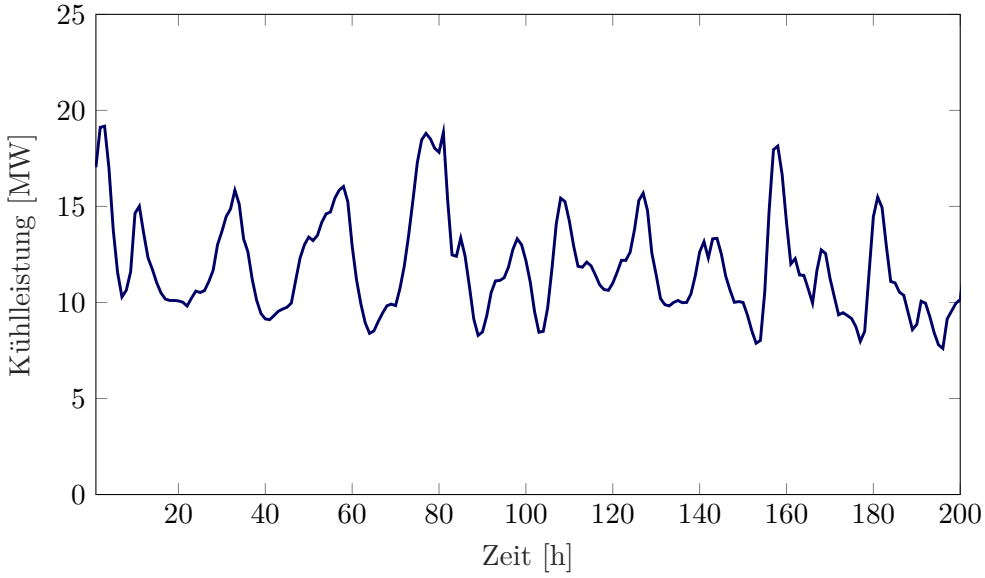
Dabei ist festzustellen, dass die ausschließliche Integration der Methode des gleitenden Medians in die Datenvorverarbeitung zu einer tendenziell erhöhten Anzahl an überlegenen ε-SVM-R Modellen führt. Dagegen bestätigt Variante II in Grundzügen das ursprüngliche Ergebnis nach Tabelle 7.10. An dieser Stelle zeigt sich in Bezug auf Variante II eine verstärkte Überlegenheit der NARX RNN mit zwei verdeckten Schichten. Demgegenüber ist aus den Ergebnissen zu Variante I eine ähnliche, aber nicht in dem Maße ausgeprägte Vorteilhaftigkeit der NARX RNN mit zwei verdeckten Schichten zu erkennen. Bezüglich des Ergebnis zu Variante III lässt sich feststellen, dass die Überlegenheit der ε-SVM-R-Modelle bzw. die Unterlegenheit der NARX RNN von der Entscheidung abhängt, ob und in welchem Umfang eine Eliminierung von Ausreißern in der Datenvorverarbeitung vorgenommen wird. Diese Schlussfolgerung deckt sich mit Erkenntnissen aus anderen Anwendungsgebieten, nach denen SVM- Modelle weniger anfällig für Übertrainingseffekte sind (Han et al., 2012; Koschwitz et al., 2018). Weiterhin lassen sich die Ergebnisse mit den unterschiedlichen Ansätzen der Definition von Basisfunktionen für KNN und SVM begründen. Während KNN als universelle Approximatoren auf adaptiven Basisfunktionen basieren, werden für SVM Kernel festgelegt (Bishop, 2006), welche für einzelne Lastgänge auf urbaner Ebene nicht passend sein können. Zusammenfassend ist an dieser Stelle festzustellen, dass die Entwicklung des autoadaptiven Prognose-Ensembles vorteilhaft gegenüber der pauschalen Verwendung eines Modells für die automatisierte Prognose thermischer Lasten auf urbaner Ebene ist. Dieses Ergebnis verstärkt sich in der Aussagekraft gerade bei Veränderungen der Datengrundlage infolge von Neu- und Rückbauten sowie Gebäudesanierungen.

Kurzfristige Lastprognosen

Auf Basis der trainierten gebäudespezifischen Modelle lassen sich kurzfristige thermische Leistungsanforderungen für einzelne Gebäude und aggregiert für Gebäudegruppen und urbane Regionen mit *AMSA* prädiktiv abschätzen und für die Quantifizierung von Gleichzeitigkeitseffekten nutzen. Abbildung 7.7 visualisiert die prognostizierte Wärme- und Kühllast des Demonstrationsquartiers für die folgenden 200 Stunden ab einem festgelegten Zeitpunkt τ_0 an. Hinsichtlich der Lastgangcharakteristika zeigen sich an dieser Stelle Unterschiede zwischen benötigter Wärme- und Kühlleistung. Während sich der funktionale Verlauf der prognostizierten Wärmeleistung für das Quartier als sinusartig mit regelmäßig in ähnlichen Abständen wiederkehrenden Leistungsspitzen beschreiben lässt, ist selbiger demgegenüber im Kühllastbereich als unregelmäßig zu charakterisieren. Dieses Ergebnis lässt sich, wie zuvor bereits erläutert, mit einzelnen gebäudespezifischen individuellen Kühllastprofilen begründen, welche auf individuellem An- und Abschaltverhalten beruhen. Diese Charakteristika in den historischen Datensätzen werden im Modelltrainingsprozess berücksichtigt und spiegeln sich folglich auch in den Prognosen wider. In Abbildung 7.8 sind dazu zwei verschiedene prognostizierte gebäudebezogene Kühllastprofile mit unterschiedlichen Verläufen dargestellt. An dieser Stelle lassen sich bereits Mustervariationen erkennen, die im Rahmen der Ähnlichkeitsanalyse zur Zusammensetzung von Gebäudegruppen genutzt werden. Darüber hinaus sind differierende Zeitpunkte der Lastspitzen mit abweichenden Prominenzen identifizierbar. Beide Effekte finden im weiteren Verlauf dieser Arbeit Eingang in die Auswertung von Implikationen für das kurzfristige Lastmanagement urbaner Energiesysteme. Zudem demonstriert die Generierung kurzfristiger Prognosen die Eignung von *AMSA* für Echtzeitanalysen.



(a) Wärmelast



(b) Kühllast

Abbildung 7.7: Exemplarische Darstellung kurzfristiger aggregierter Wärme- und Kühllastprognosen für das Demonstrationsquartier über 200 Stunden im Winter (Heizfall) bzw. Sommer (Kühlfall)

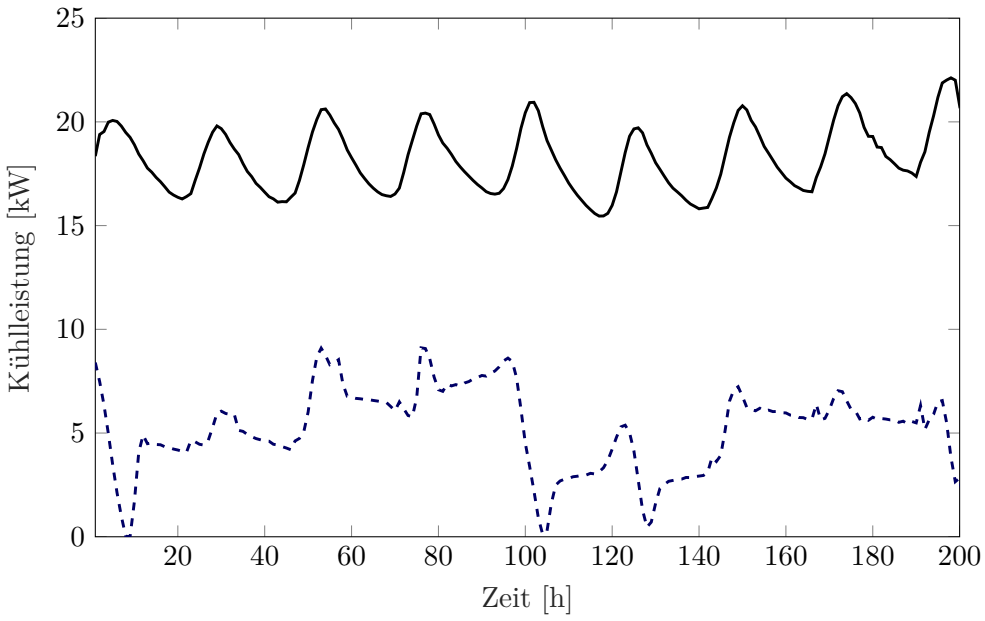
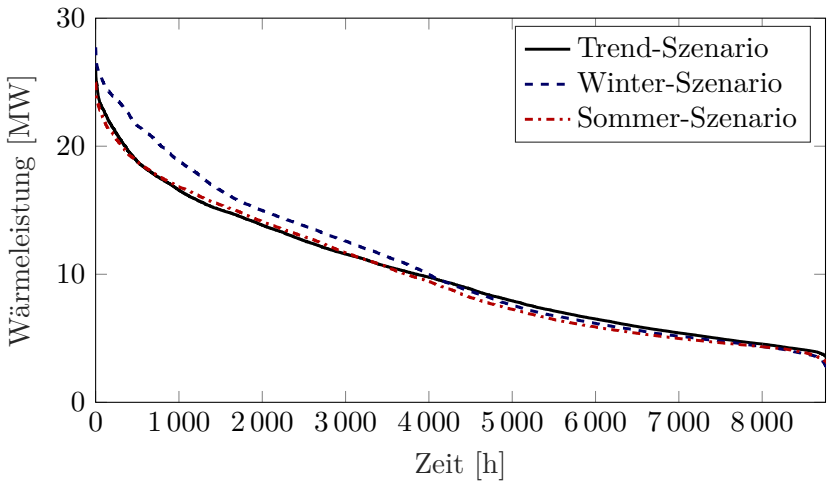


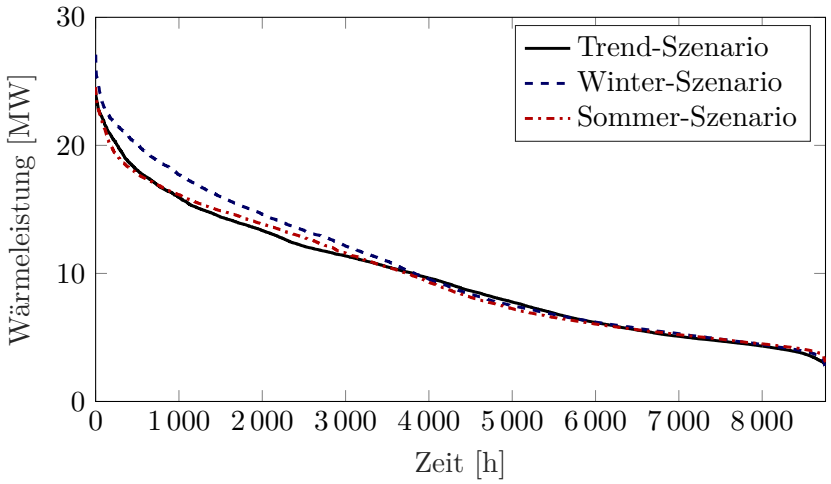
Abbildung 7.8: Exemplarische Darstellung von zwei gebäudespezifischen kurzfristigen Kühllastprognosen über 200 Stunden im Juni

Mittel- und langfristige Lastprognosen auf Basis von Wetterszenarien

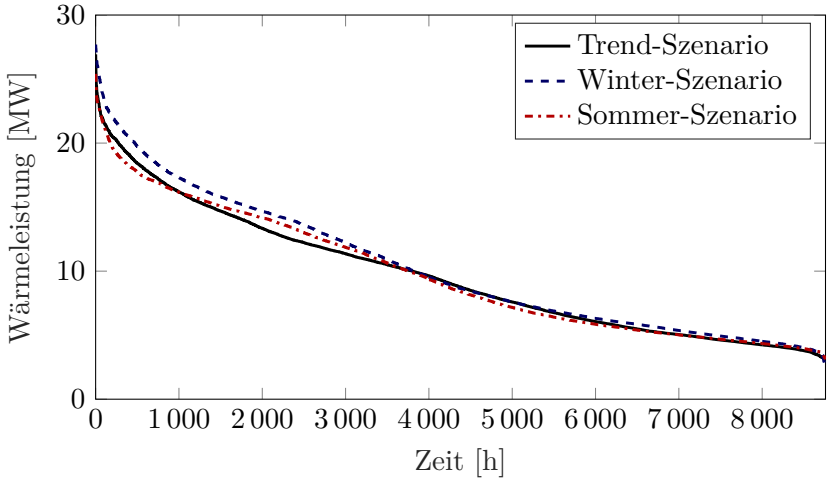
Im Hinblick auf die nachhaltige Planung von Energieversorgungssystemen im Allgemeinen sowie die Quantifizierung und Berücksichtigung von Gleichzeitigkeitseffekten im Speziellen sind lang- und mittelfristige Prognoseergebnisse unabdingbar. Um Sensitivitäten bezüglich zukünftiger Wetterbedingungen, die auf Gebäude einwirken, berücksichtigen und damit verbundene Variationen in den thermischen Lastprofilen abbilden zu können, dienen die zu Beginn dieses Kapitels vorgestellten TRJ Wetterdatensätze. Letztere fungieren jeweils als Modelleingangsparameter, wodurch verschiedene mittel- und langfristige Wärme- und Kühllastprognosen ermöglicht werden. In diesem Zusammenhang sind in den Abbildungen 7.9 und 7.10 die geordneten Jahresdauerlinien der prognostizierten Wärme- und Kühllast exemplarisch für die Jahre 2020, 2025 und 2030 in Abhängigkeit von den drei TRJ-Wetterszenarien dargestellt. Szenarioübergreifend ergibt sich für den jeweiligen Prognosehorizont hinsichtlich der zukünftig anfallenden Wärmelast des Demonstrationsquartiers eine gleichmäßige Aufteilung in Grund-, Mittel- und Spitzenlast. Hinsichtlich der Wärmelastprognosen auf Grundlage des Extrem-Winter-Szenarios mit sehr niedrigen Außenlufttemperaturen im Winter resultiert eine erhöhte Spitzen- und Mittellast für 2020, 2025 und 2030. Vor dem Hintergrund eines steigenden witterungsbedingten Heizbedarfs auf Basis dieses Szenarios ist diese Prognose als realistisch einzustufen. In Bezug auf das Extrem-Sommer-Szenario zeigen sich im Vergleich zum Trend-Wetter-Szenario kaum Unterschiede im Hinblick auf die prognostizierten Wärmelasten. Dennoch ist eine geringe Verschiebung von Spitzen- zu Mittellast festzustellen.



(a) Szenarienbasierte Lastprognosen für 2020

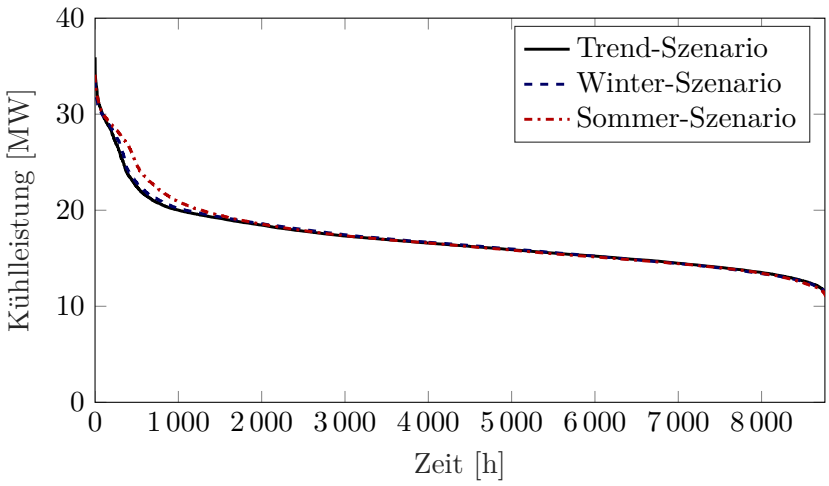


(b) Szenarienbasierte Lastprognosen für 2025

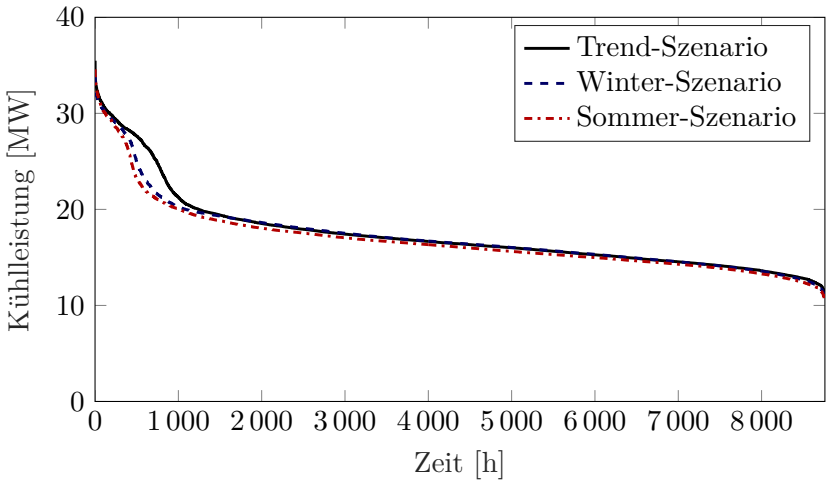


(c) Szenarienbasierte Lastprognosen für 2030

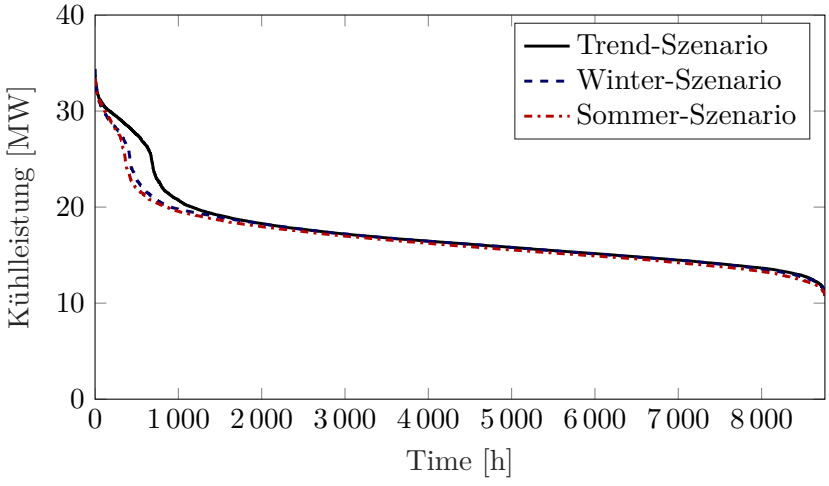
Abbildung 7.9: Darstellung der geordneten aggregierten Dauerlinien zu Wärmelastprognosen für das Demonstrationsquartier in Abhängigkeit von zukünftigen Trend- und Extremwetterszenarien für die Jahre 2020, 2025 und 2030 in Anlehnung an Koschwitz et al., 2019



(a) Szenarienbasierte Lastprognosen für 2020



(b) Szenarienbasierte Lastprognosen für 2025



(c) Szenarienbasierte Lastprognosen für 2030

Abbildung 7.10: Darstellung der geordneten aggregierten Dauerlinien zu Kühllastprognosen für das Demonstrationsquartier in Abhängigkeit von Trend- und Extremwetterszenarien für die Jahre 2020, 2025 und 2030 in Anlehnung an Koschwitz et al., 2019

Dieser Effekt ist mit höheren Außenlufttemperaturen im Winter und dadurch geringerem Wärmebedarf der Gebäude zu erklären. Die Grundlast bleibt szenarioübergreifend für die Jahre 2020, 2025 und 2030 konstant. Letzteres ist mit der stetig anfallenden Prozesswärme zu begründen, welche bei Nichtwohngebäuden häufig anfällt und mit dem entwickelten Modell quantitativ abbildbar ist.

Langfristige Wärmelastprognosen auf Basis von Sanierungsszenarien

Der Fokus der folgenden Untersuchungen liegt auf der Modellierung langfristiger Wärmelastprognosen unter Berücksichtigung des Einflusses verschiedener urbaner Sanierungsreihenfolgen auf Grundlage der in Kapitel 5 vorgestellten Methoden. Letztere und die Ergebnisse der exemplarischen Anwendung basieren auf der Studie *Long-term urban heating load predictions based on optimized retrofit orders: A cross-scenario analysis* von Koschwitz et al. (Koschwitz et al., 2019), wobei im Folgenden die wesentlichen Ergebnisse zusammengefasst aufgeführt werden. Hinsichtlich der Modelldemonstration wurden 51 der 98 betrachteten Gebäude für die Bildung von Sanierungsreihenfolgen berücksichtigt. Neben der herkömmlichen Vorgehensweise zur Ableitung von Sanierungsreihenfolgen, nach denen die Gebäude mit dem höchsten Baualter oder Energieeinsparpotential zuerst saniert werden, erfolgt die Untersuchung des Ansatzes zur Ableitung ökonomisch optimierter Sanierungsfahrpläne von Spinnräker et al. (Spinnräker et al., 2017). Hierbei ergibt sich die Reihenfolge aus der Kapitalwertminimierung für die Investition in sukzessive Gebäudesanierungen, wobei sowohl Sanierungs- als auch Energiekosten sowie ein jährliches Budget als Restriktion berücksichtigt werden (Koschwitz et al., 2019). Hinsichtlich der Lastgangmodifikationen geht damit ein spezifischer Zeitpunkt der Lastgangmodifikation für die einzelnen Gebäude im Betrachtungszeitraum einher. Im Rahmen einer übergreifenden Szenarioanalyse wird insbesondere der Einfluss unterschiedlicher Sanierungsreihenfolgen auf die thermischen Lastgänge untersucht, um Rückschlüsse auf die energetische Planung von Stadtquartieren ziehen zu können. Eingangs erfolgt die Definition der drei Reihenfolgen als Szenarien:

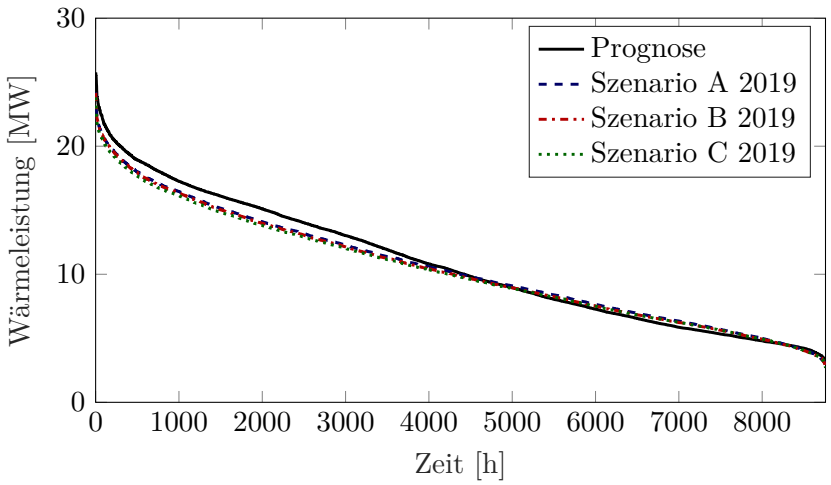
Szenario A: Reihenfolge nach dem Kriterium des höchsten Baualters

Szenario B: Reihenfolge nach dem Kriterium des größten Energieeinsparpotenzials

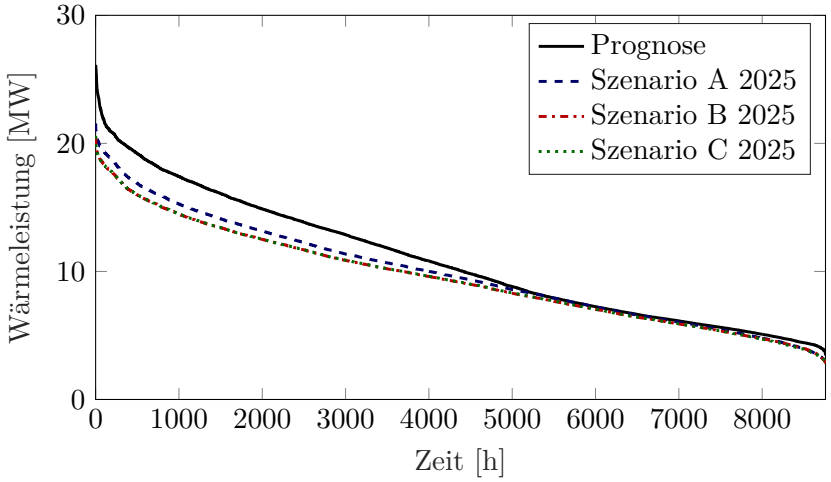
Szenario C: Ökonomisch optimierte Sanierungsreihenfolge

In Bezug auf die numerische Abbildung des Einflusses von Sanierungsmaßnahmen auf einen gebäudebezogenen Wärmelastgang ist zunächst festzulegen, ob ein Gebäude während der Sanierungsphase einen Wärmeverbrauch in einer gegebenenfalls zu spezifizierenden Höhe aufweist oder nicht. Im Kontext dieser Studie wurde die Annahme gemäß Kapitel 5 getroffen, dass kein Wärmeverbrauch während der Sanierungsphase vorliegt. Die Anpassung der Lastprognosen für ein saniertes Gebäude bei Wiederinbetriebnahme erfolgte nach der vorgestellten Methode in Kapitel 5. In diesem Zusammenhang zeigt Abbildung 7.13 die gebäudespezifische Prognose der Wärmeleistungsreduktion infolge angepasster charakteristischer Wärmeverbrauchswerte. Gemäß des ermittelten Sanierungsfahrplans ist für das Jahr 2024 die Sanierung des Gebäudes vorgesehen, sodass der Wärmeleistungsbedarf gemäß

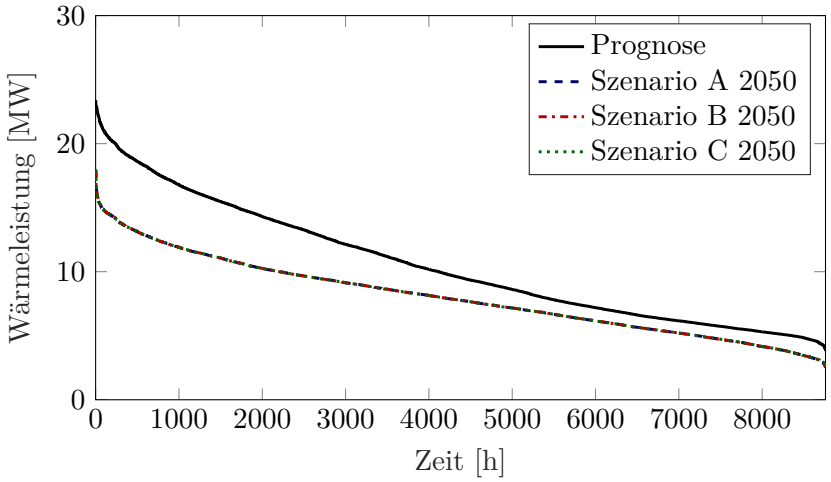
der Modellannahmen zu Null gesetzt wird. Ab einschließlich dem Jahr 2025, in dem das Gebäude wieder genutzt wird, reduziert sich der Wärmeleistungsbedarf unter der Annahme des gebäudenutzungsspezifischen Verbrauchsrichtwerts nach VDI 3807. Für jedes Gebäude des betrachteten Quartiers, welches im Sanierungsfahrplan enthalten ist, findet diese Methode Anwendung. In der Folge ergeben sich fortlaufende Modifikationen hinsichtlich des Wärmeleistungsbedarfs der gesamten Gebäudegruppe. Abbildung 7.11 zeigt dahingehend die geordneten Jahresdauerlinien auf Grundlage der langfristigen Lastprognosen für die Jahre 2019 und 2050 für drei verschiedene Sanierungsreihenfolgen. Die Ergebnisse veranschaulichen zum einen eine Lastverschiebung in 2019 mit einer reduzierten Spitzenlast und einer leicht erhöhten Grundlast. Zum anderen variiert der Wärmeleistungsbedarf in Abhängigkeit von der zugrunde liegenden Sanierungsreihenfolge. Der letztgenannte Effekt lässt sich am Beispiel der Jahresdauerlinien für das Jahr 2025 aufzeigen. Im Hinblick auf den Wärmeleistungsbedarf in 2050 ist eine verstärkte Spitzen- und Mittellast sowie eine moderate Grundlastreduktion festzustellen. Die temporäre Variation der Wärmelastzustände des Quartiers auf Grundlage der drei definierten Szenarien A bis C visualisiert Abbildung 7.12.



(a) Szenarienbasierte Lastprognosen für 2019

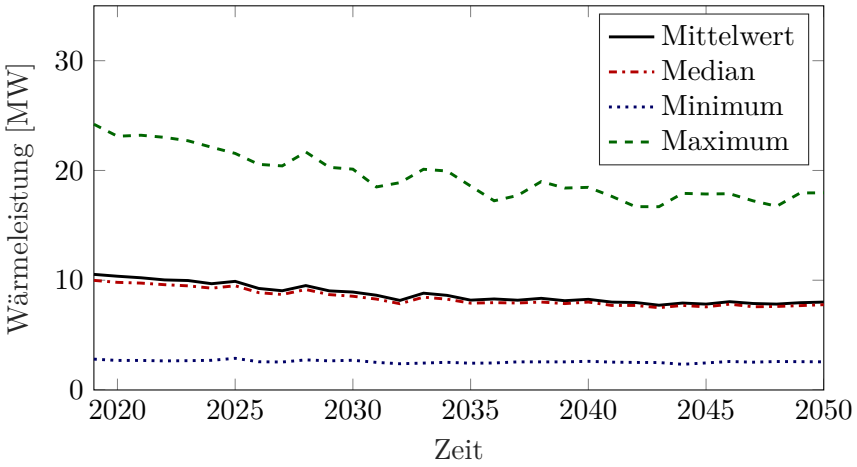


(b) Szenarienbasierte Lastprognosen für 2025

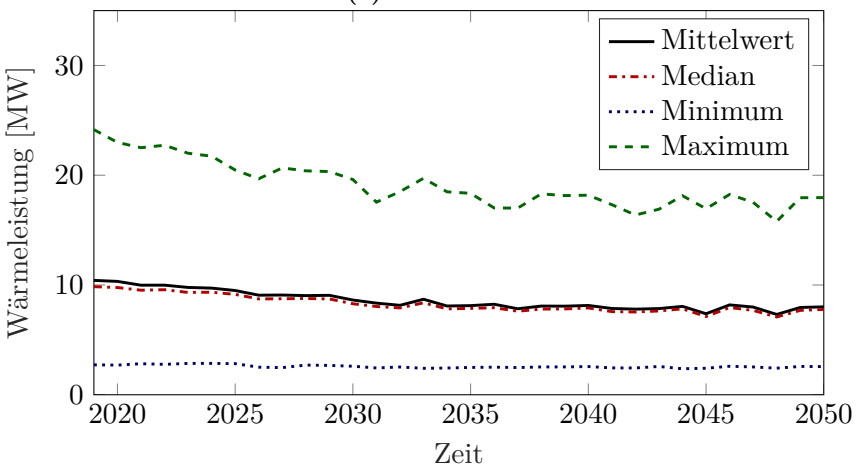


(c) Szenarienbasierte Lastprognosen für 2050

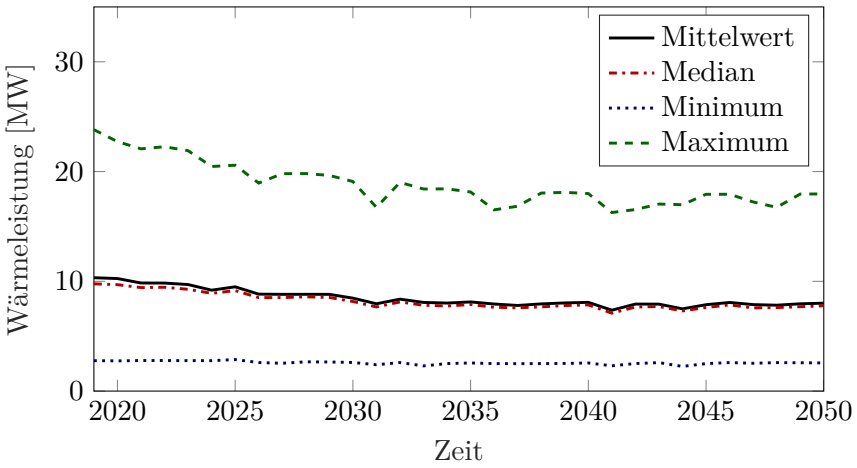
Abbildung 7.11: Lastprognosen und szenarienbasierte Lastmodifikationen infolge verschiedener Sanierungsreihenfolgen für die Jahre 2019, 2025 und 2050 in Anlehnung an Koschwitz et al., 2019



(a) Szenario A



(b) Szenario B



(c) Szenario C

Abbildung 7.12: Szenariospezifische, prognostizierte Maximum-, Minimum-, Median- und Mittelwerte für den Wärmeleistungsbedarf auf Quartiersebene in Anlehnung an Koschwitz et al., 2019

Szenarioübergreifend reduziert sich der Spitzenleistungsbedarf von 2019 bis 2050 um circa sechs Megawatt, wobei die arithmetischen Mittel- und Medianwerte eine Verringerung des Wärmeleistungsbedarfs von ungefähr drei Megawatt aufzeigen. Es ist ferner festzustellen, dass der Spitzenleistungsbedarf zu verschiedenen Zeitpunkten innerhalb des untersuchten Prognosezeitraums um circa bis zu zwei Megawatt szenariospezifisch abweicht. Dieses Ergebnis ist in Abbildung 7.14 dargestellt. Die Grundlast weist dagegen ein nahezu konstantes Niveau über den betrachteten Zeitraum hinweg auf. Dieser Effekt ist ferner auf die Modellannahme zurückzuführen, dass die historische Wärmegrundlast der Gebäude nicht von den Sanierungsmaßnahmen betroffen ist. An dieser Stelle zeigt sich zudem, dass Modifikationen der Zusammensetzung ähnlichkeitsbasierter Lastprofilgruppen sowie szenariobedingte Abweichungen in der Höhe des Gleichzeitigkeitsfaktors zu erwarten sind.

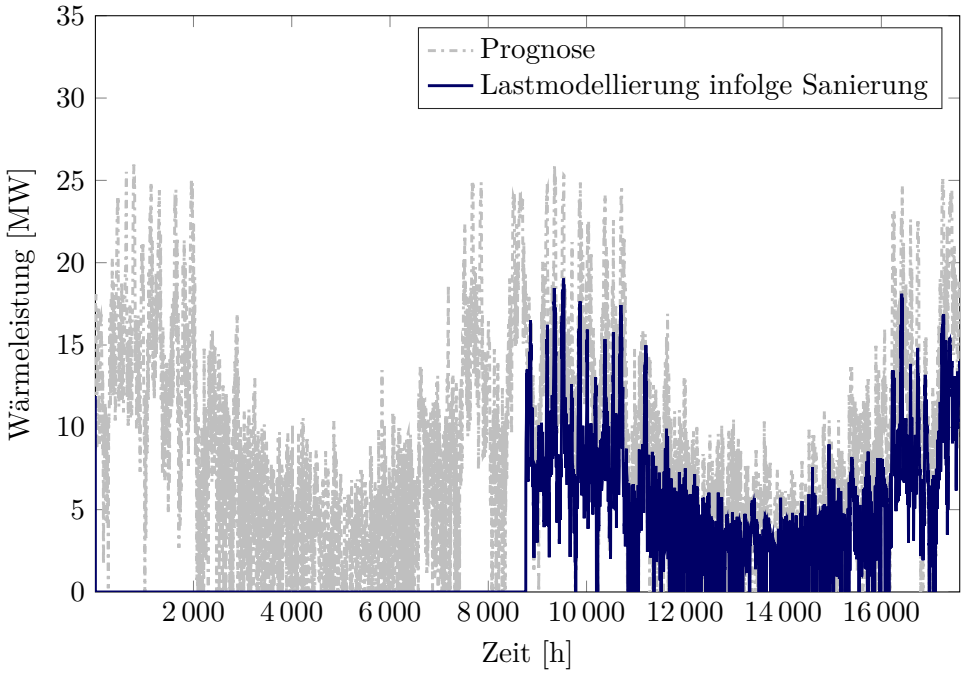


Abbildung 7.13: Exemplarisches Ergebnis der Wärmeleistungsreduktionen infolge Gebäudesanierung in Anlehnung an Koschwitz et al., 2019

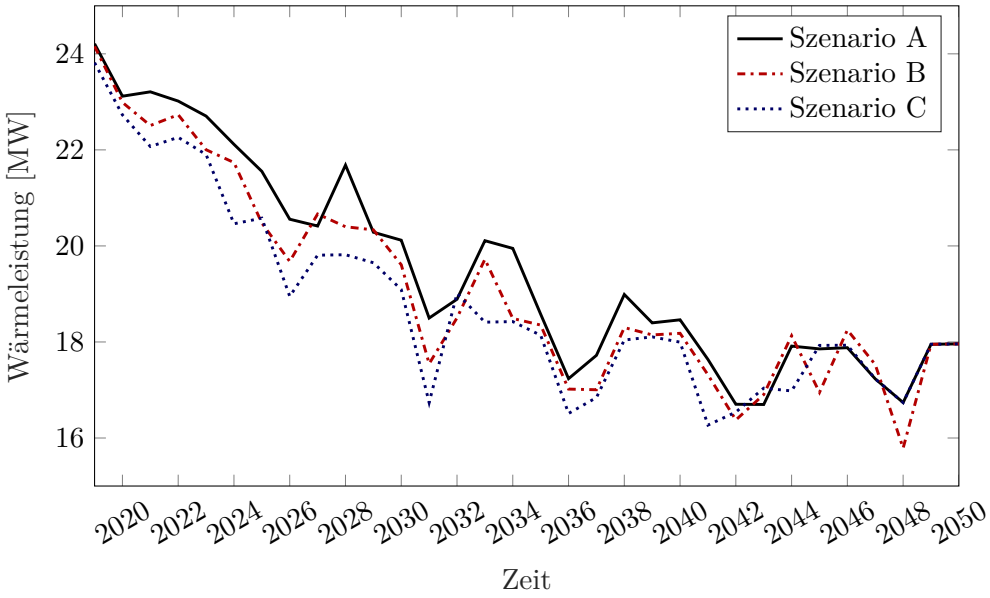


Abbildung 7.14: Vergleich der prognostizierten Fluktuationen von maximalen Wärmeleistungsspitzen in Bezug auf die Szenarien A bis C in Folge einzelner Gebäudesanierungen in Anlehnung an Koschwitz et al., 2019

7.2.3 Ergebnisse des Clusterings von Lastprofilen

Grundlage der nachfolgenden Untersuchungen sind unterschiedliche Detailgrade hinsichtlich des Clusterings der thermischen Lastprofile als Basis für die ähnlichkeitsbasierten Gruppierungen von thermischen Lastprofilen. Im Folgenden dienen die in Kapitel 5 definierten drei Detailgrade für das Clustering der weiteren Analyse. Die Untersuchung beruht auf 47 historischen Kühllastgängen, wobei, wie in Kapitel 5 erläutert, die Messdaten des Sommerhalbjahres Anwendung finden. Beginnend bei dem feinen Detailgrad zur Clusteranalyse dienen 49 Neuronen der Gruppierung von normierten Lastprofilen. Abbildung 7.15 visualisiert die Anzahl der Lastprofile, welche einem Neuron bzw. Referenzvektor zugeordnet wurden. Es ist festzustellen, dass sich auf Grundlage des Detailgrads *fein* zwei Cluster mit jeweils drei und ein Cluster mit zwei Lastgängen bilden. Die verbleibenden Cluster bestehen maximal aus einem Lastgang. Die Verwendung des mittleren Detailgrades, welcher an dieser Stelle 24 Neuronen entspricht, führt zu einem abweichenden Ergebnis des Clusterings, das in Abbildung 7.16 dargestellt ist. Infolge der Verringerung der Anzahl an Neuronen und damit auch jener der möglichen Clusterzentren werden leere Cluster vermieden und die Gruppengrößen erhöht. Die maximale Anzahl an Lastprofilen in einem Cluster beträgt 10.

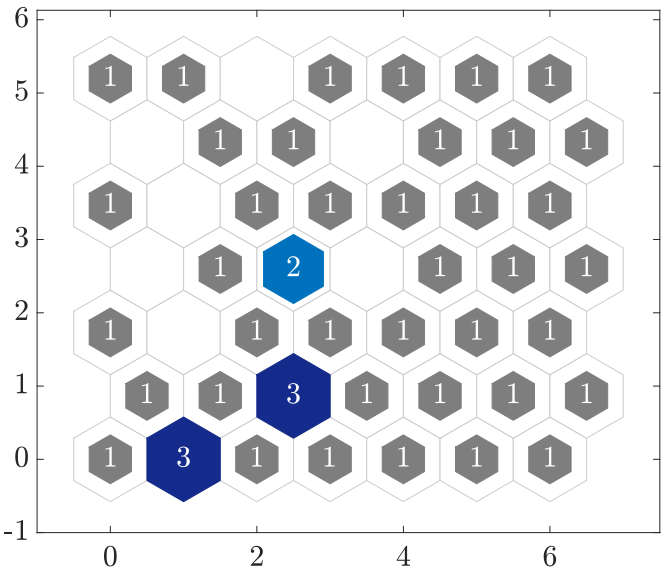


Abbildung 7.15: Darstellung der SOM-Topologie und der dazugehörigen 49 Neuronen mit der Anzahl an zugewiesenen Eingangsvektoren auf Grundlage der Gruppierung von Kühllastprofilen des Sommerquartals

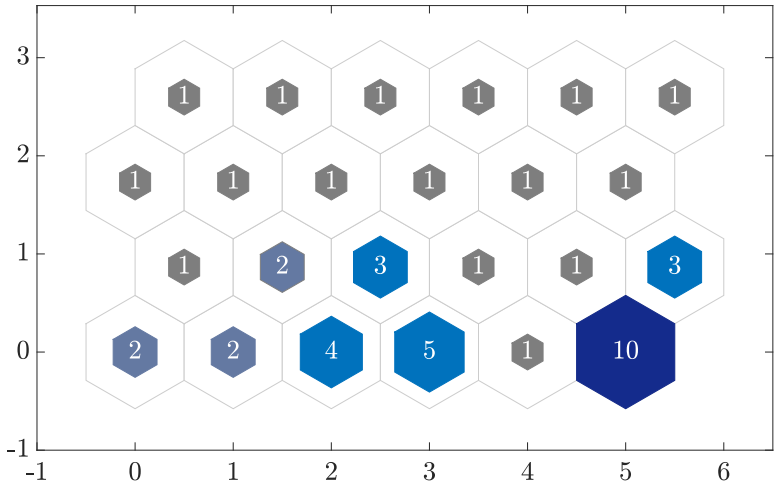


Abbildung 7.16: Darstellung der SOM-Topologie und der dazugehörigen 24 Neuronen mit der Anzahl an zugewiesenen Eingangsvektoren auf Grundlage der Gruppierung von Kühllastprofilen des Sommerquartals

Im weiteren Verlauf werden die Ergebnisse des Clusterings mit dem Detailgrad *gering* und sechs Neuronen vorgestellt, welche in Abbildung 7.17 dargestellt sind. Die Gruppengrößen lassen sich somit erhöhen, wobei gleichzeitig Lastprofile identifiziert werden, die Ausreißer darstellen und jeweils ein eigenes Cluster bilden. Für jedes Cluster kann ferner ein gemit-

teltes Standardlastprofil erstellt werden, wodurch sich Unterschiede zwischen den Clustern anhand der Lastprofilcharakteristika aufzeigen lassen. Abbildung 7.18 veranschaulicht die Standardlastprofile für sechs Cluster, wobei die Cluster 3 und 6 die Ausreißer abbilden.

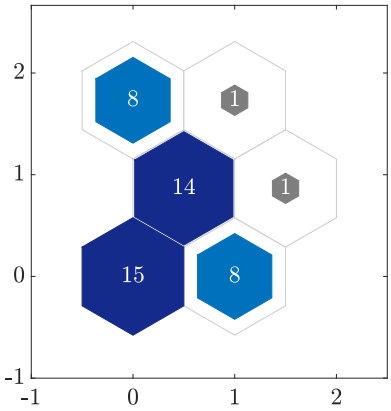
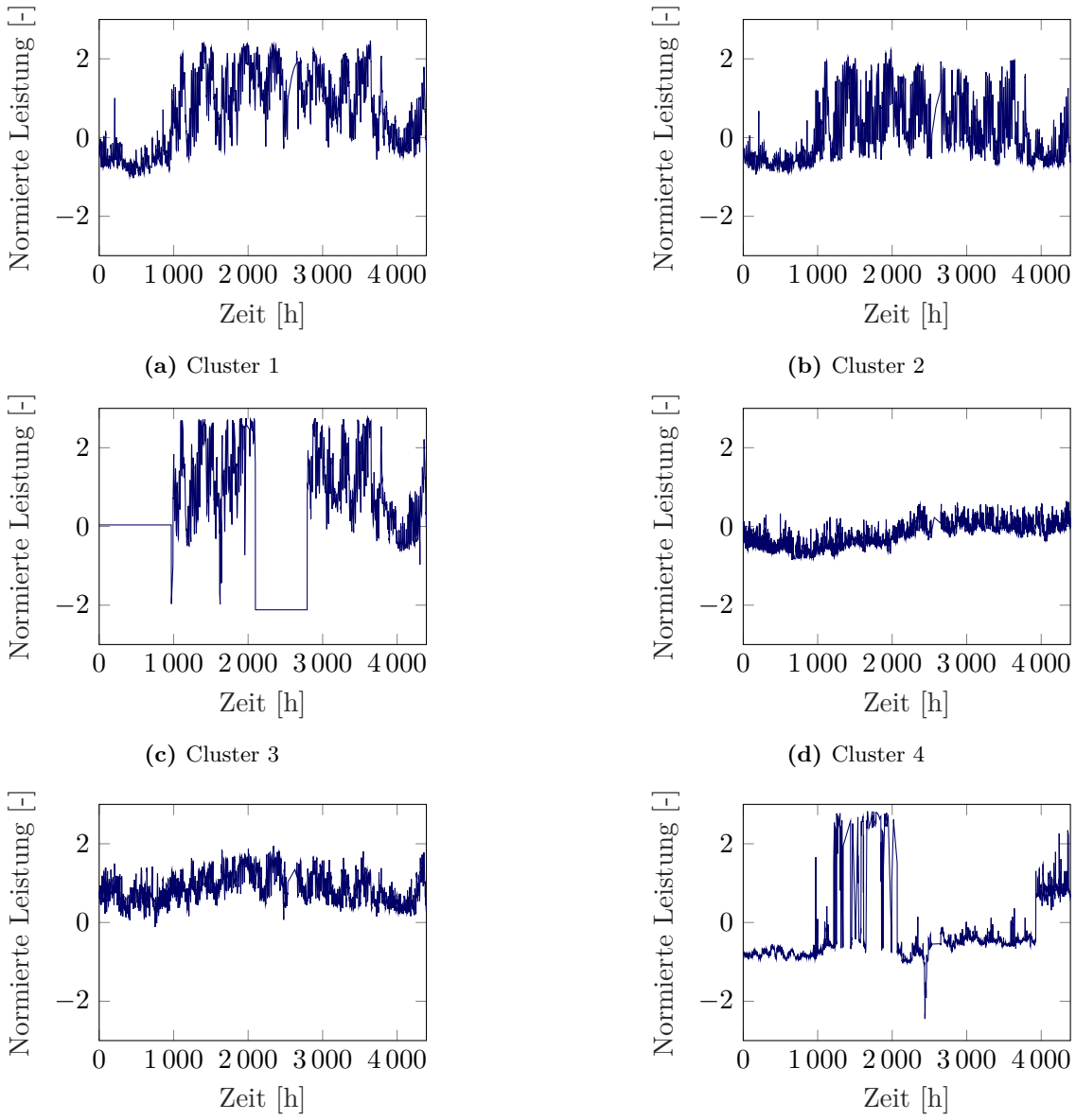


Abbildung 7.17: Darstellung der SOM-Topologie und der dazugehörigen 6 Neuronen mit der Anzahl an zugewiesenen Eingangsvektoren auf Grundlage der Gruppierung von Kühllastprofilen des Sommerquartals

Um Gruppierungen der Lastprofile auf Grundlage von Datensätzen mehrerer Jahre zu ermöglichen, werden die Clusterzugehörigkeiten im Folgenden nach dem Superpositionsprinzip überlagert.

Tabelle 7.12: Cluster für die Kälteversorgung auf Grundlage von zwei analysierten Jahren mit maximal 6 Clustern und mehr als einem Mitglied je Cluster

Gruppennummer	**Gebäudenummern**
1	4, 8, 9, 10, 15, 18, 20, 21, 24
2	5, 13, 23, 26, 29
3	34, 37, 38
4	32, 46
5	40, 44
6	6, 7, 14, 47
7	16, 19, 25
8	2, 27, 36, 43
9	11, 31, 45
10	1, 12, 22
11	17, 28, 30, 33, 42



(a) Cluster 1

(b) Cluster 2

(c) Cluster 3

(d) Cluster 4

(e) Cluster 5

(f) Cluster 6

Abbildung 7.18: Darstellung der gruppenspezifischen Standardlastprofile auf Grundlage des Clusterings mit einem geringen Detailgrad

Dadurch lassen sich Gebäudegruppen definieren, die längerfristig die selben charakteristischen Lastprofile aufweisen, wodurch das Ergebnis des Clusterings eine erhöhte Verallgemeinerbarkeit aufweist. Am Beispiel von sechs Clustern sind in Tabelle 7.12 die resultierenden 11 Gebäudegruppen mit mehr als einem Mitglied aufgelistet. Durch die Überlagerung der Cluster verdoppelt sich die Anzahl der Gebäudegruppen mit mindestens zwei Mitgliedern nahezu. Analog zu Tabelle 7.12 sind in Tabelle 7.13 die Ergebnisse der Clustersuperpositionen für maximal 24 Clusterzentren aufgeführt. In diesem Fall führt die Verfeinerung des Clusterings zu einer ähnlichen Anzahl an Gebäudegruppen mit mehr als einem Mitglied im Gegensatz zum Clustering auf Basis der einzelnen Jahresdatensätze. Ein gegenteiliger Effekt ist für die Überlagerung der Ergebnisse des Clusterings mit 49 Neuronen festzustellen, welche in Tabelle 7.14 aufgelistet sind.

Tabelle 7.13: Cluster für die Kälteversorgung auf Grundlage von zwei analysierten Jahren mit maximal 24 Clustern und mehr als einem Mitglied je Cluster

Gruppennummer	**Gebäudenummern**
1	17, 30
2	1, 22, 27
3	6, 14
4	4, 8, 9, 15, 21
5	10, 24
6	37, 38
7	5, 26, 29

Tabelle 7.14: Cluster für die Kälteversorgung auf Grundlage von zwei analysierten Jahren mit maximal 49 Clustern und mehr als einem Mitglied je Cluster

Gruppennummer	**Gebäudenummern**
1	6, 14
2	1, 22

Hier verringert sich die Anzahl der Gebäudegruppen, sodass zwei Gruppen mit jeweils zwei Gebäuden verbleiben. An dieser Stelle erreicht das feine Clustering einen sehr hohen Detailgrad, wodurch nur noch sehr ähnliche Lastprofile mit einer hohen Gleichzeitigkeit in ihren charakteristischen Verläufen zusammen gruppiert werden. Dazu ist beispielhaft ein Auszug aus zwei Lastgängen, welche einem Cluster entsprechen, in Abbildung 7.19 dargestellt.

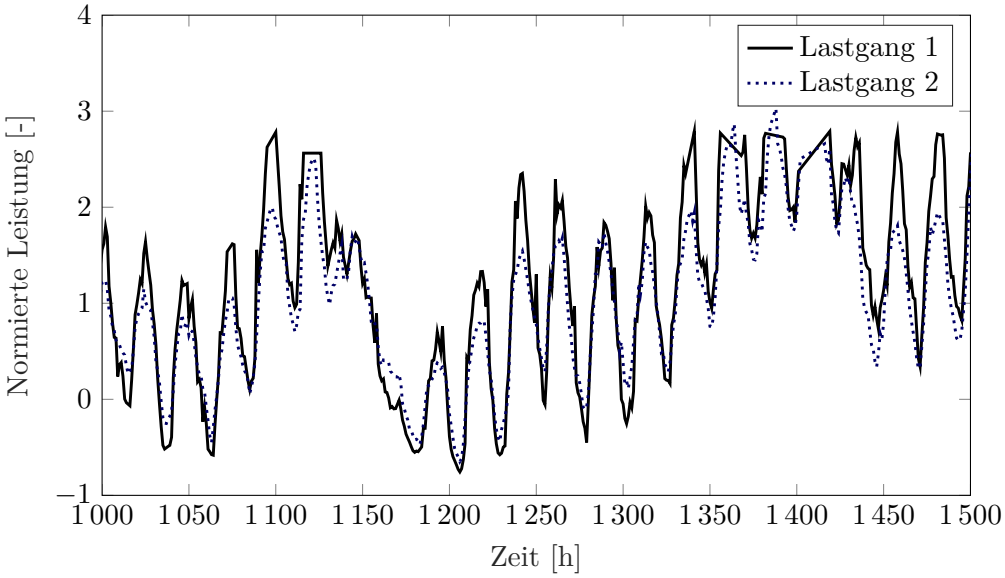


Abbildung 7.19: Exemplarischer Auszug aus den Ergebnissen der Clusterüberlagerungen mit zwei gruppierten Lastprofilen

7.3 Ergebnisse der Quantifizierung von Gleichzeitigkeitseffekten

Im Rahmen dieses Abschnitts erfolgt die Ergebnisanalyse zur Quantifizierung von Gleichzeitigkeitseffekten im Kontext des vorliegenden Anwendungsfalls. Eingangs werden Gleichzeitigkeitsfaktoren auf Basis der konventionellen Methode für die Wärme- und Kälteversorgung ermittelt. Anschließend dienen die Resultate der Anwendung von *AMSA* einerseits zum Vergleich mit konventionellen Referenzwerten, andererseits demonstrieren sie die Vorteile der im Rahmen dieser Arbeit entwickelten Erweiterung des Methodenspektrums zur Gleichzeitigkeitsanalyse.

7.3.1 Konventionelle Ermittlung von Gleichzeitigkeitsfaktoren

Im weiteren Verlauf werden die Ergebnisse zur Herleitung von Kennlinien für den Gleichzeitigkeitsfaktor anhand der gebäudespezifischen Zeitreihen zu historischen Wärme- sowie Kühllasten vorgestellt und erläutert. Die dazugehörigen methodischen Grundlagen sind in Anhang A, Abschnitt A.3, zusammengefasst. Letztere beruhen im Wesentlichen auf der allgemeinen Definition des Gleichzeitigkeitsfaktors nach Gleichung (2.1). Mit der Auswertung von Gleichzeitigkeitsfaktoren für verschiedene Abnehmergruppenzusammensetzungen und Gruppengrößen, lassen sich, wie in Kapitel 2 erläutert, Wertebereiche definieren. Die Untersuchungen basieren auf zufällig zusammengesetzten Gebäudegruppen unterschiedlicher Größe. Letztere variiert in der Wärmeversorgung in einem Wertebereich zwischen 1 und 63, wohingegen für die Kälteversorgung Gruppengrößen von 1 bis 17 betrachtet werden. Die Limitation der Gruppengrößen auf ungefähr zwei Drittel der Gesamtabnehmeranzahl

beruht auf der Überlegung, dass andernfalls zu geringe Variationsmöglichkeiten bei den Gruppenzusammensetzungen bestehen (Winter et al., 2001). Dies betrifft insbesondere größere Gruppen. In Bezug auf die Ermittlung eines gruppenspezifischen repräsentativen Gleichzeitigkeitsfaktor wird das 95%-Quantil verwendet, um den Einfluss von Ausreißern zu reduzieren. Hinsichtlich der Höhe des Gleichzeitigkeitsfaktors für die Wärmeversorgung bei variierender Abnehmerzahl ist in Abbildung 7.20 der funktionale Verlauf auf Grundlage verschiedener Stichprobengrößen visualisiert.

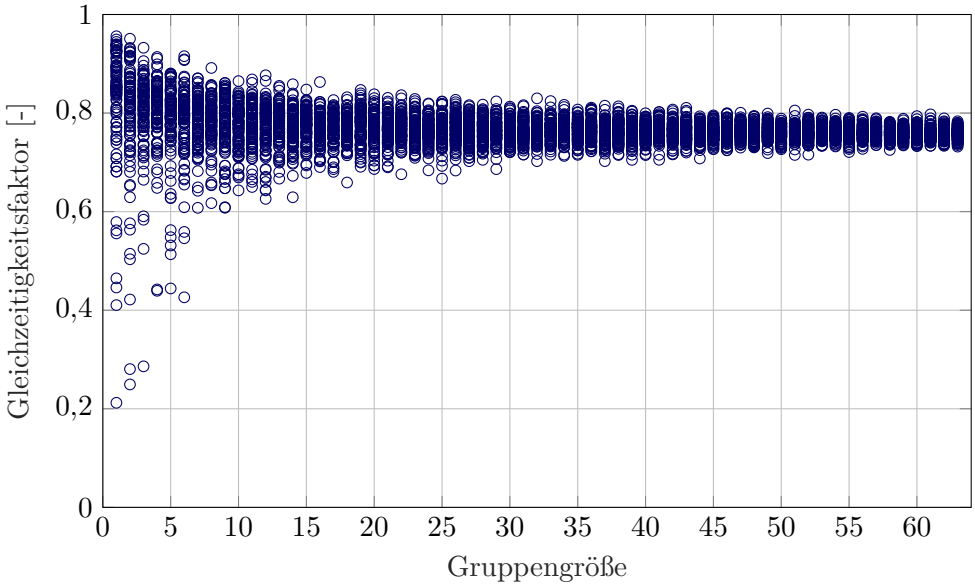


Abbildung 7.20: Ergebnis der statistischen Analyse zur Entwicklung des Gleichzeitigkeitsfaktors für die Wärmeversorgung in Abhängigkeit von der Abnehmergruppengröße

In Abbildung 7.21 sind darauf aufbauend die approximierte Näherungsfunktion auf Grundlage eines 95%-Konfidenzintervalls für die Quantifizierung des Gleichzeitigkeitsfaktors in Abhängigkeit von der Größe der Wärmeabnehmergruppe dargestellt. Auf Grundlage der Kurvenanpassung ergibt sich ein Bestimmtheitsmaß von 0,97, wobei die Funktion nach Gleichung (7.2) definiert ist.

$$f(x) = \begin{cases} C_1 \cdot x^{C_2} + C_3 & x \geq 2 \\ \\ 1 & x < 2 \end{cases} \tag{7.2}$$

Die dazugehörigen Funktionsparameter C_1, C_2 und C_3 sind nachfolgend aufgeführt.

$$C_1 = 0{,}5542$$
$$C_2 = -0{,}1162$$
$$C_3 = 0{,}4448$$

Abbildung 7.22 zeigt den Vergleich der Näherungsfunktionen für gebäudespezifische Nennleistungen auf Grundlage einer Norminnentemperatur von 20 und 22 °C im Rahmen einer durchgeführten Sensitivitätsanalyse. Für letztere ist in Anhang D, Abschnitt D.2 das Ergebnis der statistischen Auswertung aufgeführt. Eine Verschiebung der Norminnentemperatur im Zuge der Ermittlung der Normheizlast mittels des Regressionsverfahrens um zwei Kelvin bedingt eine Abweichung beider Kennlinien voneinander um 0,04. Dementsprechend bewirkt die Erhöhung der angenommenen Norminnentemperatur für den gesamten Gebäudebestand eine Reduzierung des Gleichzeitigkeitsfaktors um 0,04. Abbildung 7.23 veranschaulicht zudem den Vergleich zwischen den Kennlinien von Winter et al. und Tol et al. sowie der im Rahmen dieser Arbeit hergeleiteten Funktion des Gleichzeitigkeitsfaktors in Abhängigkeit von der Wärmeabnehmeranzahl.

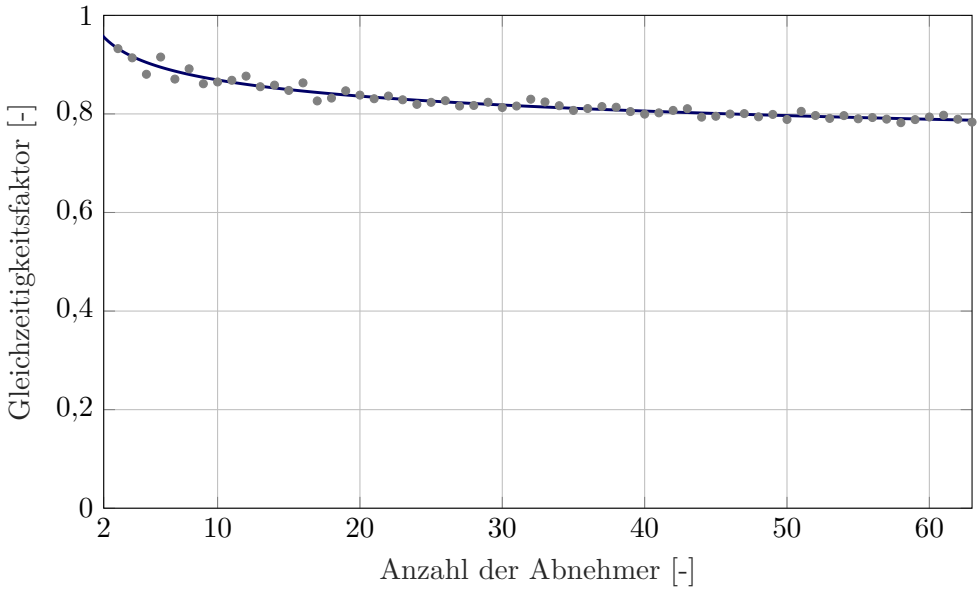


Abbildung 7.21: Darstellung der Näherungsfunktion zur Approximation des Gleichzeitigkeitsfaktors für die Wärmeversorgung in Abhängigkeit von der Größe der Abnehmergruppe

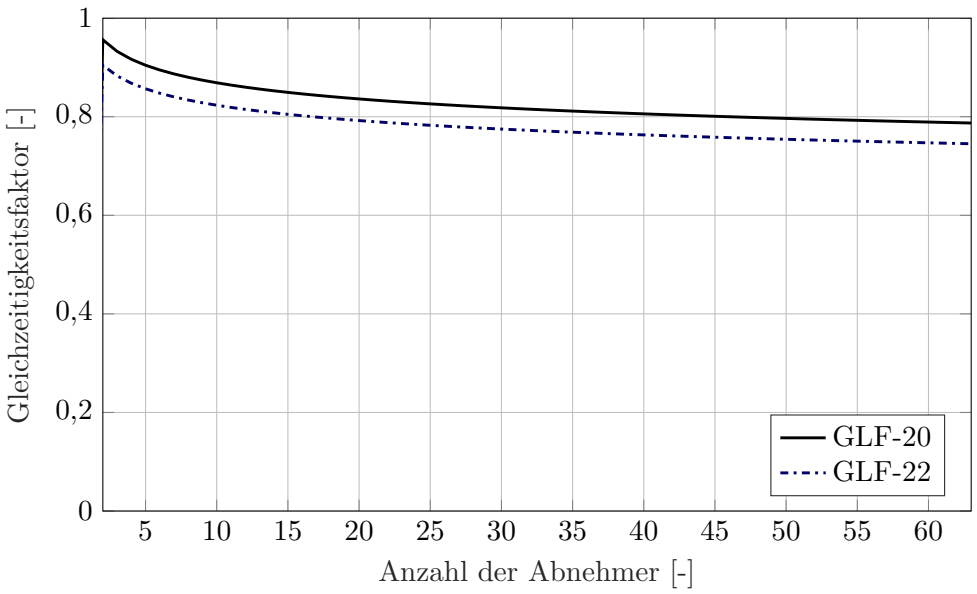


Abbildung 7.22: Vergleich der Näherungsfunktionen zur Approximation des Gleichzeitigkeitsfaktors für die Wärmeversorgung in Abhängigkeit von der Größe der Abnehmergruppe und der gewählten Norminnentemperatur im Hinblick auf die Bestimmung der Nennleistungen

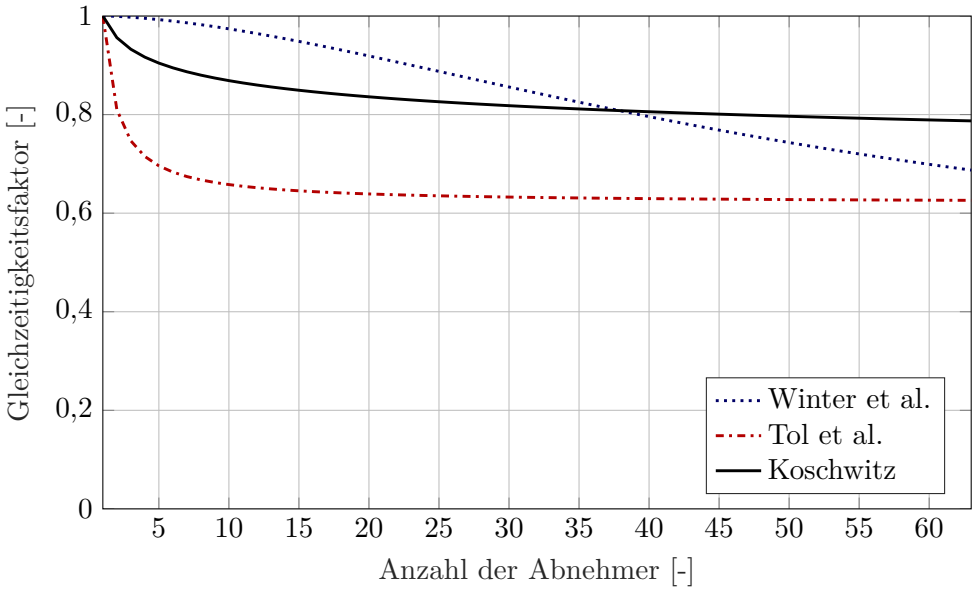


Abbildung 7.23: Vergleich der im Rahmen dieser Arbeit hergeleiteten Kennlinie für den Gleichzeitigkeitsfaktor in Abhängigkeit von der Abnehmeranzahl mit jenen auf Grundlage der Studien von Winter et al. und Tol et al.

Letztere weist einen ähnlichen charakteristischen Verlauf zu der Kennlinie von Tol et al. auf, wobei der Wertebereich 0,2 bis 0,3 erhöht ist. Ferner ist festzustellen, dass der Gleichzeitigkeitsfaktor für das in dieser Arbeit untersuchte Gebiet der Kennlinie von Tol et al. folgend als zu gering angesetzt werden würde. Dies liegt in erster Linie darin begründet, dass der funktionale Zusammenhang für Wohngebäude hergeleitet wurde, während die Gleichzeitigkeitsfaktoren in dieser Arbeit für Nichtwohngebäude berechnet werden. Weiterhin ist die erhöhte Gleichzeitigkeit im betrachteten Forschungszentrum Jülich auf unter anderem homogene Gebäudenutzungszeiten infolge ähnlicher Arbeits- und Ferienzeiten zurückzuführen, wodurch Wärmespitzen in einem begrenzten Zeitraum zu erwarten sind. In Bezug auf den Vergleich mit der Kennlinie von Winter et al. ist ferner festzustellen, dass der Gleichzeitigkeitsfaktor für Gruppengrößen bis ungefähr 37 höher anzusetzen ist. Demgegenüber ist der Gleichzeitigkeitsfaktor für das vorliegende Gebiet für Gruppen mit mehr als 37 Gebäuden zu gering dimensioniert. Zusammenfassend resultieren diese Ergebnisse in den folgenden Schlussfolgerungen:

I. Für das vorliegende Versorgungsgebiet führt die Quantifizierung des Gleichzeitigkeitsfaktors nach Tol et al. zu einer Unterdimensionierung von zentralen Wärmeerzeugungsanlagen und Wärmenetzen, wodurch Versorgungsengpässe entstehen können.

II. Für das vorliegende Versorgungsgebiet führt die Quantifizierung des Gleichzeitigkeitsfaktors nach Winter et al. zu Überdimensionierungen bei kleinen und Unterdimensionierungen bei großen Versorgungsgruppen, wodurch im ersten Fall ökonomische sowie ökologische Ineffizienzen und im zweiten Fall Versorgungsengpässe zu erwarten sind.

Daher ist an dieser Stelle festzustellen, dass eine Übertragbarkeit von Kennlinien nicht ausreichend für die Planung von Versorgungsgebieten ist. Analog zur Wärmeversorgung ist in Abbildung 7.24 der funktionale Verlauf des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der untersuchten Stichprobengröße dargestellt. Abbildung 7.25 zeigt die approximierte Näherungsfunktion im Rahmen des 95%-Konfidenzintervalls für die Quantifizierung des Gleichzeitigkeitsfaktors in Abhängigkeit von der Größe der Kälteabnehmergruppe. Auf Grundlage der Kurvenanpassung ergibt sich ein Bestimmtheitsmaß von 0,91, wobei die Funktion nach Gleichung (7.3) definiert ist.

$$f(x) = \begin{cases} C_1 \cdot x^{C_2} + C_3 & x \geq 3 \\ \\ 1 & x < 3 \end{cases} \tag{7.3}$$

Die dazugehörigen Funktionsparameter C_1, C_2 und C_3 sind nachfolgend aufgeführt.

$C_1 = 0{,}4618$
$C_2 = -0{,}5640$
$C_3 = 0{,}7515$

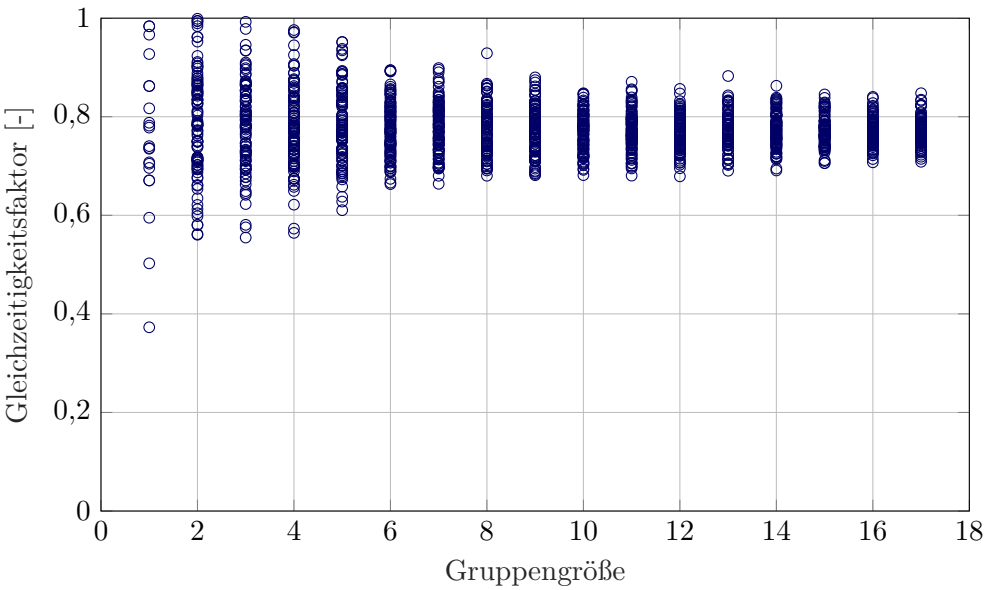


Abbildung 7.24: Ergebnis der statistischen Analyse zur Entwicklung des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der Gruppengröße

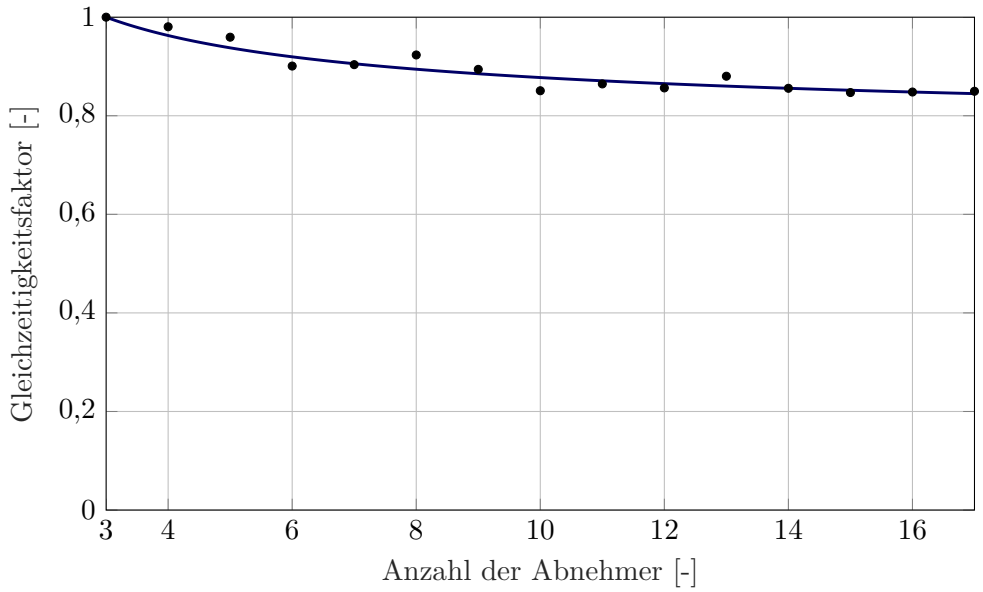


Abbildung 7.25: Darstellung der Näherungsfunktion zur Approximation des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der Größe der Abnehmergruppe

Abbildung 7.26 zeigt den Vergleich der Näherungsfunktionen für gebäudespezifische Nennleistungen auf Grundlage einer Auslegungstemperatur von 33 und 35 °C. Die Ergebnisse der statistischen Auswertung für die letztgenannte Außentemperatur ist ferner Anhang

D, Abschnitt D.2 zu entnehmen. Eine eindeutige Trennung beider Leistungsanteile ist im Nichtwohngebäudebereich oftmals schwer möglich, da der Wärmebedarf beispielsweise durch nacht- und wochendspezifische Regelungsstrategien beeinflusst wird (Frederiksen et al., 2014), wodurch sich von dem obigen Schema abweichende Datenpunkte ergeben. An dieser Stelle können beispielhafte Visualisierungen dieses Problems Anhang D, Abschnitt D.1 entnommen werden. Daher muss die Heizgrenztemperatur manuell festgelegt werden. Im Rahmen von durchgeführten Energieaudits liegen weitere Informationen für einige Gebäude vor. In diesem Zusammenhang wurde unter anderem der gebäudespezifische Wärmebedarf in einen statischen und in einen dynamischen Anteil aufgeteilt. Dahingehend zeigen sich Unterschiede zwischen den Gebäuden im Hinblick auf die Aufteilung zwischen statischem und dynamischen Wärmebedarf, welche im Detail Tabelle C.4 zu entnehmen sind. Es sind dabei sowohl Gebäude mit ausschließlich statischem Wärmebedarf als auch Gebäude mit 95% dynamischem Wärmebedarf vorhanden. Damit ist neben den unterschiedlichen Baualtersklassen auch eine Heterogenität in der Charakteristik der Lastprofile für den Wärmebedarf anzunehmen.

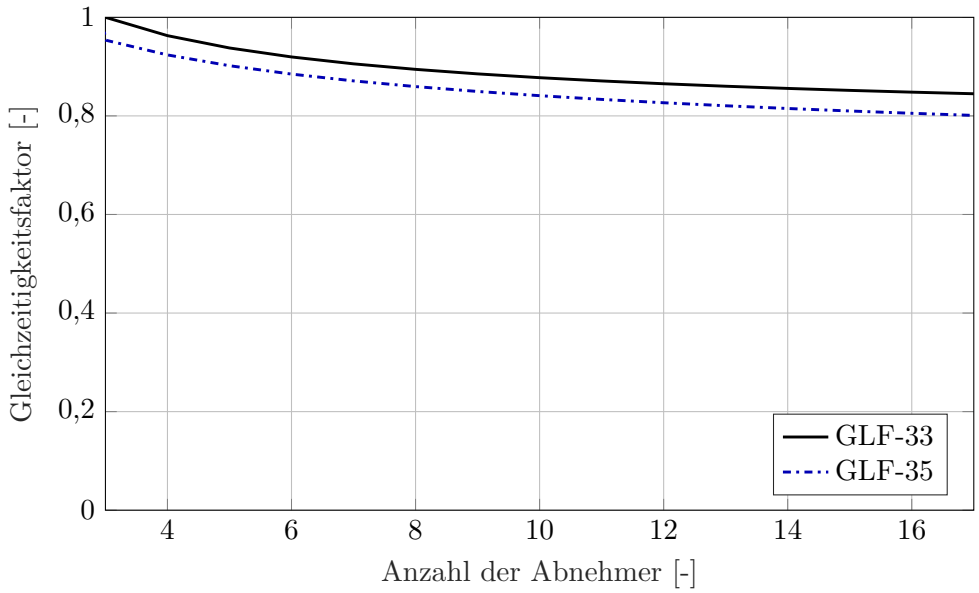


Abbildung 7.26: Vergleich der Näherungsfunktionen zur Approximation des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der Größe der Abnehmergruppe und der gewählten Auslegungstemperatur im Hinblick auf die Bestimmung der Nennleistungen

7.3.2 Quantifizierung von Gleichzeitigkeitseffekten mit AMSA

Der folgende Unterabschnitt beinhaltet abschließend die Ergebnisse zur Quantifizierung von Gleichzeitigkeitseffekten mit *AMSA* im Rahmen exemplarischer Analysen. Im Hinblick auf die vorangestellte Herleitung der Kennlinien werden im Folgenden die Resultate der

szenariobasierten Ermittlung von Gleichzeitigkeitsfaktoren unter Berücksichtigung von Unsicherheit auf Quartiersebene analysiert. Dazu dienen für die Wärmeversorgung der Gebäude exemplarisch die folgenden Szenarien zur Modelldemonstration:

W1. Gleichzeitigkeitsfaktoren auf Grundlage der Analyse historischer Messdaten und 10 untersuchter Wärmeleistungsspitzen

W2. Gleichzeitigkeitsfaktoren auf Grundlage der Analyse von Wärmelastprognosen im Zeitraum von 2019 bis 2025 unter Verwendung des Trend-TRJ und 50 untersuchter Leistungsspitzen

W3. Gleichzeitigkeitsfaktoren auf Grundlage der Analyse von Wärmelastprognosen im Zeitraum von 2019 bis 2025 unter Verwendung des Extrem-Winter-TRJ und 50 untersuchter Leistungsspitzen

W4. Gleichzeitigkeitsfaktoren auf Grundlage der Analyse von Wärmelastprognosen im Zeitraum von 2019 bis 2025 unter Verwendung des Trend-TRJ und einer ökonomisch optimierten Sanierungsreihenfolge sowie 50 untersuchter Leistungsspitzen

Die Ergebnisse sind in Abbildung 7.27 mittels Boxplots dargestellt.

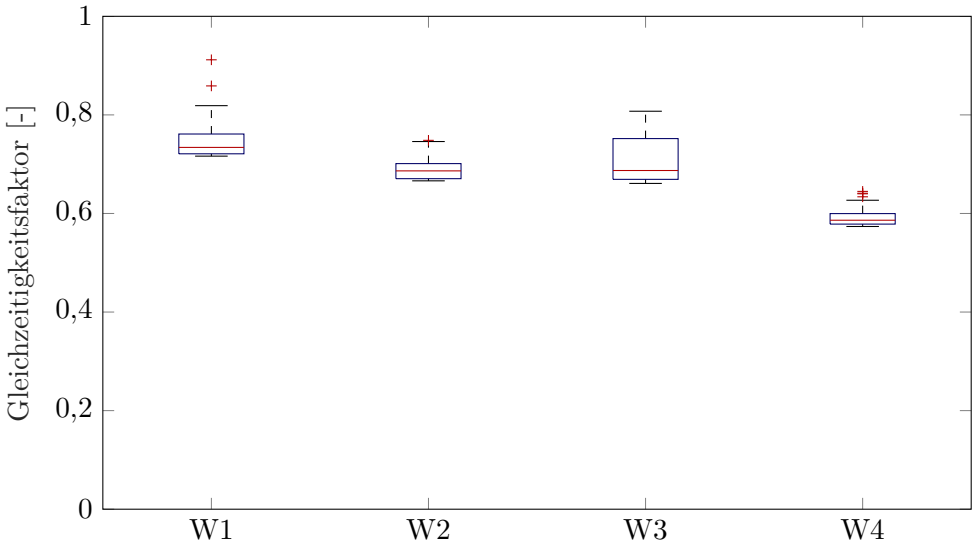


Abbildung 7.27: Spektrum des Gleichzeitigkeitsfaktors auf Grundlage der 10 höchsten Wärmeleistungsspitzen im Kontext historischer Wärmeleistungsdaten aus den Jahren 2016 und 2017 sowie auf Basis der 50 höchsten Wärmeleistungsspitzen in den Wärmebedarfsprognosen bis 2025

Es ist festzustellen, dass die Maximalwerte für den Gleichzeitigkeitsfaktor szenarioübergreifend zwischen ungefähr 0,91 in W1 und 0,65 in W4 variieren. Im Vergleich dazu ergibt sich auf Basis der hergeleiteten Kennlinie für die betreffende Gesamtgruppengröße ein

Wert von ungefähr 0,77. Somit würden im Hinblick auf W1 sowohl thermische Energieerzeuger als auch Verteilungssysteme unterdimensioniert werden. Unter Berücksichtigung zukünftiger Lastzustände, welche an dieser Stelle exemplarisch für den Zeitraum bis 2025 prognostiziert sind, ergibt sich ein geringerer Maximalwert für den Gleichzeitigkeitsfaktor unter Berücksichtigung von Gebäudesanierungen. Dadurch sind Ineffizienzen infolge überdimensionierter Erzeuger und Netze zu erwarten, sofern der Kennlinienwert angesetzt wird. Ferner variiert auch das Wertespektrum als Entscheidungskorridor in den Szenarien. Die Ergebnisse implizieren an dieser Stelle, dass die Spitzenlast beispielsweise bei Umsetzung der Sanierungsreihenfolge im Vergleich zum konventionell angesetzten Kennlinienwert um 12% sinkt.

Im Kontext der Planung zur Kälteversorgung des Quartiers werden für die Modellevaluation die folgenden Szenarien exemplarisch analysiert:

C1. Gleichzeitigkeitsfaktoren auf Grundlage der Analyse historischer Messdaten und 10 untersuchter Kühlleistungsspitzen

C2. Gleichzeitigkeitsfaktoren auf Grundlage der Analyse von Kühllastprognosen im Zeitraum von 2019 bis 2025 unter Verwendung des Trend-TRJ und 50 untersuchter Leistungsspitzen

C3. Gleichzeitigkeitsfaktoren auf Grundlage der Analyse von Kühllastprognosen im Zeitraum von 2019 bis 2025 unter Verwendung des Extrem-Sommer-TRJ und 50 untersuchter Leistungsspitzen

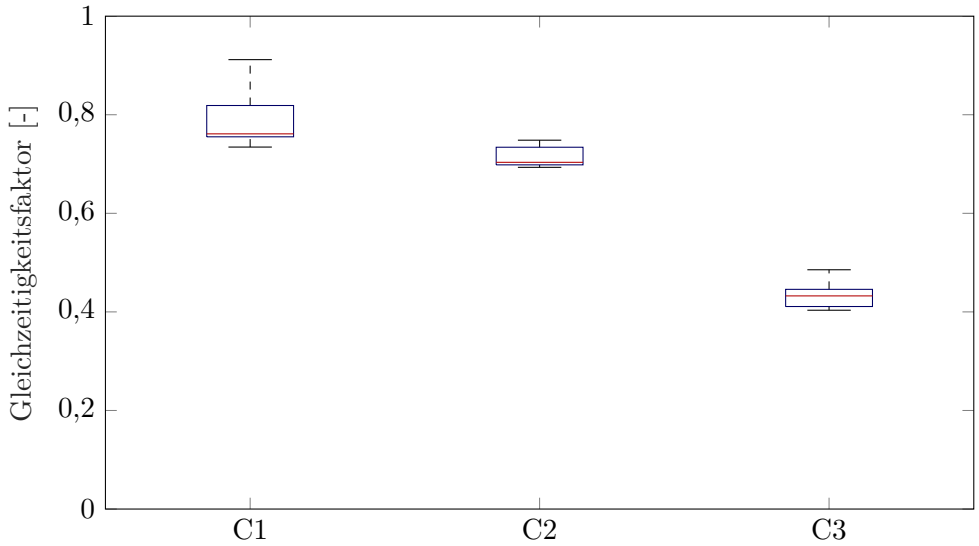


Abbildung 7.28: Spektrum des Gleichzeitigkeitsfaktors auf Grundlage der 10 höchsten Kühlleistungsspitzen im Kontext historischer Kühlleistungsdaten aus den Jahren 2016 und 2017 sowie auf Basis der 50 höchsten Kühlleistungsspitzen in den Kühlbedarfsprognosen bis 2025

Abbildung 7.28 visualisiert die mit *AMSA* generierten Ergebnisse. Auf Grundlage der hergeleiteten Näherungsgleichung ergibt sich ein Gleichzeitigkeitsfaktor von ungefähr 0,8, wobei mit *AMSA* Maximalwerte im Bereich zwischen circa 0,91 und 0,49 in Abhängigkeit vom betrachteten Szenario generiert werden. An dieser Stelle zeigt sich die starke Abhängigkeit des Gleichzeitigkeitsfaktors von den angenommen Randbedingungen. In der Planung des Energiesystems könnte im vorliegenden Fall beispielsweise ein Wert im Bereich 0,49 für die Dimensionierung eines zentralen Kälteversorgungssystems angesetzt werden und Speichersysteme zur Verschiebung von Lastspitzen fungieren. Alternativ könnten lokal installierte Kältemaschinen ergänzend zur zentralen Erzeugung für die Abdeckung von Lastspitzen dienen. In diesem Zusammenhang ist beispielsweise auch zu prüfen, ob Abwärme zur Verfügung steht, die mittels Absorptionskältemaschinen genutzt werden kann. Es ist zusammenfassend an dieser Stelle festzustellen, dass mit *AMSA* differenziertere und zukunftsfähigere Entscheidungen in der Dimensionierungsplanung von urbanen Energiesystemen möglich sind. So lassen sich verschiedene Szenarien zur Entwicklung des Gleichzeitigkeitsfaktors in den Vergleich von Versorgungsvarianten integrieren.

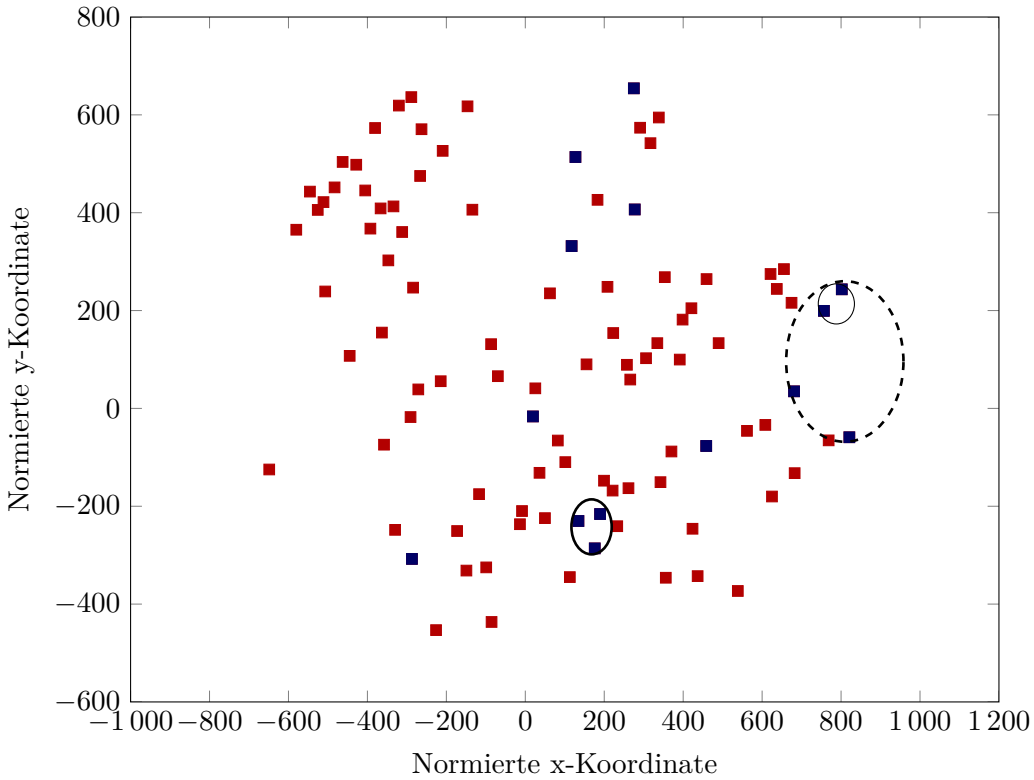


Abbildung 7.29: Darstellung eines exemplarischen Ergebnisses zur Identifizierung von Versorgungsinseln auf Grundlage der Gleichzeitigkeitsanalyse mit *AMSA*

In Bezug auf die Auswertung der Ähnlichkeitsanalysen zeigt Abbildung 7.29 exemplarisch

die Verortung einer Gebäudegruppe (blau markiert) mit ähnlichen Lastprofilen im betrachteten Quartier dieser Arbeit. Es lassen sich am Beispiel des vorliegenden Ergebnisses mindestens drei mögliche Versorgungsinseln identifizieren. Als Indikatoren dienen dabei die räumliche Nähe der Gebäude zueinander und die Ähnlichkeit der Lastprofile, welche kontinuierliche Spitzenlastverschiebungen der Gebäudegruppe durch die Integration zentraler Speichersysteme ermöglichen.

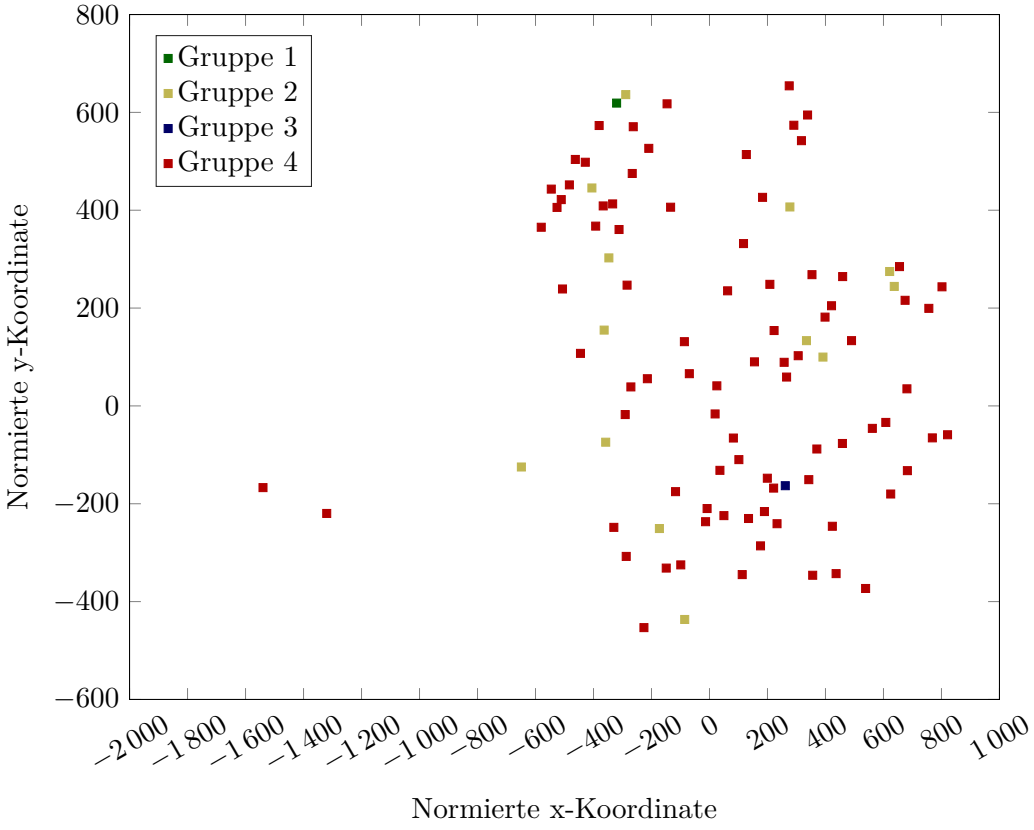


Abbildung 7.30: Darstellung der Gebäudegruppierungen auf Grundlage von Wärmelastprognosen für das Jahr 2025

Im Hinblick auf die Analyse langfristiger Lastszenarien sind in Abbildung 7.30 die Gebäudegruppierungen auf Grundlage von Wärmelastprognosen für das Jahr 2025 dargestellt. Abbildung 7.31 visualisiert ergänzend dazu die Modifikation der Gruppenzugehörigkeiten infolge der Berücksichtigung der ökonomisch optimierten Sanierungsreihenfolge. Somit können szenariobasiert Insellösungen für die Energiesystemvarianten identifiziert werden. Darüber hinaus lassen sich Anhaltspunkte für Lastausgleichspotenziale auf Grundlage der Gebäudegruppierungen generieren, welche in den Unterschieden zwischen Lastprofilen der einzelnen Versorgungsgruppen begründet liegen. In diesem Zusammenhang eignet sich insbesondere der Detailgrad *fein*, der im Zuge der Festlegung der Cluster-Granularität in

AMSA ausgewählt werden kann.

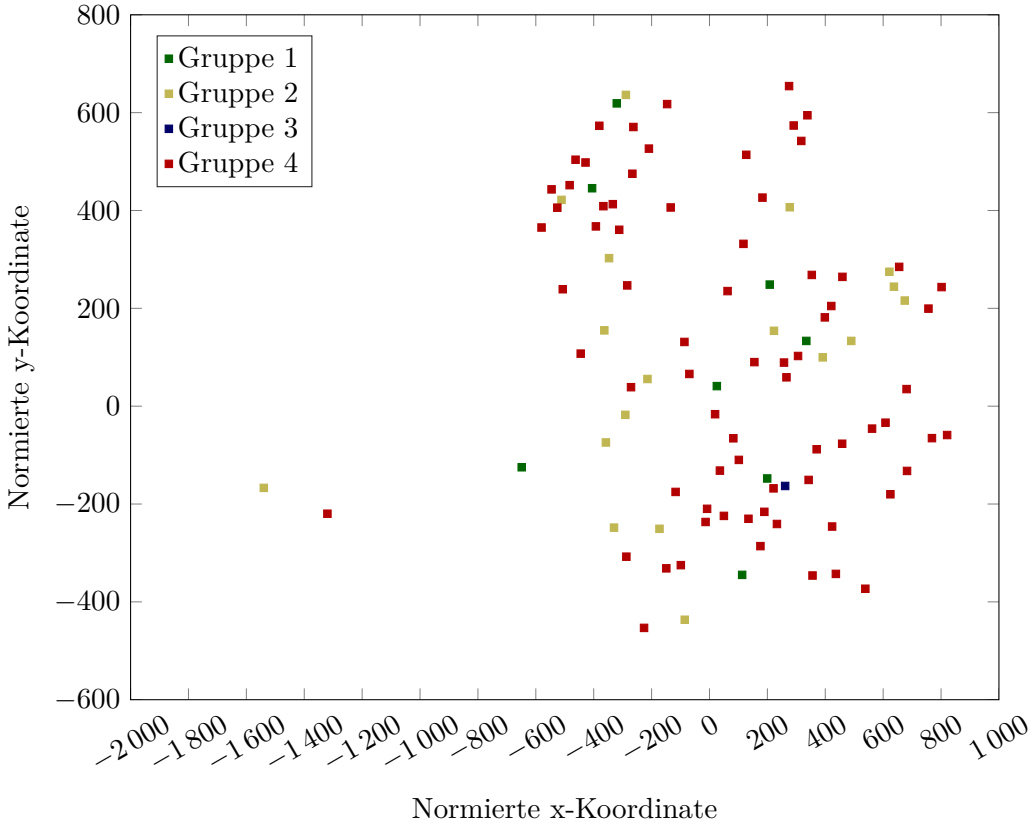


Abbildung 7.31: Darstellung der Gebäudegruppierungen auf Grundlage von Wärmelastprognosen für das Jahr 2025 unter Berücksichtigung der ökonomisch optimierten Sanierungsreihenfolge im Quartier

In Kombination mit der Quantifizierung von gruppenspezifischen Anteilen an den Lastspitzen im Quartier dient die Ähnlichkeitsanalyse ferner der Quantifizierung des Lastverschiebungspotenzials. Für die weiteren Analysen in diesem Abschnitt gelten die folgenden Bezeichnungen:

LV. Gebäudegruppe mit dem größten kontinuierlichen Lastverschiebungspotenzial

KLV1. Gebäudegruppe mit geringerem kontinuierlichen Lastverschiebungspotenzial

KLV2. Gebäudegruppe mit geringerem kontinuierlichen Lastverschiebungspotenzial

KLV3. Gebäudegruppe mit geringerem kontinuierlichen Lastverschiebungspotenzial

Abbildungen 7.32 und 7.33 veranschaulichen exemplarisch das Ergebnis vor dem Hintergrund von Lastprognosen für das Jahr 2025. Hierbei zeigt sich, dass mit *AMSA* die Gruppe *LV* auf Grundlage des kontinuierlichen ganzheitlichen Lastverschiebungspotenzials

identifiziert wird, da sowohl historisch als auch auf Basis der Lastprognosen konstant hohe Spitzenlastanteile quantifiziert wurden. Gruppe *KLV* weist dabei ein etwas geringfügigeres Potenzial auf. Es ist jedoch darüber hinaus auch weiterhin festzustellen, dass sich neben der Gruppe *LV* auch weitere Lastprofilgruppen für Lastverschiebungen eignen.

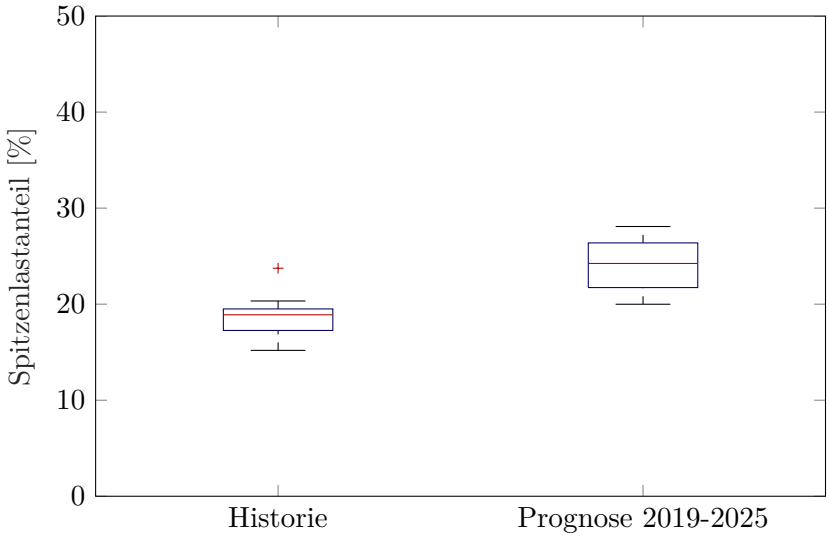


Abbildung 7.32: Gebäudegruppe (*LV*) mit dem größten kontinuierlichen Lastverschiebungspotenzial auf Grundlage von Wärmelastprognosen für das Jahr 2025

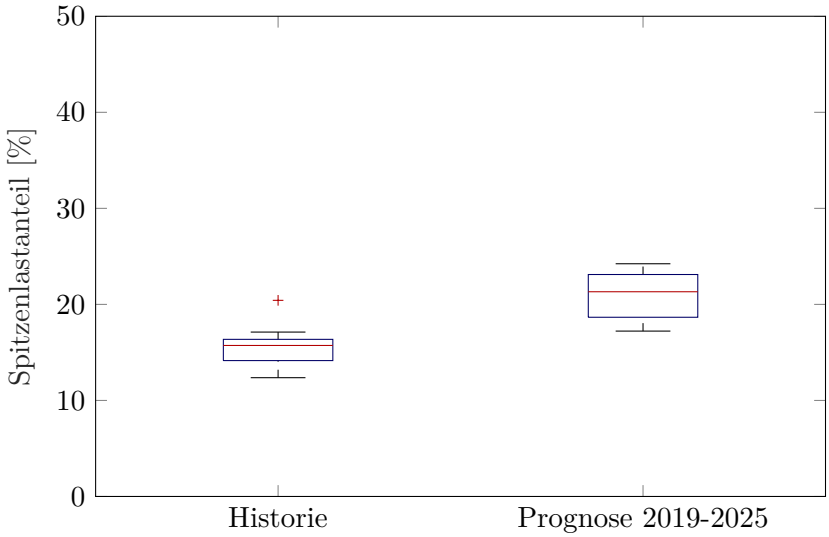


Abbildung 7.33: Gebäudegruppe (*KLV*) mit dem zweitgrößten kontinuierlichen Lastverschiebungspotenzial auf Grundlage von Wärmelastprognosen für das Jahr 2025

In Bezug auf die Analyse der Gleichzeitigkeitseffekte in der Lastverteilung auf Grundlage

kurzfristiger Wärme- und Kühllastprognosen sind in den Abbildungen 7.34 und 7.35 die maximalen Anteile von jeweils vier klassifizierten Gebäudegruppen am Spitzenwärme- bzw. Spitzenkühlleistungsbedarf dargestellt. Es wird dabei ferner zwischen folgenden Varianten unterschieden.

24h. Untersuchung von einer Leistungsspitze im Betrachtungszeitraum

72h. Untersuchung von fünf Leistungsspitzen im Betrachtungszeitraum

168h. Untersuchung von zehn Leistungsspitzen im Betrachtungszeitraum

Die blau gefärbten Säulen repräsentieren die Gruppe, deren aggregiertes Lastprofil auf Basis der Klassifikation mit *AMSA* die größte Ähnlichkeit zu jenem auf Quartiersebene aufweist. Diese sind zudem mit der Abkürzung *LV* (Lastverschiebung) gekennzeichnet. Die verbleibenden Gruppen weisen den Zusatz *KLV* (Keine Lastverschiebung) auf. Weiterhin ist bei der Interpretation der Säulendiagramme zu beachten, dass die Anteile für die 72- und 168-Stunden-Prognosezeiträume jeweils in Summe nicht hundert Prozent ergeben, da mehr als eine Leistungsspitze analysiert wurde und der maximale Spitzenlastanteil damit zu unterschiedlichen Zeitpunkten auftreten kann.

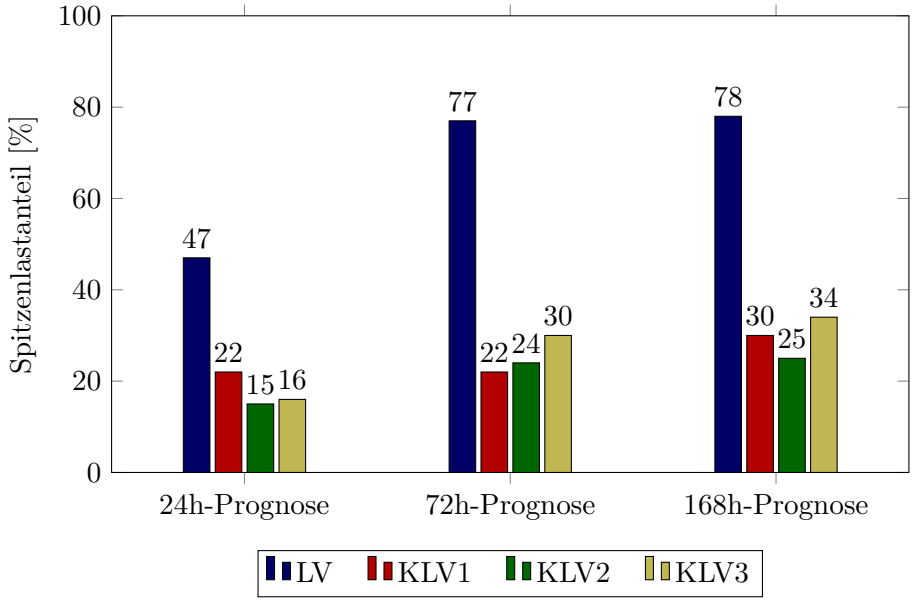


Abbildung 7.34: Prozentualer Anteil des gruppenspezifischen Wärmeleistungsbedarfs zum Zeitpunkt des Spitzenwärmeleistungsbedarfs auf Quartiersebene in Abhängigkeit vom Prognosezeithorizont und analysierter Leistungsspitzenzahl

Anhand des gruppierten Säulendiagramms für die Analyse der Wärmelastverteilungen ist festzustellen, dass Gruppe 1 mit 47%, 77% und 78% die jeweils größten Anteile am Wärmespitzenleistungsbedarf auf Quartiersebene besitzt. Die verbleibenden klassifizierten Gebäudegruppen weisen maximal einen 34%-Anteil auf. Dieses Ergebnis findet infolge der Analyse der Kühllastverteilungen Bestätigung: Der maximale Anteil von Gruppe 1 am

Spitzenkühlleistungsbedarf des gesamten Quartiers beträgt in Abhängigkeit vom jeweiligen Prognosehorizont 56%, 74% bzw. 80%, während die verbleibenden Gruppen maximal einen Anteil von 23% abdecken.

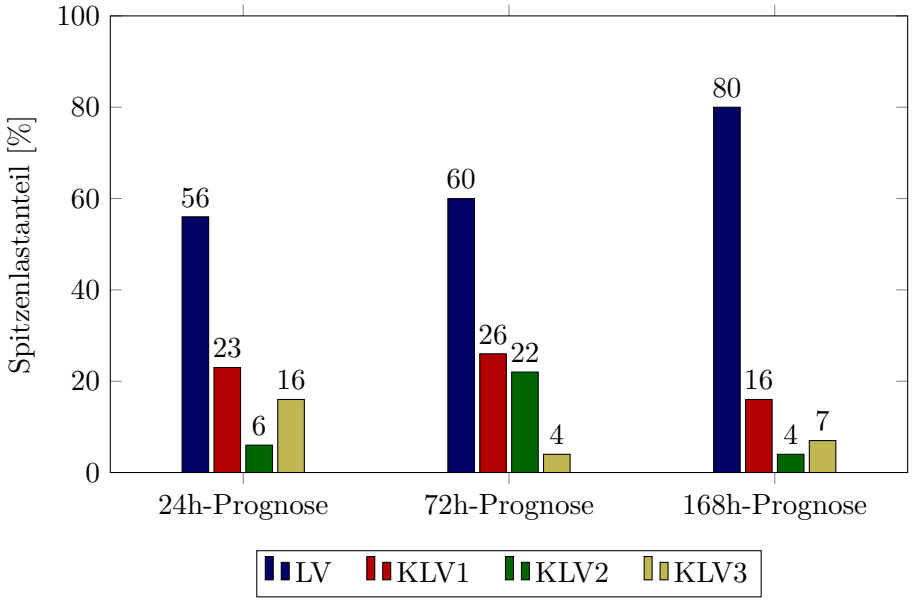


Abbildung 7.35: Prozentualer Anteil des gruppenspezifischen Kühlleistungsbedarfs zum Zeitpunkt des Spitzenkühlleistungsbedarfs auf Quartiersebene in Abhängigkeit vom Prognosezeithorizont und der Leistungsspitzenzahl

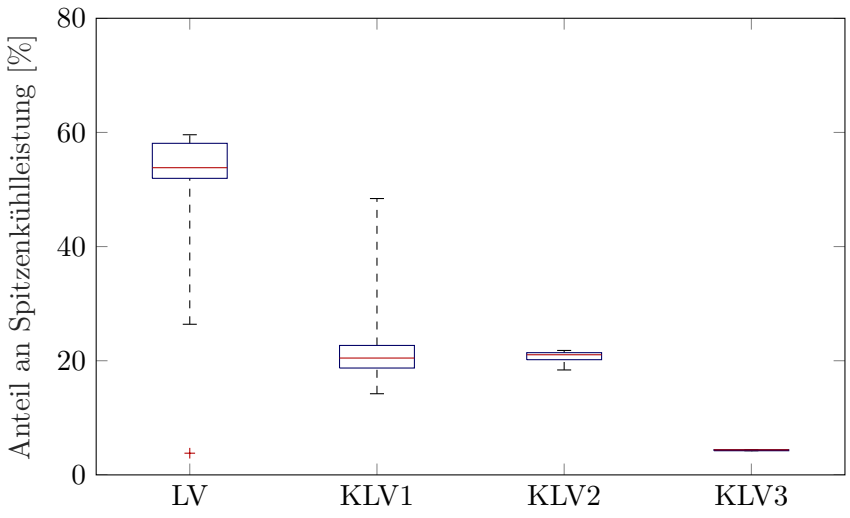


Abbildung 7.36: Gruppenspezifische Anteile an Spitzenkühllasten auf Quartiersebene auf Basis von 72-Stunden-Prognosen und fünf untersuchten Leistungsspitzen

Die Ergebnisse dieses Unterabschnitts lassen folglich den Schluss zu, dass mittels Clustering und Klassifikation Gebäudegruppen identifiziert werden können, deren temporäre Spitzenlastverschiebung als Gruppe einen sehr hohen Einfluss auf den Spitzenlastbedarf auf Quartiersebene haben. Darüber hinaus bestätigen die Ergebnisse der Unsicherheitsanalysen auf Basis kurzfristiger Lastprognosen, welche in Abbildung 7.36 dargestellt sind, dass neben der Gruppe *LV* auch in den anderen Gruppen Potenziale für Lastverschiebungen vorhanden sind.

Im Hinblick auf die eingangs dieser Arbeit thematisierte Transformation urbaner Energiesysteme vor dem Hintergrund komplexer dynamischer Informationsgrundlagen lassen sich die vorangestellten Ergebnisse und die Vorteilhaftigkeit der Anwendung von *AMSA* in der Energiesystemplanung wie folgt exemplarisch zusammen fassen.

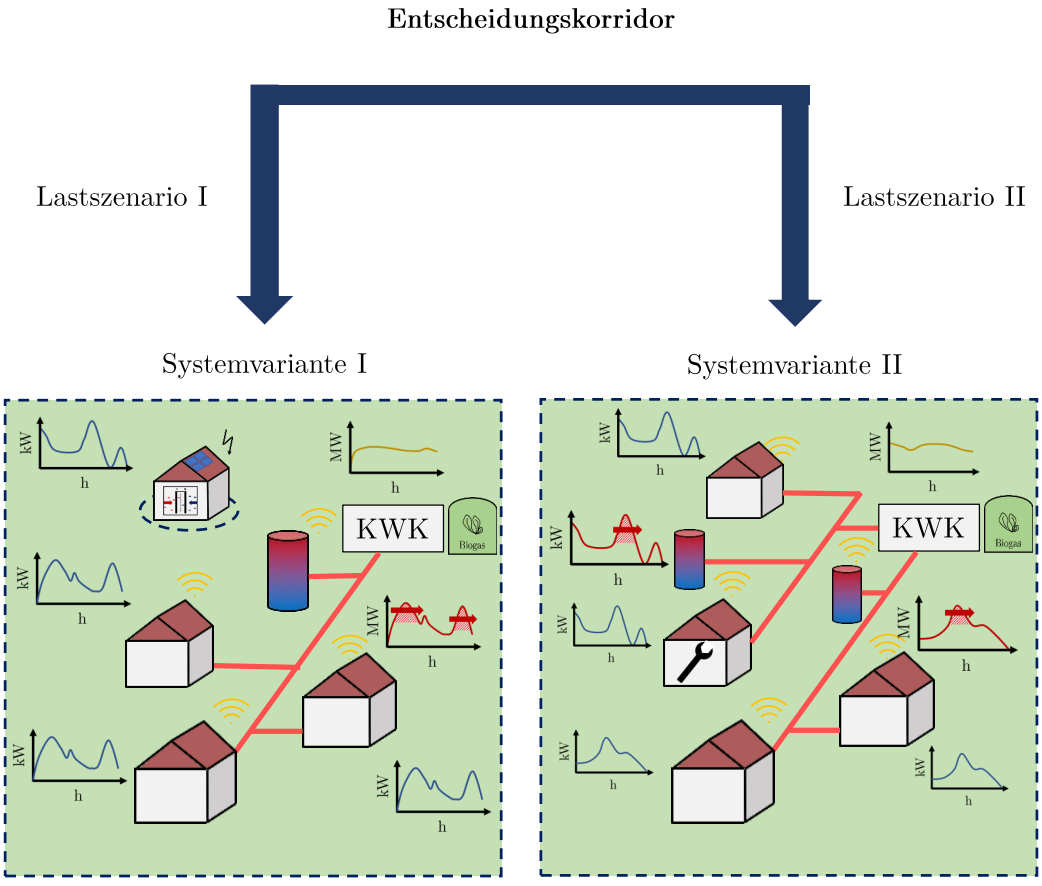


Abbildung 7.37: Exemplarische systemische Darstellung der Energiesystemvarianten infolge zukünftiger Lastprofilmodifikationen unter Berücksichtigung des szenarioabhängigen Lastverschiebungspotenzials

In Abbildung 7.37 sind dazu zwei verschiedene Energiesystemvarianten als Entscheidungskorridor auf Basis unterschiedlicher Lastszenarien vereinfacht dargestellt. Die Quantifizierung von Lastszenario I und die anschließende Ähnlichkeitsanalyse bezüglich der individuellen Lastprofilcharakteristika resultieren in der Empfehlung eines Zusammenschlusses von drei Gebäuden zu einem Subsystem mit einem zentralen Energieerzeugungs- und speicherungssystem. Infolge dessen lassen sich kontinuierlich Lastspitzen in diesem Subsystem verschieben, sodass der Energieerzeuger hauptsächlich eine Grundlast bereitstellt und somit der teure Betrieb des Spitzenlastkessels reduziert wird (Frederiksen et al., 2014). Lastszenario II berücksichtigt die Sanierung eines Gebäudes und beispielsweise modifizierte Wetterrahmenbedingungen. Dadurch ergibt sich eine abweichende Systemempfehlung mit zwei Subsystemen. Mittels eines jeweils installierten Speichers lassen sich die aggregierten Lastprofile der Subsysteme letzterer und somit auch das Gesamtlastprofil glätten.

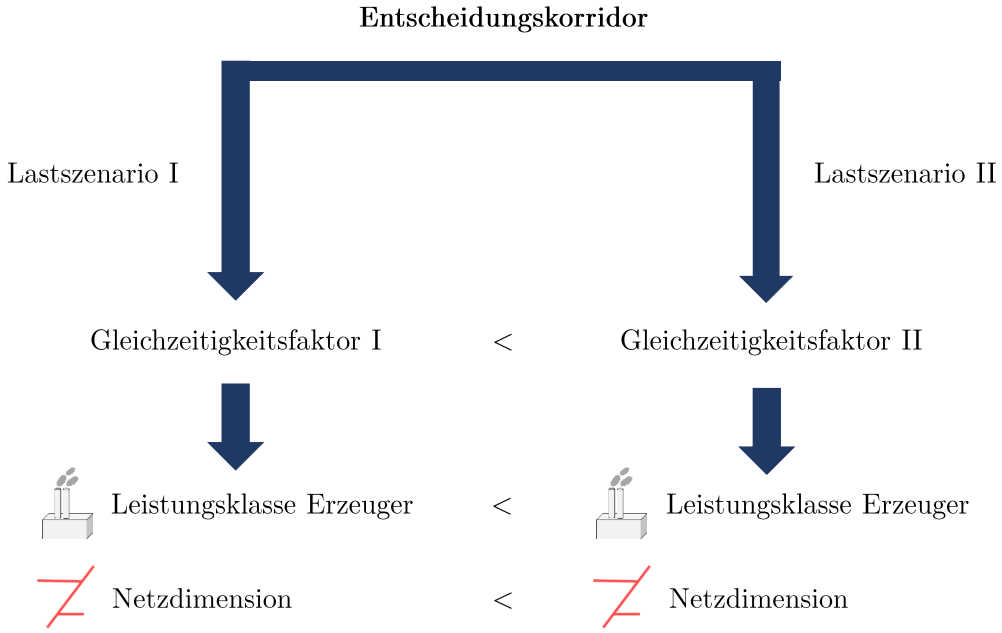


Abbildung 7.38: Exemplarische systemische Darstellung der Implikationen für die Dimensionierung zentraler Energieinfrastrukturen infolge der prädiktiven Lastszenarioanalysen und resultierender Gleichzeitigkeitsfaktoren

Der Gleichzeitigkeitsfaktor als weiterer Entscheidungsparameter für die ökonomisch und ökologisch optimierte Auslegung von zentralen Energieerzeugungsanlagen und Netzen ist ebenso abhängig von dem gewählten Lastszenario. Im Sinne der Planung und des Betriebs eines zukunftsfähigen optimierten Energiesystems ist dieser ein wesentlicher Bestandteil im Entscheidungskorridor, da sowohl die Wahl der Leistungsklasse eines Energieerzeugers als auch die Festlegung von Netzdimensionen von der Höhe abhängig ist. Abbildung 7.38 zeigt exemplarisch, dass auf Grundlage von Lastszenario I der Gleichzeitigkeitsfaktor

geringer als jener basierend auf Lastszenario II ist. Diese Analyse ist mit *AMSA* auch auf Subsysteme übertragbar, sodass ein differenzierteres Bild entsteht. Es wurde somit gezeigt, dass die Berücksichtigung von Gleichzeitigkeitseffekten in Lastverteilungen für die Wahl und den Betrieb von Energiesystemen einen Mehrwert darstellt. Gleichwohl sind für die optimierte Systemwahl auch weitere Analysen, beispielsweise in Bezug auf geeignete Energieerzeugerkombinationen, notwendig.

7.4 Zusammenfassung

Die Ausführungen dieses Kapitels zielten auf die Demonstration der Eignung von *AMSA* für die Adressierung des abgeleiteten Forschungsbedarfs zur Quantifizierung von Gleichzeitigkeitseffekten ab. Als Informationsgrundlage dienten Monitoringdaten des Forschungszentrums Jülich zum gebäudespezifischen Wärme- und Kälteverbrauch im Quartier. Zudem ergänzten Zeitreihen der Wettervariablen die historische Datengrundlage. Um Nutzungszeiten der Gebäude für Lastprognosen berücksichtigen zu können, wurden binär kodierte Anwesenheitsprofile in Abhängigkeit von den durchschnittlich angenommenen Arbeitszeiten erstellt, wobei explizit Wochenend- und Feiertage sowie Ferienzeiten Berücksichtigung fanden. In Ergänzung dazu dienten kurzfristige Wetterprognosen und TRJ des DWD sowie Kennwerte zum spezifischen Wärmeverbrauch von Gebäuden der Datenanreicherung. Ferner wurden Sanierungsreihenfolgen von Spinnräker et al. als Basis für die Modifizierung zukünftiger Lastprognosen im Rahmen von Szenarioanalysen in Bezug auf die Entwicklung des zukünftigen Energiebedarfs zur Abbildung disruptiver Ereignisse in der Planung urbaner Energiesysteme integriert.

Zur Demonstration der Modellkomponenten wurden eingangs die Ergebnisse der Datenvorverarbeitung hinsichtlich der Fehlerbereinigung und Normalisierung vorgestellt sowie insbesondere die Notwendigkeit der Ausreißereliminierung hervorgehoben. Anschließend erfolgte die komparative Analyse der Prognosemethoden des entwickelten Ensembles, wobei sich die NARX RNN gegenüber den *epsilon*-SVM-R Modellen als mehrheitlich überlegen gezeigt haben. Dennoch konnten in Einzelfällen mit der zweitgenannten Modellkategorie bessere Ergebnisse bezüglich der Prognosegüte erzielt werden. Die Resultate begründen an dieser Stelle die Vorteilhaftigkeit des Ensemble-Konzepts. Zudem wurde der Einfluss von Variationen in der Datengrundlage auf die automatisierte Modellwahl untersucht. In Bezug auf die Ergebnisse konnte gezeigt werden, dass die Häufigkeit der Wahl eines trainierten Modells für eine gebäudespezifische Prognoseaufgabe von der Zeitreihencharakteristik im Allgemeinen und der gewählten Methoden zur Datenvorverarbeitung im Speziellen abhängt. Diese Schlussfolgerung bedingt implizit die Entwicklung und Nutzung von autoadaptiven Modellen wie *AMSA* vor dem Hintergrund hochdynamischer Informationsquellen in intelligenten vernetzen Energiesystemen. Weiterhin wurden unterschiedliche Prognosezeiträume untersucht, wobei einerseits die Eignung von *AMSA* für kurzfristige Echtzeitprognosen und andererseits für die Generierung mittel- und langfristige Lastszenarien demonstriert wurde. In Bezug auf letztere erfolgte die Analyse des Einflusses von drei verschiedenen

Sanierungsreihenfolgen im betrachteten Quartier auf die Lastprognosen. Es resultierten dabei temporäre Lastverschiebungen und fluktuierende Leistungsspitzen über den Betrachtungszeitraum, welche sich auf die Gleichzeitigkeitsfaktoren auswirken. Ferner wurden Lastvariationen infolge der Verwendung von Extrem-Wetterdatensätzen aufgezeigt, welche Lastprofilmodifikationen hervorrufen und folglich einen ganzheitlichen Einfluss auf die Quantifizierung von Gleichzeitigkeitseffekten ausüben. Zusammenfassend lässt sich an dieser Stelle zudem festhalten, dass *AMSA* ebenso auf Einzelgebäude- und Quartiersebene, wie auch auf urbaner Ebene für die Generierung von Lastprognosen eingesetzt werden kann. Im Hinblick auf Demonstration der Modellkomponente zur automatisierten ähnlichkeitsbasierten Gebäudegruppierung wurde die Eignung von SOM für den unüberwachten Lernprozess ohne vorgegebene Clusteranzahl nachgewiesen. Im Ergebnis lassen sich gruppenspezifische aggregierte Standardlastprofile abbilden, welche sich untereinander in Abhängigkeit von den gewählten Dimensionen der SOM-Topologie mehr oder weniger stark unterscheiden. Für Analysezeiträume, die sich über mehr als ein Jahr erstrecken, ist das eingeführte Superpositionsprinzip zur Clusterüberlagerung als sinnvoll zu bewerten, um längerfristige Tendenzen der regelmäßigen gleichzeitigen Lastabnahme von Gebäuden zu identifizieren.

Im Rahmen der exemplarischen Quantifizierung von Gleichzeitigkeitseffekten auf Grundlage der Datenbasis im Demonstrationsquartier wurde gezeigt, dass mit *AMSA* für den Gleichzeitigkeitsfaktor im Vergleich zur Anwendung konventioneller Methoden differenziertere Ergebnisse erzielt werden können, welche für die Planung zukünftiger Energiesysteme unabdingbar sind. Weiterhin implizieren die Ergebnisse der Ähnlichkeitsanalysen die Möglichkeit der Identifizierung von geeigneten Versorgungsinseln, Gebäudegruppen mit dem größten Lastverschiebungspotenzial und Anhaltspunkte für geeignete Lastausgleichsstrategien. Hinsichtlich der operativen Ebene unterstützen die Gleichzeitigkeitsanalysen mit *AMSA* die prädiktive Betriebsführung, indem Gleichzeitigkeitsfaktoren die vorzuhaltenden Leistungskapazitäten regulieren und kurzfristige Ähnlichkeiten in den Lastprofilen der Gebäude für Lastverschiebungsstrategien im Sinne des Demand Side Managements genutzt werden können.

KAPITEL 8

Schlussbetrachtung

Das im Rahmen der vorliegenden Arbeit entwickelte autoadaptive prädiktive Modell *AMSA* adressiert den Forschungsbedarf zur Quantifizierung von Gleichzeitigkeitseffekten in Lastverteilungen urbaner Energiesysteme auf Grundlage hoch komplexer variierender Datengrundlagen. Während die Abschätzung des Gleichzeitigkeitsfaktors in der Planung zentraler Versorgungssysteme bislang auf Erfahrungswerten und statistischen Analysen historischer Lastgangdaten beruht, wird mit *AMSA* der Betrachtungswinkel um kurz-, mittel- und langfristige, szenariobasierte Prognosen zur Abbildung der dynamischen Entwicklung energetischer Laststrukturen erweitert. In der konventionellen Planung erfolgt die Berücksichtigung des Gleichzeitigkeitsfaktors vor dem Hintergrund der Reduktion anzusetzender Maximallastzustände für die Dimensionierung zentraler Energieerzeugungsanlagen und Netze infolge der Diversität des individuell auftretenden Spitzenleistungsbedarfs innerhalb von Gebäudegruppen. Mit dem entwickelten Modell dieser Arbeit wird der Begriff des Gleichzeitigkeitsfaktors erweitert, indem Unsicherheitsanalysen und prädiktive Szenarien integriert sowie Lastprofilcharakteristiken von Gebäuden im Hinblick auf Ähnlichkeiten automatisiert, analysiert und gruppiert werden. Dadurch lassen sich Entscheidungskorridore ableiten und Informationen zu Lastverschiebungspotenzialen und möglichen Versorgungsinseln generieren. Im weiteren Verlauf dieses Kapitels erfolgt in Abschnitt 8.1 die Zusammenfassung der wichtigsten Ergebnisse, wobei die in Kapitel 1 formulierten Hypothesen adressiert werden. Abschnitt 8.2 beinhaltet einen Ausblick zu Einsatzgebieten und Weiterentwicklungspotenzialen von *AMSA*.

8.1 Zusammenfassung der wichtigsten Ergebnisse

Vor dem Hintergrund der Herleitung des Forschungsbedarfs im Themenbereich der Gleichzeitigkeitsanalyse wurden zunächst Methoden und Modelle zur Ermittlung des thermischen Energiebedarfs als Grundlage für die Quantifizierung von Gleichzeitigkeitseffekten vorgestellt sowie Vor- und Nachteile verschiedener Ansätze diskutiert. Aufgrund der Informationsknappheit zu gebäudephysikalischen und nutzungsspezifischen Parametern auf urbaner Ebene sowie einer erhöhten Generierung von Messdaten infolge des vermehrten Einsatzes intelligenter vernetzter Energiesysteme, sind datenbasierte Methoden für den Anwendungszweck der vorliegenden Arbeit vorteilhaft. Überdies ergibt sich die Möglichkeit der Kombination mit dynamischen Gebäudesimulationen, wobei die resultierenden

Ergebnisse als Eingangsgrößen fungieren. Im Zuge der Analyse des derzeitigen Stands der Technik und Wissenschaft zur Gleichzeitigkeitsanalyse, zeigte sich, dass bisherige Quantifizierungen mehrheitlich auf Erfahrungswerten und empirischen Analysen beruhen. Vor dem Hintergrund komplexer variierender Datengrundlagen sowie Anforderungen an prädiktiv ausgerichtete nachhaltige Planungen von Energiesystemen greifen die bisherigen Ansätze zu kurz. Daraufhin wurde der Forschungsbedarf hinsichtlich selbstlernender Modelle zur Identifizierung komplexer Zusammenhänge zwischen gebäudeindividuellen Lastcharakteristika unter Berücksichtigung zukünftiger szenariobasierter Lastzustände und Unsicherheiten für die Nutzung von Gleichzeitigkeitseffekten abgeleitet.

Mit dem Ziel der Integration zukünftiger Lastzustände in die Gleichzeitigkeitsanalyse unter optionaler Berücksichtigung von Szenarien infolge externer Einflussfaktoren, wurden Prognosemethoden und Szenariotechniken analysiert. Als Grundlage zur Identifizierung geeigneter Ansätze dienten Aufarbeitungen der Stärken und Schwächen von Methodenkategorien sowie anwendungsbasierte Studien zur Evaluierung der Methodenauswahl für den Anwendungszweck in der vorliegenden Arbeit. Im Ergebnis zeigte sich, dass NARX RNN und ε-SVM-R Modelle anderen Methoden im Kontext der datenbasierten energetischen Lastprognosen überlegen sind. Zudem wurden im Rahmen einer Vorstudie beide Prognosemodelle miteinander verglichen. Im Ergebnis lieferten die NARX RNN überwiegend Prognosen mit höherer Genauigkeit. Dennoch ließen Einzelresultate den Rückschluss zu, dass gebäudespezifisch auch ε-SVM-R Modelle in Abhängigkeit von der individuellen Lastgangcharakteristik vorteilhaft sein können. Folglich eignet sich ein Ensemble-Modell aus beiden Methodenkategorien bestehend für automatisierte gebäudeindividuelle Lastprognosen auf Quartiersebene. Zur Identifizierung geeigneter Methoden für die numerische Abbildung von Szenarien, wurden verschiedene Konzepte vorgestellt und auf ihre jeweilige Eignung für den vorliegenden Anwendungsfall geprüft. Dabei folgt die Szenariomodellierung in dieser Arbeit der Idee der *PMT School*. In diesem Zusammenhang wird der Frage *Was-wäre-wenn* nachgegangen, indem Lastmodifikationen bei Gebäudesanierungen und klimatischen Veränderungen numerisch abgebildet werden.

Für die automatisierte Mustererkennung in unbekannten Datensätzen sind Clustering-Algorithmen geeignet, welche zur Gruppenbildung von energetischen Lastprofilen auf Basis charakteristischer Ähnlichkeiten in den Lastprofilen dienen. Das Ergebnis ist wiederum als Eingangsgröße und Trainingsgrundlage für Methoden der Klassifikation geeignet. Zur Identifizierung passender Algorithmen wurden Methodenkategorien sowie Stärken und Schwächen einzelner zugehöriger Methoden herausgearbeitet. Zusammen mit Forschungsergebnissen anwendungsbasierter Studien hat sich im ersten Schritt das Konzept der SOM als vorteilhaft für die Modellentwicklung herausgestellt. Dies ist insbesondere auf die Eigenschaft zurückzuführen, dass im Gegensatz zu anderen Clustering-Methoden kein *a priori* Wissen zur Clusteranzahl erforderlich ist. Zudem ist eine hohe Ergebnisgüte auf Grundlage der analysierten Forschungsarbeiten zu erwarten. Im zweiten Schritt wurden Klassifikationsmethoden analysiert, wobei sich das Konzept der LVQ aufgrund der intuitiven Kombinierbarkeit mit SOM als vorteilhaft erwiesen hat. Zudem beschränkt sich

der Anwendungsbereich nicht ausschließlich auf binäre Klassifikationsprobleme und im Vergleich zu SVM-Ansätzen werden Klassenrepräsentanten durch typische Vektoren an Stelle von Extremvektoren abgebildet. Diese Eigenschaften dokumentieren die explizite Eignung von LVQ für die Modellierung.

Die Entwicklung des autoadaptiven prädiktiven Modells basiert zum einen hinsichtlich der Zielvariablen auf dem hergeleiteten Forschungsbedarf im Themenfeld der Gleichzeitigkeitsanalyse, zum anderen auf den Implikationen für die Modellierung, welche in den Gebieten der Prognosemodelle und Szenariotechniken sowie Clustering und Klassifikation energetischer Lastprofile zur Adressierung der Fragestellungen dieser Arbeit abgeleitet wurden. In Bezug auf den Gleichzeitigkeitsfaktor wurde die konventionelle Definition erweitert sowie ein Ähnlichkeitsfaktor in Verbindung mit Spitzenlastanteilen von Gebäudegruppen als numerische Zielvariablen eingeführt. Im weiteren Verlauf wurden die mathematischen Grundlagen und gewählten Parameter für die Modellentwicklung vorgestellt und erläutert.

AMSA wurde in Matlab auf Grundlage einer mehrteiligen, hierarchisch aufgebauten Klassenarchitektur implementiert, die es erlaubt, Erweiterungen im Softwarekonzept vorzunehmen, ohne den numerischen Rechenkern modifizieren zu müssen. Dadurch ist eine Integration in bestehende Energie-Informationssysteme möglich. Die Hauptelemente von *AMSA* umfassen zwei grafische Benutzeroberflächen zur Steuerung des Programms und der interpretierbaren Aufbereitung von Ergebnissen in Form von Visualisierungen, grafischen Darstellungen und Kennzahlen. Somit lässt sich die Komplexität von analysierten Datengrundlagen im Rahmen eines Prozesses der Informationsextraktion auf Kernparameter für die Dimensionierung, Parzellierung und den Betrieb von urbanen Energiesystemen reduzieren.

Zur Demonstration des entwickelten Modells wurden Messdaten zu gebäudespezifischen Energieverbräuchen des Forschungszentrums Jülich, ergänzt durch Daten des DWD und VDI, verwendet sowie mittels der vorgestellten Methoden zur Datenvorverarbeitung vorbereitet. Auf Grundlage der Untersuchungsergebnisse infolge der Anwendung des Prognose-Ensembles wurde die Eignung zur hinreichend präzisen Simulation zukünftiger thermischer Lasten von Gebäuden sowie zur numerischen Abbildung des kurz-, mittel- und langfristigen Energiebedarfs von Gebäuden und Quartieren nachgewiesen. Weiterhin haben Analysen zur langfristigen Entwicklung der thermischen Leistungsabnahme auf Quartiersebene unter Berücksichtigung verschiedener Sanierungsreihenfolgen und Wetterrandbedingungen im Rahmen dieser Arbeit gezeigt, dass in diesem Kontext Lastvariationen, speziell im Bereich der Leistungsspitzen, zu erwarten und folglich für die Quantifizierung von Gleichzeitigkeitseffekten zu berücksichtigen sind. Zur automatisierten Gruppierung von gebäudespezifischen thermischen Lastgängen wurde die Modellkombination aus SOM und LVQ evaluiert. Auf Basis der Ergebnisse exemplarischer Anwendungen eignet sich diese für das vorliegende Einsatzgebiet. In Bezug auf die exemplarischen Quantifizierungen von Gleichzeitigkeitseffekten für das Demonstrationsquartier wurden die Vorteile der eingeführten Definitionserweiterung nachgewiesen. Die Ermittlung des Gleichzeitigkeitsfaktors mit *AMSA* unter Berücksichtigung zukünftiger Lastzustände und Unsicherheit bietet dabei

abgegrenzte Entscheidungskorridore, die im Gegensatz zu konventionell angesetzten Kennwerten eine präzisere und zukunftsfähige Dimensionierung von Energieversorgungssystemen ermöglichen. Darüber hinaus wurde gezeigt, dass sich Gebäudegruppierungen auf Grundlage ähnlicher charakteristischer Lastprofile für die Identifizierung und Lokalisierung von Lastverschiebungspotenzial für einen optimierten Betrieb von Energieversorgungssystemen eignet, wobei in Abhängigkeit vom gewählten Detailgrad des Clusterings Gebäude, kleinere oder größere Gebäudegruppen, ausgemacht werden können. Hinsichtlich längerfristiger Planungen zur Quartiersentwicklung zeigte sich, dass letztgenannte Effekte als Entscheidungshilfe für die Festlegung von Versorgungsinseln dienen können. Ergebnisübergreifend ergibt sich ferner eine breite Anwendbarkeit von *AMSA* aufgrund der demonstrierten autoadaptiven Eigenschaften.

Vor dem Hintergrund der Schlussbetrachtung zur vorliegenden Arbeit haben sich die eingangs gestellten Hypothesen bestätigt.

8.2 Ausblick

In Anlehnung an die zusammenfassenden Ausführungen im vorangestellten Abschnitt ist festzustellen, dass sich *AMSA* für die detaillierte Quantifizierung von Gleichzeitigkeitseffekten eignet. Hinsichtlich zukünftiger Herausforderungen in der Energiewende adressiert das entwickelte Modell Indikatoren für Lösungsansätze der folgenden Herausforderungen: In Zukunft ist die Etablierung smarter Energiesysteme mit kleinteiligen dezentralen Insellösungen vermehrt zu erwarten. Die Integration intelligenter Zählersysteme bedingt hochdynamisch wachsende Informationsgrundlagen, die es im Sinne einer kurzfristigen Betriebs- und einer längerfristigen szenariobasierten Dimensionierungsplanung zu analysieren gilt, um wirtschaftlich und ökologisch optimierte Strategien abzuleiten. Darüber hinaus implizieren strukturelle Veränderungen im Gebäudebestand infolge wachsender Anforderungen an Energiestandards, Neu- und Rückbauvorhaben innerhalb urbaner Regionen Veränderungen der bestehenden Energieversorgungsstrukturen. Zusammen mit veränderten klimatischen Randbedingungen ergeben sich modifizierte Lastverteilungen, die es für eine nachhaltige Auslegung von Energiesystemen zu berücksichtigen gilt. Um Ressourcen möglichst effizient nutzen zu können, sind bei der Planung zukünftiger Systeme verstärkt Gleichzeitigkeitsfaktoren zur Reduzierung von Erzeugerleistungen und Netzverlusten zu berücksichtigen. Darüber hinaus bieten ähnlichkeitsbasierte Subgruppenbildungen in der zentralen Energieversorgung die Möglichkeit, Planungen im Sinne des Demand-Side-Managements anzupassen und Speichersysteme effizient einsetzen und verorten zu können. Die genannten Aspekte bilden eine breite Grundlage für zukünftige Anwendungsfelder von *AMSA*, welche für die Validierung des Modells genutzt werden sollten.

In Bezug auf aktuelle Forschungsarbeiten bieten sich darüber hinaus verschiedene explizite Anknüpfungspunkte für *AMSA*. Neben einer Vergrößerung der Integrationstiefe in das bestehende Energie-Informationssystem für urbane Regionen, welches in Kapitel 6 vorge-

stellt wurde, ist eine Schnittstelle zu GeTIS (Fichter et al., 2017; Weck-Ponten et al., 2018) anzustreben, um Lastprognosen und Gleichzeitigkeitseffekte auch für die optimierte Auslegung von geothermiebasierten Systemen und kalten Nahwärmenetzen nutzen zu können. In Kapitel 7 wurde ferner im Zuge der Auswertung von Wärme- und Kühllastprognosen der Einfluss des Nutzerverhaltens auf die Ergebnisse hervorgehoben. In diesem Zusammenhang könnte eine Kopplung von *AMSA* mit Vorhersagen zum Nutzerverhalten auf Basis von numerischen Menschmodellen wie *MORPHEUS* (Wölki, 2017) eine Verbesserung der Lastprognosen bewirken. Insgesamt ergeben sich durch die im Rahmen des siebten Energieforschungsprogramms eingeführten *Reallabore* (van Treeck, 2019) vielfältige Forschungsplattformen, auf denen das entwickelte Modell der vorliegenden Arbeit eingesetzt und beispielsweise vor dem Hintergrund der Anwendung im Stromsektor weiterentwickelt werden kann.

Literatur

Agrawal, Rakesh, Johannes Gehrke, Dimitrios Gunopulos & Prabhakar Raghavan (1998): „Automatic subspace clustering of high dimensional data for data mining applications“. *the 1998 ACM SIGMOD international conference*. Hrsg. von Haas, Laura, Pamela Drew, Ashutosh Tiwary & Michael Franklin: S. 94–105 (siehe S. 50).

Al-Hamadi, H. M. & S. A. Soliman (2005): „Long-term/mid-term electric load forecasting based on short-term correlation and annual growth“. *Electric Power Systems Research*, Bd. 74(3): S. 353–361 (siehe S. 41).

Al-Homoud, Mohammad Saad (2001): „Computer-aided building energy analysis techniques“. Bd. (siehe S. 12, 13, 35, 114).

Amasyali, Kadir & Nora M. El-Gohary (2018): „A review of data-driven building energy consumption prediction studies“. *Renewable and Sustainable Energy Reviews*, Bd. 81: S. 1192–1205 (siehe S. 30, 32, 34–36, 88, 210).

Amer, Muhammad, Tugrul U. Daim & Antonie Jetter (2013): „A review of scenario planning“. *Futures*, Bd. 46: S. 23–40 (siehe S. 37–39).

Ankerst, Mihael, Markus M Breunig, Hans-Peter Kriegel & Jörg Sander (1999): „OPTICS: ordering points to identify the clustering structure“. *ACM Sigmod record*. Bd. 28. 2. ACM: S. 49–60 (siehe S. 49).

Asare-Bediako, B., W. L. Kling & P. F. Ribeiro (2014): „Future residential load profiles: Scenario-based analysis of high penetration of heavy loads and distributed generation“. *Energy and Buildings*, Bd. 75: S. 228–238 (siehe S. 40).

Bauer, M. & J.-L. Scartezzini (1998): „A simplified correlation method accounting for heating and cooling loads in energy-efficient buildings“. *Energy and Buildings*, Bd. 27(2): S. 147–154 (siehe S. 14, 15).

Bhavsar, Hetal & Amit Ganatra (2012): „A comparative study of training algorithms for supervised machine learning“. *International Journal of Soft Computing and Engineering (IJSCE)*, Bd. 2(4): S. 2231–2307 (siehe S. 55).

Bishop, C. M. (2006): *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer (siehe S. 30, 45, 57, 58, 74–77, 80–82, 84, 86, 133).

Bishop, Peter, Andy Hines & Terry Collins (2007): „The current state of scenario development: an overview of techniques“. *foresight*, Bd. 9(1): S. 5–25 (siehe S. 39).

Björsell, Niclas, Axel Bring, Lars Eriksson, Pavel Grozman, Magnus Lindgren, Per Sahlin, Alexander Shapovalov & Mika Vuolle (1999): „IDA indoor climate and energy“. *Proc. of the 6-th IBPSA Conference*: S. 1035–1042 (siehe S. 11).

Blesl, M. (2002): *Räumlich hoch aufgelöste Modellierung leitungsgebundener Energieversorgungssysteme zur Deckung des Niedertemperaturwärmebedarfs*. Universität Stuttgart, Institut für Energiewirtschaft und Rationelle Energieanwendung. IER, Bibliothek (siehe S. 24, 25).

Bobric, Elena C., Gheorghe Cartina & Gheorghe Grigoras (2009): „Clustering techniques in load profile analysis for distribution stations“. *Advances in electrical and computer engineering*, Bd. 9(1): S. 63–66 (siehe S. 52, 53).

Börjeson, Lena, Mattias Höjer, Karl-Henrik Dreborg, Tomas Ekvall & Göran Finnveden (2006): „Scenario types and techniques: Towards a user's guide“. *Futures*, Bd. 38(7): S. 723–739 (siehe S. 39).

Boser, Bernhard E., Isabelle M. Guyon & Vladimir N. Vapnik (1992): „A training algorithm for optimal margin classifiers“. *the fifth annual workshop*. Hrsg. von Haussler, David: S. 144–152 (siehe S. 86).

Bradfield, Ron, George Wright, George Burt, George Cairns & Kees van der Heijden (2005): „The origins and evolution of scenario techniques in long range business planning“. *Futures*, Bd. 37(8): S. 795–812 (siehe S. 37, 38).

Bradley, Elizabeth & Holger Kantz (2015): „Nonlinear time-series analysis revisited“. *Chaos (Woodbury, N.Y.)*, Bd. 25(9): S. 097610 (siehe S. 73).

Breiman, Leo (1996): „Bagging predictors“. *Machine learning*, Bd. 24(2): S. 123–140 (siehe S. 58).

Bueno, Bruno, Leslie Norford, Grégoire Pigeon & Rex Britter (2012): „A resistance-capacitance network model for the analysis of the interactions between the energy performance of buildings and the urban climate“. *Building and Environment*, Bd. 54: S. 116–125 (siehe S. 15).

Bunte, Kerstin, Barbara Hammer, Thomas Villmann, Michael Biehl & Axel Wismüller (2010): „Exploratory observation machine (xom) with kullback-leibler divergence for dimensionality reduction and visualization“. *ESANN*. Bd. 10: S. 87–92 (siehe S. 50).

Burhenne, Sebastian (2013): *Monte Carlo based uncertainty and sensitivity analysis for building performance simulation: Zugl.: Karlsruhe, KIT, Diss., 2013*. Schriftenreihe der Reiner Lemoine-Stiftung. Aachen: Shaker (siehe S. 216).

Burt, George (2007): „Why are we surprised at surprises? Integrating disruption theory and system analysis with the scenario methodology to help identify disruptions and discontinuities“. *Technological Forecasting and Social Change*, Bd. 74(6): S. 731–749 (siehe S. 38).

Cartalis, C., A. Synodinou, M. Proedrou, A. Tsangrassoulis & M. Santamouris (2001): „Modifications in energy demand in urban areas as a result of climate changes: An assessment for the southeast Mediterranean region“. *Energy Conversion and Management*, Bd. 42(14): S. 1647–1656 (siehe S. 40).

Caruana, Rich, Steve Lawrence & C. Lee Giles (2001): „Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping“. *Advances in neural information processing systems*: S. 402–408 (siehe S. 80).

Catalina, Tiberiu, Vlad Iordache & Bogdan Caracaleanu (2013): „Multiple regression model for fast prediction of the heating energy demand“. *Energy and Buildings*, Bd. 57: S. 302–312 (siehe S. 35).

Chaouch, Mohamed (2014): „Clustering-Based Improvement of Nonparametric Functional Time Series Forecasting: Application to Intra-Day Household-Level Load Curves“. *IEEE Transactions on Smart Grid*, Bd. 5(1): S. 411–419 (siehe S. 52).

Chen, S., S. A. Billings & P. M. Grant (1990): „Non-linear system identification using neural networks“. *International Journal of Control*, Bd. 51(6): S. 1191–1214 (siehe S. 36).

Cherkassky, Vladimir & Yunqian Ma (2004): „Practical selection of SVM parameters and noise estimation for SVM regression“. *Neural Networks*, Bd. 17(1): S. 113–126 (siehe S. 83).

Chermack, Thomas J., Susan A. Lynham & Wendy E. A. Ruona (2001): „A review of scenario planning literature“. *Futures Research Quarterly*, Bd. 17(2): S. 7–32 (siehe S. 38).

Chicco, G., R. Napoli & F. Piglione (2006): „Comparisons Among Clustering Techniques for Electricity Customer Classification“. *IEEE Transactions on Power Systems*, Bd. 21(2): S. 933–940 (siehe S. 48, 49, 53).

– (2001): „Load pattern clustering for short-term load forecasting of anomalous days“. *2001 Power Tech*: S. 6 (siehe S. 52, 53).

Chicco, G., R. Napoli, F. Piglione, P. Postolache, M. Scutariu & C. Toader (2005): „Emergent electricity customer classification“. *IEE Proceedings - Generation, Transmission and Distribution*, Bd. 152(2): S. 164 (siehe S. 53).

– (2004): „Load Pattern-Based Classification of Electricity Customers“. *IEEE Transactions on Power Systems*, Bd. 19(2): S. 1232–1239 (siehe S. 53).

Chicco, Gianfranco (2012): „Overview and performance assessment of the clustering methods for electrical load pattern grouping“. *Energy*, Bd. 42(1): S. 68–80 (siehe S. 53).

Cho, Sung-Hwan, Won-Tae Kim, Choon-Soeb Tae & M. Zaheeruddin (2004): „Effect of length of measurement period on accuracy of predicted annual heating energy consumption of buildings“. *Energy Conversion and Management*, Bd. 45(18-19): S. 2867–2878 (siehe S. 15).

Christenson, M., H. Manz & D. Gyalistras (2006): „Climate warming impact on degree-days and building energy demand in Switzerland“. *Energy Conversion and Management*, Bd. 47(6): S. 671–686 (siehe S. 2).

Clarke, Stella M., Jan H. Griebsch & Timothy W. Simpson (2005): „Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses“. *Journal of the American Statistical Association*, Bd. 127(6): S. 1077 (siehe S. 84, 88).

Cleff, Thomas (2015): *Deskriptive Statistik und Explorative Datenanalyse: Eine computergestützte Einführung mit Excel, SPSS und STATA*. Wiesbaden: Gabler Verlag (siehe S. 69).

Coakley, Daniel, Paul Raftery & Marcus Keane (2014): „A review of methods to match building energy simulation models to measured data“. *Renewable and Sustainable Energy Reviews*, Bd. 37: S. 123–141 (siehe S. 10, 13–15).

Connolly, D., H. Lund, B. V. Mathiesen, S. Werner, B. Möller, U. Persson, T. Boermans, D. Trier, P. A. Østergaard & S. Nielsen (2014): „Heat Roadmap Europe: Combining district heating with heat savings to decarbonise the EU energy system“. *Energy Policy*, Bd. 65: S. 475–489 (siehe S. 1).

Connolly, David, H. Lund & B. V. Mathiesen (2016): „Smart Energy Europe: The technical and economic impact of one potential 100% renewable energy scenario for the European Union“. *Renewable and Sustainable Energy Reviews*, Bd. 60: S. 1634–1653 (siehe S. 40).

Connor, Jerome, Les E. Atlas & Douglas R. Martin (1992): „Recurrent networks and NARMA modeling“. *Advances in Neural Information Processing Systems*: S. 301–308 (siehe S. 36).

Crawley, Drury B., Linda K. Lawrie, Frederick C. Winkelmann, W. F. Buhl, Y.Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte & Jason Glazer (2001): „EnergyPlus: Creating a new-generation building energy simulation program“. *Energy and Buildings*, Bd. 33(4): S. 319–331 (siehe S. 11).

Dan Foresee, F. & M. T. Hagan (1997): „Gauss-Newton approximation to Bayesian learning“. *International Conference on Neural Networks (ICNN'97)*: S. 1930–1935 (siehe S. 80).

Dankert, Jürgen & Helga Dankert (2011): *Technische Mechanik: Statik, Festigkeitslehre, Kinematik/Kinetik*. 6., überarbeitete Auflage. Wiesbaden: Vieweg+Teubner Verlag / Springer Fachmedien Wiesbaden GmbH Wiesbaden (siehe S. 216, 217).

David C. Hoaglin (2003): „John W. Tukey and Data Analysis“. *Statistical Science*, Bd. 18(3): S. 311–318 (siehe S. 68, 69).

Deutscher Wetterdienst & Bundesinstitut für Bau-, Stadt- und Raumforschung (2017): *Handbuch - Ortsgenaue Testreferenzjahre von Deutschland für mittlere, extreme und zukünftige Witterungsverhältnisse*. Offenbach (siehe S. 120, 121).

Dhar, A., T. A. Reddy & D. E. Claridge (1999): „A Fourier Series Model to Predict Hourly Heating and Cooling Energy Use in Commercial Buildings With Outdoor Temperature as the Only Weather Variable“. *Journal of Solar Energy Engineering*, Bd. 121(1): S. 47 (siehe S. 15).

– (1998): „Modeling Hourly Energy Use in Commercial Buildings With Fourier Series Functional Forms“. *Journal of Solar Energy Engineering*, Bd. 120(3): S. 217 (siehe S. 15).

Diaconescu, Eugen (2008): „The use of NARX neural networks to predict chaotic time series“. *Wseas Transactions on computer research*, Bd. 3(3): S. 182–191 (siehe S. 36, 73).

DIN 12831 Beiblatt 2 (2012): *Heizungsanlagen in Gebäuden –Verfahren zur Berechnung der Norm-Heizlast –Beiblatt 2: Vereinfachtes Verfahren zur Ermittlung der Gebäude-Heizlastund der Wärmeerzeugerleistung*. Berlin (siehe S. 13, 14, 212).

DIN 1988-300 (2012): *Technische Regeln für Trinkwasser-Installationen - Teil 300: Ermittlung der Rohrdurchmesser; Technische Regel des DVGW* (siehe S. 18).

DIN EN 12831-1 (2017): *Energetische Bewertung von Gebäuden –Verfahren zur Berechnung der Norm-Heizlast –Teil 1: Raumheizlast, Modul M3-3;Deutsche Fassung EN 12831-1:2017*. Berlin (siehe S. 11).

DIN EN 806-3 (2006): *Technische Regeln für Trinkwasser-Installationen –Teil 3: Berechnung der Rohrinnendurchmesser –Vereinfachtes Verfahren;Deutsche Fassung EN 806-3:2006* (siehe S. 18).

DIN V 18599-1 (2018): *Energetische Bewertung von Gebäuden –Berechnung des Nutz-, End- und Primärenergiebedarfs für Heizung,Kühlung, Lüftung, Trinkwarmwasser und Beleuchtung –Teil 1: Allgemeine Bilanzierungsverfahren, Begriffe, Zonierung undBewertung der Energieträger*. Berlin (siehe S. 12).

Dotzauer, Erik (2002): „Simple model for prediction of loads in district-heating systems“. *Applied Energy*, Bd. 73(3-4): S. 277–284 (siehe S. 30, 31, 35).

Drucker, Harris, Christopher J. C. Burges, Linda Kaufman, Alex J. Smola & Vladimir Vapnik (1997): „Support vector regression machines“. *Advances in neural information processing systems*: S. 155–161 (siehe S. 81–85).

Duan, Pan, Kaigui Xie, Tingting Guo & Xiaogang Huang (2011): „Short-Term Load Forecasting for Electric Power Systems Using the PSO-SVR and FCM Clustering Techniques“. *Energies*, Bd. 4(1): S. 173–184 (siehe S. 52, 53).

Dunn, J. C. (1973): „A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters“. *Journal of Cybernetics*, Bd. 3(3): S. 32–57 (siehe S. 52).

Edwards, Richard E., Joshua New & Lynne E. Parker (2012): „Predicting future hourly residential electrical consumption: A machine learning case study“. *Energy and Buildings*, Bd. 49: S. 591–603 (siehe S. 32).

Enami, Neda, Reza Askari Moghadam, Kourosh Dadashtabar & Mojtaba Hoseini (2010): „Neural network based energy efficiency in wireless sensor networks: A survey“. *International Journal of Computer Science & Engineering Survey*, Bd. 1(1): S. 39–53 (siehe S. 51).

Enzer, Selwyn (1981): „Exploring long-term business climates and strategies with interax“. *Futures*, Bd. 13(6): S. 468–482 (siehe S. 38).

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu et al. (1996): „A density-based algorithm for discovering clusters in large spatial databases with noise“. *Kdd*. Bd. 96: S. 226–231 (siehe S. 49, 50).

Estivill-Castro, Vladimir & Jianhua Yang (2000): „Fast and Robust General Purpose Clustering Algorithms“. *PRICAI 2000 Topics in Artificial Intelligence*. Hrsg. von Goos, G., J. Hartmanis, J. van Leeuwen, Riichiro Mizoguchi & John Slaney. Bd. 1886. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg: S. 208–218 (siehe S. 47).

Fahrmeir, Ludwig, Christian Heumann, Rita Künstler, Iris Pigeot & Gerhard Tutz (2016): *Statistik: Der weg zur datenanalyse*. Springer-Verlag (siehe S. 210, 211).

Fan, J. Y. & J. D. McDonald (1994): „A real-time implementation of short-term load forecasting for distribution power systems“. *IEEE Transactions on Power Systems*, Bd. 9(2): S. 988–994 (siehe S. 31, 35).

Farid, Dewan Md., Li Zhang, Chowdhury Mofizur Rahman, M. A. Hossain & Rebecca Strachan (2014): „Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks“. *Expert Systems with Applications*, Bd. 41(4): S. 1937–1946 (siehe S. 59).

Fichter, Eric, Sebastian Weck, Ralf Becker, Jan Derksen, Stephan Düber, Jérôme Frisch, David Koppmann, Robert Albertus Petrus Löhring, Jörg Blankenbach, Christoph Alban van Treeck & Martin Ziegler (2017): „Geothermal Information System for Potential Studies in Subsurface Soil Layers“. *[Proceedings of Building Simulation 2017: 15th Conference of IBPSA, BS2017, 2017-08-07 - 2017-08-09, San Francisco, USA / Editors: Charles S. Barnaby and Michael Wetter]*: S. 662–671 (siehe S. 173).

Figueiredo, V., F. Rodrigues, Z. Vale & J. B. Gouveia (2005): „An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques“. *IEEE Transactions on Power Systems*, Bd. 20(2): S. 596–602 (siehe S. 59).

Frederiksen, Svend & Sven Werner (2014): *District heating and cooling*. Lund: Studentlitteratur (siehe S. 2, 17, 22, 155, 165, 235).

Freund, Yoav & Robert E. Schapire (1997): „A decision-theoretic generalization of on-line learning and an application to boosting“. *Journal of computer and system sciences*, Bd. 55(1): S. 119–139 (siehe S. 58).

– (1999): „Large margin classification using the perceptron algorithm“. *Machine learning*, Bd. 37(3): S. 277–296 (siehe S. 55).

Friedman, Jerome H. (1989): „Regularized discriminant analysis“. *Journal of the American Statistical Association*, Bd. 84(405): S. 165–175 (siehe S. 57).

Gaderer, Matthias (2008): „Wärmeversorgung mit fester Biomasse bei kleiner Leistung“. Diss. Technische Universität München (siehe S. 20, 26, 214, 215).

Gao, Yang, Emmanuel Tumwesigye, Brian Cahill & Karsten Menzel (2010): „Using data mining in optimisation of building energy consumption and thermal comfort management“. *The 2nd International Conference on Software Engineering and Data Mining*: S. 434–439 (siehe S. 60).

Garland, Joshua & Elizabeth Bradley (2015): „Prediction in projection“. *Chaos (Woodbury, N.Y.)*, Bd. 25(12): S. 123108 (siehe S. 73).

Geysen, Davy, Oscar de Somer, Christian Johansson, Jens Brage & Dirk Vanhoudt (2018): „Operational thermal load forecasting in district heating networks using machine learning and expert advice“. *Energy and Buildings*, Bd. 162: S. 144–153 (siehe S. 35).

Glück, Bernd (1985): *Heizwassernetze für Wohn- und Industriegebiete*. Frankfurt (Main): Verlags- und Wirtschaftsgesellschaft der Elektrizitätswerke (siehe S. 18).

Gordon, Theodore Jay (1994): „Trend impact analysis“. *Futures Research Methodology*, Bd.: S. 1–9 (siehe S. 38).

Granell, Ramon, Colin J. Axon & David C. H. Wallom (2015): „Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles“. *IEEE Transactions on Power Systems*, Bd. 30(6): S. 3217–3224 (siehe S. 53).

Grubler, Arnulf, Xuemei Bai, Thomas Buettner, Shobhakar Dhakal, David J. Fisk, Toshiaki Ichinose, James E. Keirstead, G. Sammmer, David Satterthwaite, Niels B. Schulz et al. (2012): „Urban energy systems“. Bd. (siehe S. 40).

Guan, Jun, Natasa Nord & Shuqin Chen (2016): „Energy planning of university campus building complex: Energy usage and coincidental analysis of individual buildings with a case study“. *Energy and Buildings*, Bd. 124: S. 99–111 (siehe S. 2, 22, 26, 66, 67).

Guerra Santin, Olivia, Laure Itard & Henk Visscher (2009): „The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock“. *Energy and Buildings*, Bd. 41(11): S. 1223–1232 (siehe S. 2).

Guha, Sudipto, Rajeev Rastogi & Kyuseok Shim (1998): „CURE“. *the 1998 ACM SIGMOD international conference*. Hrsg. von Haas, Laura, Pamela Drew, Ashutosh Tiwary & Michael Franklin: S. 73–84 (siehe S. 48).

Gupta, Manish, Jing Gao, Charu C. Aggarwal & Jiawei Han (2014): „Outlier Detection for Temporal Data: A Survey“. *IEEE Transactions on Knowledge and Data Engineering*, Bd. 26(9): S. 2250–2267 (siehe S. 68).

Guyon, Isabelle & André Elisseeff (2003): „An introduction to variable and feature selection“. *Journal of machine learning research*, Bd. 3(Mar): S. 1157–1182 (siehe S. 121).

Hagan, M. T. & M. B. Menhaj (1994): „Training feedforward networks with the Marquardt algorithm“. *IEEE transactions on neural networks*, Bd. 5(6): S. 989–993 (siehe S. 78).

Haida, T. & S. Muto (1994): „Regression based peak load forecasting using a transformation technique“. *IEEE Transactions on Power Systems*, Bd. 9(4): S. 1788–1794 (siehe S. 31).

Hammer, Barbara, Marc Strickert & Thomas Villmann (2004): „Relevance LVQ versus SVM“. *Artificial Intelligence and Soft Computing - ICAISC 2004*. Hrsg. von Hutchison, David et al. Bd. 3070. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg: S. 592–597 (siehe S. 58).

Han, Jiawei, Micheline Kamber & Jian Pei (2012): *Data mining: Concepts and techniques, third edition*. 3rd ed. The Morgan Kaufmann series in data management systems. Waltham, Mass.: Morgan Kaufmann Publishers (siehe S. 32, 33, 45–52, 54, 56, 58, 59, 68, 70, 83, 98, 133, 210, 211).

Hartigan, John A. (1975): „Clustering algorithms“. Bd. (siehe S. 45).

Haykin, Simon (op. 1999): *Neural networks: A comprehensive foundation*. 2nd ed. London: Prentice Hall International (siehe S. 73, 75).

Haykin, Simon S. (2009): *Neural networks and learning machines*. 3. ed. New York: Pearson (siehe S. 82, 86).

Hinneburg, Alexander, Daniel A. Keim et al. (1998): „An efficient approach to clustering in large multimedia databases with noise“. *KDD*. Bd. 98: S. 58–65 (siehe S. 49).

Hippert, H. S., C. E. Pedreira & R. C. Souza (2001): „Neural networks for short-term load forecasting: A review and evaluation“. *IEEE Transactions on Power Systems*, Bd. 16(1): S. 44–55 (siehe S. 30).

Hodge, Victoria & Jim Austin (2004): „A Survey of Outlier Detection Methodologies“. *Artificial Intelligence Review*, Bd. 22(2): S. 85–126 (siehe S. 68).

Hofmann, Daniela & Barbara Hammer (2012): „Kernel Robust Soft Learning Vector Quantization“. *Artificial Neural Networks in Pattern Recognition*. Hrsg. von Hutchison,

David et al. Bd. 7477. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg: S. 14–23 (siehe S. 58).

Hong, Tao, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli & Rob J. Hyndman (2016): „Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond“. *International Journal of Forecasting*, Bd. 32(3): S. 896–913 (siehe S. 40).

Hong, Tao, Jason Wilson & Jingrui Xie (2014): „Long Term Probabilistic Load Forecasting and Normalization With Hourly Information“. *IEEE Transactions on Smart Grid*, Bd. 5(1): S. 456–462 (siehe S. 40).

Howard, B., L. Parshall, J. Thompson, S. Hammer, J. Dickinson & V. Modi (2012): „Spatial distribution of urban building energy consumption by end use“. *Energy and Buildings*, Bd. 45: S. 141–151 (siehe S. 9).

Howlett, Robert J. & Lakhmi C. Jain (2013): *Radial basis function networks 2: new advances in design*. Bd. 67. Physica (siehe S. 56).

Hsu, Chih-Wei & Chih-Jen Lin (2002): „A comparison of methods for multiclass support vector machines“. *IEEE transactions on neural networks*, Bd. 13(2): S. 415–425 (siehe S. 57).

Huang, J. P., K. L. Poh & B. W. Ang (1995): „Decision analysis in energy and environmental modeling“. *Energy*, Bd. 20(9): S. 843–855 (siehe S. 38).

Huang, S. R. (1997): „Short-term load forecasting using threshold autoregressive models“. *IEE Proceedings - Generation, Transmission and Distribution*, Bd. 144(5): S. 477 (siehe S. 31).

Huang, Zhexue (1998): „Extensions to the k-means algorithm for clustering large data sets with categorical values“. *Data mining and knowledge discovery*, Bd. 2(3): S. 283–304 (siehe S. 47).

Huss, William R. & E. J. Honton (1987a): „Alternative methods for developing business scenarios“. *Technological Forecasting and Social Change*, Bd. 31(3): S. 219–238 (siehe S. 38).

Huss, William R. & Edward J. Honton (1987b): „Scenario planning—What style should you use?“ *Long Range Planning*, Bd. 20(4): S. 21–29 (siehe S. 38).

Hyde, O. & P. F. Hodnett (1997): „An adaptable automated procedure for short-term electricity load forecasting“. *IEEE Transactions on Power Systems*, Bd. 12(1): S. 84–94 (siehe S. 31).

IBPSA-USA: *Building Energy Software Tools Directory* (siehe S. 11).

Im Sobol & Yu L. Levitan (1999): „A pseudo-random number generator for personal computers“. *Computers & Mathematics with Applications*, Bd. 37(4-5): S. 33–40 (siehe S. 216).

In Hyeob Yu, Jin Ki Lee, Jong Min Ko & Sun Ic Kim (2005): „A method for classification of electricity demands using load profile data“. *Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05)*: S. 164–168 (siehe S. 52).

Infield, D. G. & D. C. Hill (1998): „Optimal smoothing for trend removal in short term electricity demand forecasting“. *IEEE Transactions on Power Systems*, Bd. 13(3): S. 1115–1120 (siehe S. 31).

Jadhav, Sayali D. & H. P. Channe (2016): „Comparative study of K-NN, naive Bayes and decision tree classification techniques“. *International Journal of Science and Research*, Bd. 5(1): S. 1842–1845 (siehe S. 55).

Jain, A. K., M. N. Murty & P. J. Flynn (1999): „Data clustering: A review“. *ACM Computing Surveys*, Bd. 31(3): S. 264–323 (siehe S. 45).

Jain, Amit & B. Satish (2009): „Clustering based Short Term Load Forecasting using Support Vector Machines“. *2009 IEEE Bucharest PowerTech (POWERTECH)*: S. 1–8 (siehe S. 52).

Jain, Anil K. (2010): „Data clustering: 50 years beyond K-means“. *Pattern Recognition Letters*, Bd. 31(8): S. 651–666 (siehe S. 45).

Jain, Anil K. & Richard C. Dubes (1988): „Algorithms for clustering data“. Bd. (siehe S. 45, 48).

Jetter, Antonie & Willi Schweinfort (2011): „Building scenarios with Fuzzy Cognitive Maps: An exploratory study of solar energy“. *Futures*, Bd. 43(1): S. 52–66 (siehe S. 39).

Jiménez, M. J. & M. R. Heras (2005): „Application of multi-output ARX models for estimation of the U and g values of building components in outdoor testing“. *Solar Energy*, Bd. 79(3): S. 302–310 (siehe S. 15).

Kalamees, Targo (2004): „IDA ICE: the simulation tool for making the whole building energy and HAM analysis“. *Annex*, Bd. 41: S. 12–14 (siehe S. 11).

Kalman, Rudolph Emil (1960): „A new approach to linear filtering and prediction problems“. *Journal of basic Engineering*, Bd. 82(1): S. 35–45 (siehe S. 31).

Karypis, G., Eui-Hong Han & V. Kumar (1999): „Chameleon: Hierarchical clustering using dynamic modeling“. *Computer*, Bd. 32(8): S. 68–75 (siehe S. 48).

Kaufman, Leonard & Peter J. Rousseeuw (2009): *Finding groups in data: an introduction to cluster analysis*. Bd. 344. John Wiley & Sons (siehe S. 45).

Keirstead, James, Mark Jennings & Aruna Sivakumar (2012): „A review of urban energy system models: Approaches, challenges and opportunities“. *Renewable and Sustainable Energy Reviews*, Bd. 16(6): S. 3847–3866 (siehe S. 15).

Klein, S. A. & W. A. Beckman (2007): „TRNSYS 16: A transient system simulation program: mathematical reference“. Bd. 5: S. 389–396 (siehe S. 11).

Ko, Chia-Nan & Cheng-Ming Lee (2013): „Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter“. *Energy*, Bd. 49: S. 413–422 (siehe S. 31).

Kohonen, T. (1990a): „Improved versions of learning vector quantization“. *1990 IJCNN International Joint Conference on Neural Networks*. IEEE: 545–550 vol.1 (siehe S. 57, 98).

– (1990b): „The self-organizing map“. *Proceedings of the IEEE*, Bd. 78(9): S. 1464–1480 (siehe S. 51, 93, 94).

Kohonen, Teuvo (1981): „Automatic Formation of TopologicalMaps of Patterns in a Self-Organizing System“. Bd. (siehe S. 50).

– (2001): *Self-Organizing Maps*. Third Edition. Bd. 30. Springer Series in Information Sciences. Berlin & Heidelberg: Springer (siehe S. 57, 95–98).

Konstantin, Panos (2013): *Praxisbuch Energiewirtschaft: Energieumwandlung, -transport und -beschaffung im liberalisierten Markt*. 3., aktual. Aufl. 2013. VDI-Buch. Berlin, Heidelberg & s.l.: Springer Berlin Heidelberg (siehe S. 24).

Koschwitz, D., J. Frisch & C. van Treeck (2018): „Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale“. *Energy*, Bd. 165: S. 134–142 (siehe S. 36, 70, 73, 80, 82, 84, 87, 119, 121, 127–133, 238).

Koschwitz, Daniel, Mark-Alexander Brüntjen, Aitor Chivite, Jérôme Frisch & Christoph Alban van Treeck (2016): „Softwaregestützte Wärmebedarfsermittlung bei Informations- und Ressourcenknappheit“. *BauSIM 2016 : September 14-16, 2016, Dresden, Germany, September 14-16, 2016, Dresden, Germany / Hrsg.: Technische*

Universität Dresden; John Grunewald. Stuttgart: Fraunhofer IRB Verlag: S. 691–696 (siehe S. 114).

Koschwitz, Daniel, Eric Spinnräker, Jérôme Frisch & Christoph van Treeck (2019): „Long-term urban heating load predictions based on optimized retrofit orders: A cross-scenario analysis“. *Energy and Buildings*, Bd.: S. 109637 (siehe S. 90–92, 122, 137–139, 141–144).

Koschwitz, Daniel, Eric Spinnräker, Jérôme Frisch & Christoph van Treeck (2017): „Urbane Energieversorgungskonzepte und Sanierungsfahrpläne“. *HLH : Lüftung, Klima, Heizung, Sanitär, Gebäudetechnik*, Bd. 68(10): S. 19–22 (siehe S. 114).

Kosko, Bart (1986): „Fuzzy cognitive maps“. *International journal of man-machine studies*, Bd. 24(1): S. 65–75 (siehe S. 39).

Kotsiantis, S. B., I. D. Zaharakis & P. E. Pintelas (2006): „Machine learning: a review of classification and combining techniques“. *Artificial Intelligence Review*, Bd. 26(3): S. 159–190 (siehe S. 56).

Kotsiantis, Sotiris B., I. Zaharakis & P. Pintelas (2007): „Supervised machine learning: A review of classification techniques“. *Emerging artificial intelligence applications in computer engineering*, Bd. 160: S. 3–24 (siehe S. 55, 57).

Kwok, James T. (2001): „Linear dependency between ε and the input noise in ε-support vector regression“. *International Conference on Artificial Neural Networks*: S. 405–410 (siehe S. 83).

Labeeuw, Wouter & Geert Deconinck (2013): „Residential Electrical Load Model Based on Mixture Model Clustering and Markov Models“. *IEEE Transactions on Industrial Informatics*, Bd. 9(3): S. 1561–1569 (siehe S. 53).

Lahdelma, Risto, Pekka Salminen & Joonas Hokkanen (2000): „Using Multicriteria Methods in Environmental Planning and Management“. *Environmental Management*, Bd. 26(6): S. 595–605 (siehe S. 38).

Lance, G. N. & W. T. Williams (1967): „A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems“. *The Computer Journal*, Bd. 9(4): S. 373–380 (siehe S. 48).

LeCun, Yann, Leon Bottou, Genevieve B. Orr & Klaus-Robert Müller (1998): „Efficient BackProp“. *Neural Networks: Tricks of the Trade*. Hrsg. von Goos, Gerhard, Juris Hartmanis, Jan van Leeuwen, Genevieve B. Orr & Klaus-Robert Müller. Bd. 1524. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg: S. 9–50 (siehe S. 70).

Levenberg, Kenneth (1944): „A method for the solution of certain non-linear problems in least squares“. *Quarterly of applied mathematics*, Bd. 2(2): S. 164–168 (siehe S. 78).

Li, Qiong, Qinglin Meng, Jiejin Cai, Hiroshi Yoshino & Akashi Mochida (2009): „Applying support vector machine to predict hourly cooling load in the building“. *Applied Energy*, Bd. 86(10): S. 2249–2256 (siehe S. 35).

Li, Zhi Yong (2012): „An Empirical Study of Knowledge Discovery on Daily Electrical Peak Load Using Decision Tree“. *Advanced Materials Research*, Bd. 433-440: S. 4898–4902 (siehe S. 59).

Lin, T., B. G. Horne, P. Tino & C. L. Giles (1996): „Learning long-term dependencies in NARX recurrent neural networks“. *IEEE transactions on neural networks*, Bd. 7(6): S. 1329–1338 (siehe S. 36).

Littlestone, Nick (1988): „Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm“. *Machine learning*, Bd. 2(4): S. 285–318 (siehe S. 55).

Liu, Bidong, Jakub Nowotarski, Tao Hong & Rafal Weron (2015a): „Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts“. *IEEE Transactions on Smart Grid*, Bd.: S. 1 (siehe S. 14).

Liu, Jiying, Mohammad Heidarinejad, Stefan Gracik & Jelena Srebric (2015b): „The impact of exterior surface convective heat transfer coefficients on the building energy consumption in urban neighborhoods with different plan area densities“. *Energy and Buildings*, Bd. 86: S. 449–463 (siehe S. 40).

Lloyd, Gavin R., Richard G. Brereton, Rita Faria & John C. Duncan (2007): „Learning vector quantization for multiclass classification: application to characterization of plastics“. *Journal of chemical information and modeling*, Bd. 47(4): S. 1553–1563 (siehe S. 59).

Lohr, S. L. (1999): *Sampling: Design and Analysis*. Sampling: Design and Analysis. Duxbury Press (siehe S. 72).

Lourakis, Manolis I. A. (2005): „A brief description of the Levenberg-Marquardt algorithm implemented by levmar“. *Foundation of Research and Technology*, Bd. 4(1): S. 1–6 (siehe S. 78–80).

Lund, Henrik, Sven Werner, Robin Wiltshire, Svend Svendsen, Jan Eric Thorsen, Frede Hvelplund & Brian Vad Mathiesen (2014): „4th Generation District Heating (4GDH)“. *Energy*, Bd. 68: S. 1–11 (siehe S. 1, 16).

Lundin, Mikael, Staffan Andersson & Ronny Östin (2004): „Development and validation of a method aimed at estimating building performance parameters“. *Energy and Buildings*, Bd. 36(9): S. 905–914 (siehe S. 16).

MacQueen, James et al. (1967): „Some methods for classification and analysis of multivariate observations“. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Bd. 1: S. 281–297 (siehe S. 47).

Magoulès, F. & Hai-xiang Zhao (2016): *Data mining and machine learning in building energy analysis*. Computer engineering series. London, UK, ISTE Ltd. & Hoboken, NJ, USA: John Wiley & Sons, Inc (siehe S. 9, 11–13, 15, 16, 30, 32–35, 71).

Marquardt, Donald W. (1963): „An Algorithm for Least-Squares Estimation of Nonlinear Parameters“. *Journal of the Society for Industrial and Applied Mathematics*, Bd. 11(2): S. 431–441 (siehe S. 78).

Mavrogianni, A., M. Davies, M. Kolokotroni & I. Hamilton (2009): „A GIS-based bottom-up space heating demand model of the London domestic stock“. *Proceddings 11th International IBPSA Conference, Building Simulation* (siehe S. 15).

May-Ostendorp, Peter T., Gregor P. Henze, Balaji Rajagopalan & Charles D. Corbin (2013): „Extraction of supervisory building control rules from model predictive control of windows in a mixed mode building“. *Journal of Building Performance Simulation*, Bd. 6(3): S. 199–219 (siehe S. 60).

Mbamalu, G.A.N. & M. E. El-Hawary (1993): „Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation“. *IEEE Transactions on Power Systems*, Bd. 8(1): S. 343–348 (siehe S. 31).

McLoughlin, Fintan, Aidan Duffy & Michael Conlon (2015): „A clustering approach to domestic electricity load profile characterisation using smart metering data“. *Applied Energy*, Bd. 141: S. 190–199 (siehe S. 52, 53).

Mena, R., F. Rodríguez, M. Castilla & M. R. Arahal (2014): „A prediction model based on neural networks for the energy consumption of a bioclimatic building“. *Energy and Buildings*, Bd. 82: S. 142–155 (siehe S. 36).

Mika, Sebastian, Gunnar Ratsch, Jason Weston, Bernhard Scholkopf & Klaus-Robert Mullers (1999): „Fisher discriminant analysis with kernels“. *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*: S. 41–48 (siehe S. 57).

Molina-Solana, Miguel & María J. Ros (2017): „Data science for building energy management: A review“. *Renewable and Sustainable Energy Reviews*, Bd. 70: S. 598–609 (siehe S. 59, 60).

Mori, H. & A. Yuihara (2001): „Deterministic annealing clustering for ANN-based short-term load forecasting“. *IEEE Transactions on Power Systems*, Bd. 16(3): S. 545–551 (siehe S. 52).

Müller, D., M. Lauster, A. Constantin, M. Fuchs & P. Remmen (2016): „AixLib-An open-source modelica library within the IEA-EBC annex 60 framework“. *BauSIM 2016*, Bd.: S. 3–9 (siehe S. 11).

Murtagh, F. (1983): „A Survey of Recent Advances in Hierarchical Clustering Algorithms“. *The Computer Journal*, Bd. 26(4): S. 354–359 (siehe S. 48).

Narendra, K. S. & K. Parthasarathy (1991): „Learning automata approach to hierarchical multiobjective analysis“. *IEEE Transactions on Systems, Man, and Cybernetics*, Bd. 21(1): S. 263–272 (siehe S. 75).

Nazarko, J., A. Jurczuk & W. Zalewski (2005): „ARIMA models in load modelling with clustering approach“. *2005 IEEE Russia Power Tech*: S. 1–6 (siehe S. 52).

Neto, Alberto Hernandez & Flávio Augusto Sanzovo Fiorelli (2008): „Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption“. *Energy and Buildings*, Bd. 40(12): S. 2169–2176 (siehe S. 35).

Ng, R. T. & Jiawei Han (2002): „CLARANS: A method for clustering objects for spatial data mining“. *IEEE Transactions on Knowledge and Data Engineering*, Bd. 14(5): S. 1003–1016 (siehe S. 49).

Nguyen, Anh-Tuan, Sigrid Reiter & Philippe Rigo (2014): „A review on simulation-based optimization methods applied to building performance analysis“. *Applied Energy*, Bd. 113: S. 1043–1058 (siehe S. 15).

Nova, David & Pablo A. Estévez (2014): „A review of learning vector quantization classifiers“. *Neural Computing and Applications*, Bd. 25(3-4): S. 511–524 (siehe S. 57, 58).

Nytsch-Geusen, Christoph, C. Banhardt, A. Inderfurth, K. Mucha, J. Möckel, J. Rädler, M. Thorade & C. Ribas Tugores (2016): „BuildingSystems-Eine Modular Hierarchische Modell-Bibliothek zur Energetischen Gebäude- und Anlagensimulation“. *proceedings of BauSIM 2016*, Bd. (siehe S. 11).

Nytsch-Geusen, Christoph, Katharina Mucha, Alexander Inderfurth & Jörg Rädler (2014): „Entwicklung eines räumlich und physikalisch adaptiven energetischen Gebäudemodells in Modelica“. *Proceedings of the 5th German-Austrian IBPSA Conference, 22.-24. September* (siehe S. 11).

Obernberger, Ingwald (1997): „Möglichkeiten der technologischen und wirtschaftlichen Optimierung von Nahwärme-und Mikronetzen“. *Altener Konferenz "Mikronetze-Gebäudeübergreifende Wärmeversorgung auf Biomassebasis"* (siehe S. 2).

Olofsson, T. & S. Andersson (2002): „Overall heat loss coefficient and domestic energy gain factor for single-family buildings“. *Building and Environment*, Bd. 37(11): S. 1019–1026 (siehe S. 16).

Olonscheck, Mady, Anne Holsten & Jürgen P. Kropp (2011): „Heating and cooling energy demand and related emissions of the German residential building stock under climate change“. *Energy Policy*, Bd. 39(9): S. 4795–4806 (siehe S. 2).

Papalexopoulos, A. D. & T. C. Hesterberg (1990): „A regression-based approach to short-term system load forecasting“. *IEEE Transactions on Power Systems*, Bd. 5(4): S. 1535–1547 (siehe S. 31).

Park, J. H., Y. M. Park & K. Y. Lee (1991): „Composite modeling for adaptive short-term load forecasting“. *IEEE Transactions on Power Systems*, Bd. 6(2): S. 450–457 (siehe S. 31).

Parsons, Lance, Ehtesham Haque & Huan Liu (2004): „Subspace clustering for high dimensional data“. *ACM SIGKDD Explorations Newsletter*, Bd. 6(1): S. 90–105 (siehe S. 45).

Pfafferott, Jens, Sebastian Herkel & Jeannette Wapler (2005): „Thermal building behaviour in summer: Long-term data evaluation using simplified models“. *Energy and Buildings*, Bd. 37(8): S. 844–852 (siehe S. 15).

Platt, John C., Nello Cristianini & John Shawe-Taylor (2000): „Large margin DAGs for multiclass classification“. *Advances in neural information processing systems*: S. 547–553 (siehe S. 57).

Pohekar, S. D. & M. Ramachandran (2004): „Application of multi-criteria decision making to sustainable energy planning—A review“. *Renewable and Sustainable Energy Reviews*, Bd. 8(4): S. 365–381 (siehe S. 38).

Powell, Kody M., Akshay Sriprasad, Wesley J. Cole & Thomas F. Edgar (2014): „Heating, cooling, and electrical load forecasting for a large-scale district energy system“. *Energy*, Bd. 74: S. 877–885 (siehe S. 36, 74).

Rehau (2012): „Gleichzeitigkeit - der unterschätzte Faktor. Effiziente Planung von Nahwärmenetzen“. *BWK. Das Energie-Fachmagazin*, Bd. 64(12): S. 41–43 (siehe S. 17).

Reinhart, Christoph F. & Carlos Cerezo Davila (2016): „Urban building energy modeling – A review of a nascent field“. *Building and Environment*, Bd. 97: S. 196–202 (siehe S. 9, 10, 15, 16).

Reitermanova, Z. (2010): „Data splitting“. *WDS*. Bd. 10: S. 31–36 (siehe S. 72).

Rodrigues, Fátima, Jorge Duarte, Vera Figueiredo, Zita Vale & M. Cordeiro (2003): „A Comparative Analysis of Clustering Algorithms Applied to Load Profiling“. *Machine Learning and Data Mining in Pattern Recognition*. Hrsg. von Perner, Petra & Azriel Rosenfeld. Bd. 2734. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg: S. 73–85 (siehe S. 52, 53).

Roth, U. (1980): *Wechselwirkungen zwischen der Siedlungsstruktur und Wärmeversorgungssystemen*. 06: Schriftenreihe des Bundesministers für Raumordnung, Bauwesen und Städtebau. Bundesminister für Raumordnung, Bauwesen u. Städtebau (siehe S. 9).

Runkler, Thomas A. (2015): *Data Mining: Modelle und Algorithmen intelligenter Datenanalyse*. 2., aktualisierte Auflage. Computational Intelligence. Wiesbaden: Springer Vieweg (siehe S. 68, 70).

Saez, Doris, Fernand Avila, Daniel Olivares, Claudio Canizares & Luis Marin (2015): „Fuzzy Prediction Interval Models for Forecasting Renewable Resources and Loads in Microgrids“. *IEEE Transactions on Smart Grid*, Bd. 6(2): S. 548–556 (siehe S. 40).

Sakawa, Masatoshi, Kosuke Kato & Satoshi Ushiro (2001): „Cooling load prediction in a district heating and cooling system through simplified robust filter and multilayered neural network“. *Applied Artificial Intelligence*, Bd. 15(7): S. 633–643 (siehe S. 35).

Salminen, Pekka, Joonas Hokkanen & Risto Lahdelma (1998): „Comparing multicriteria methods in the context of environmental problems“. *European Journal of Operational Research*, Bd. 104(3): S. 485–496 (siehe S. 38).

Saltelli, Andrea, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana & Stefano Tarantola (2008): *Global sensitivity analysis: the primer*. John Wiley & Sons (siehe S. 10, 216).

Sargunaraj, S., D. P. Sen Gupta & S. Devi (1997): „Short-term load forecasting for demand side management“. *IEE Proceedings - Generation, Transmission and Distribution*, Bd. 144(1): S. 68 (siehe S. 31).

Sathiracheewin, Supalak & Vichai Surapatana (2011): „Daily typical load clustering of residential customers“. *2011 8th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON 2011)*: S. 797–800 (siehe S. 53).

Schellong, Wolfgang (2016): *Analyse und Optimierung von Energieverbundsystemen*. Berlin & Heidelberg: Springer Vieweg (siehe S. 30, 31).

Schramek, Ernst-Rudolf & Hermann Recknagel (2007): *Taschenbuch für Heizung+Klimatechnik 07/08*. Bd. 73. Oldenbourg Industrieverlag (siehe S. 18).

Semeraro, Luca, Emanuele Crisostomi, Alessandro Franco, Alberto Landi, Marco Raugi, Mauro Tucci & Giuseppe Giunta (2014): „Electrical load clustering: The Italian case“. *Proc. IEEE PES Conf. Innovative Smart Grid Technologies (ISGT) Europe* (siehe S. 53).

Shamshirband, Shahaboddin, Dalibor Petković, Rasul Enayatifar, Abdul Hanan Abdullah, Dušan Marković, Malrey Lee & Rodina Ahmad (2015): „Heat load prediction in district heating systems with adaptive neuro-fuzzy method“. *Renewable and Sustainable Energy Reviews*, Bd. 48: S. 760–767 (siehe S. 35).

Sheikholeslami, Gholamhosein, Surojit Chatterjee & Aidong Zhang (2000): „WaveCluster: A wavelet-based clustering approach for spatial data in very large databases“. *The VLDB Journal The International Journal on Very Large Data Bases*, Bd. 8(3-4): S. 289–304 (siehe S. 50).

Shu, Haiwen, Lin Duanmu, Chaohui Zhang & Yingxin Zhu (2010): „Study on the decision-making of district cooling and heating systems by means of value engineering“. *Renewable Energy*, Bd. 35(9): S. 1929–1939 (siehe S. 25).

Si, J., S. Lin & M.-A Vuong (2000): „Dynamic topology representing networks“. *Neural Networks*, Bd. 13(6): S. 617–627 (siehe S. 50).

Soliman, S. A., S. Persaud, K. El-Nagar & M. E. El-Hawary (1997): „Application of least absolute value parameter estimation based on linear programming to short-term load forecasting“. *International Journal of Electrical Power & Energy Systems*, Bd. 19(3): S. 209–216 (siehe S. 31).

Spinnräker, Eric, Daniel Koschwitz, Felix Kirchmann, Jérôme Frisch & Christoph van Treeck (2018): „Webbasiertes Energie-Informationssystem für urbane Räume“. *Bauingenieur*, Bd. 93(2): S. 68–72 (siehe S. 114, 115, 118).

Spinnräker, Eric, Daniel Koschwitz, Romana Markovic, Jérôme Frisch & Christoph van Treeck (2017): „Software-supported identification of an economically

optimized retrofit order by minimizing life-cycle costs using a genetic algorithm including constraints“. *Energy Procedia*, Bd. 122: S. 739–744 (siehe S. 114, 139).

Stern, F. (2013): „Chapter 10: peak demand and time-differentiated energy savings cross-cutting protocols“. *The uniform methods project: methods for determining energy efficiency savings for specific measures*, Bd.: S. 1–13 (siehe S. 67).

Swedish District Heating Association (2004): *District Heating Substations - Design and Installation - Technical requirements* (siehe S. 22).

Tajine, M. & D. Elizondo (1998): „Growing methods for constructing recursive deterministic perceptron neural networks and knowledge extraction“. *Artificial Intelligence*, Bd. 102(2): S. 295–322 (siehe S. 32).

Taşdemir, Kadim, Pavel Milenov & Brooke Tapsall (2011): „Topology-based hierarchical clustering of self-organizing maps“. *IEEE transactions on neural networks*, Bd. 22(3): S. 474–485 (siehe S. 52).

Taylor, J. W. & R. Buizza (2002): „Neural network load forecasting with weather ensemble predictions“. *IEEE Transactions on Power Systems*, Bd. 17(3): S. 626–632 (siehe S. 41).

Tian, Wei (2013): „A review of sensitivity analysis methods in building energy analysis“. *Renewable and Sustainable Energy Reviews*, Bd. 20: S. 411–419 (siehe S. 15).

Tol, Hakan Ibrahim & Svend Svendsen (2012a): „Operational planning of low-energy district heating systems connected to existing buildings“. *2nd International Conference on Renewable Energy: Generation and Applications* (siehe S. 20).

Tol, H.İ. & S. Svendsen (2012b): „Improving the dimensioning of piping networks and network layouts in low-energy district heating systems connected to low-energy buildings: A case study in Roskilde, Denmark“. *Energy*, Bd. 38(1): S. 276–290 (siehe S. 20).

Tsekouras, George J., Nikos D. Hatziargyriou & Evangelos N. Dialynas (2007): „Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers“. *IEEE Transactions on Power Systems*, Bd. 22(3): S. 1120–1128 (siehe S. 47).

Vagropoulos, Stylianos I., Evaggelos G. Kardakos, Christos K. Simoglou, Anastasios G. Bakirtzis & João P.S. Catalão (2016): „ANN-based scenario generation methodology for stochastic variables of electric power systems“. *Electric Power Systems Research*, Bd. 134: S. 9–18 (siehe S. 40).

van der Maaten, Laurens & Geoffrey Hinton (2008): „Visualizing data using t-SNE“. *Journal of machine learning research*, Bd. 9(Nov): S. 2579–2605 (siehe S. 50).

van Treeck, Christoph (2004): „Introduction to Building Performance Modeling and Simulation“. Priv. Doz. Technische Universität München (siehe S. 11).

– (2010): *Introduction to Building Performance Modeling and Simulation.* Habilitation: TU München (siehe S. 11).

van Treeck, Christoph Alban (2019): „Workshop 2: Bilanzräume, Systemgrenzen und Informationsaustauschanforderungen in Reallaboren“. *Dokumentation des 5. Projektleitertreffens Kälte und Kühlung - aktive und passive Konzepte : Detmold, 27. - 28. November 2018.* Energiewendebauen. Aachen: Wissenschaftliche Begleitforschung ENERGIEWENDEBAUEN, RWTH Aachen University, Lehrstuhl für Gebäude- und Raumklimatechnik: S. 23–26 (siehe S. 173).

van Vught, F. A. (1987): „Pitfalls of forecasting“. *Futures*, Bd. 19(2): S. 184–196 (siehe S. 38).

Vapnik, Vladimir N. (1995): *The Nature of Statistical Learning Theory.* New York, NY: Springer New York (siehe S. 81).

VDI 2067 Blatt 10 (2013): *Wirtschaftlichkeit gebäudetechnischer AnlagenEnergiebedarf von Gebäuden für Heizen, Kühlen,Be- und Entfeuchten.* Berlin (siehe S. 12).

VDI 2072 (2018): *Wärmeübergabestation mit Wasser-Wasser-Wärmeübertrager für Durchfluss-Trinkwassererwärmung/ Raumwärmeversorgung.* Berlin (siehe S. 18).

VDI 2078 (2015): *Berechnung der thermischen Lasten und Raumtemperaturen (Auslegung Kühllast und Jahressimulation).* Düsseldorf (siehe S. 11).

VDI 3807 Blatt 1 (2013): *Verbrauchskennwerte für Gebäude - Grundlagen* (siehe S. 14).

VDI 3807 Blatt 2 (2014): *Verbrauchskennwerte für Gebäude - Verbrauchskennwerte für Heizenergie, Strom und Wasser* (siehe S. 121).

Verdu, S. V., M. O. Garcia, C. Senabre, A. G. Marin & F.J.G. Franco (2006): „Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps“. *IEEE Transactions on Power Systems*, Bd. 21(4): S. 1672–1682 (siehe S. 50).

Vesanto, J. & E. Alhoniemi (2000): „Clustering of the self-organizing map“. *IEEE transactions on neural networks*, Bd. 11(3): S. 586–600 (siehe S. 95).

Vivarelli, Francesco & Christopher K.I Williams (2001): „Comparing Bayesian neural network algorithms for classifying segmented outdoor images“. *Neural Networks*, Bd. 14(4-5): S. 427–437 (siehe S. 56).

Wang, Meiping & Qi Tian (2013): „Application of wavelet neural network on thermal load forecasting“. *International Journal of Wireless and Mobile Computing*, Bd. 6(6): S. 608 (siehe S. 53).

Wang, Wei, Jiong Yang, Richard Muntz et al. (1997): „STING: A statistical information grid approach to spatial data mining“. *VLDB*. Bd. 97: S. 186–195 (siehe S. 50).

Weck-Ponten, S., R. Becker, S. Herle, J. Blankenbach, J. Frisch & C. van Treeck (2018): „Automatisierte Datenaggregation zur Einbindung einer dynamischen Gebäudesimulation in ein Geoinformationssystem“. *International Journal of Sustainable Development and Planning*, Bd. 13(1): S. 84–95 (siehe S. 173).

Wei Wang, Jiong Yang & R. Muntz (1999): „STING+: an approach to active spatial data mining“. *Proceedings 15th International Conference on Data Engineering (Cat. No.99CB36337)*: S. 116–125 (siehe S. 50).

Weissmann, Claudia, Tianzhen Hong & Carl-Alexander Graubner (2017): „Analysis of heating load diversity in German residential districts and implications for the application in district heating systems“. *Energy and Buildings*, Bd. 139: S. 302–313 (siehe S. 22).

Werner, Sven (2016): „European space cooling demands“. *Energy*, Bd. 110: S. 148–156 (siehe S. 2).

– (2017): „International review of district heating and cooling“. *Energy*, Bd. 137: S. 617–631 (siehe S. 1).

Westergren, Karl-Erik, Hans Högberg & Urban Norlén (1999): „Monitoring energy consumption in single-family houses“. *Energy and Buildings*, Bd. 29(3): S. 247–257 (siehe S. 15).

Weston, Jason, Chris Watkins et al. (1999): „Support vector machines for multi-class pattern recognition“. *Esann*. Bd. 99: S. 219–224 (siehe S. 56).

Wilson, D. R. & T. R. Martinez (2001): „The need for small learning rates on large problems“. *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222)*. IEEE: S. 115–119 (siehe S. 77, 78).

Winter, W., T. Haslauer & I. Obernberger (2001): „Untersuchungen der Gleichzeitigkeit in kleinen und mittleren Nahwärmenetzen“. *Euroheat & Power*, Bd. 09&10/2001: S. 1–17 (siehe S. 2, 9, 17, 19, 20, 150, 214–216).

W.M. Ma, Eden & Tommy W.S. Chow (2004): „A new shifting grid clustering algorithm“. *Pattern Recognition*, Bd. 37(3): S. 503–514 (siehe S. 50).

Wölki, Daniel (2017): *MORPHEUS: Modelica-based implementation of a numerical human model involving individual human aspects* (siehe S. 173).

Wu, Leon, Gail Kaiser, David Solomon, Rebecca Winter, Albert Boulanger & Roger Anderson (2012): „Improving efficiency and reliability of building systems using machine learning and automated online evaluation“. *2012 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*: S. 1–6 (siehe S. 35, 50).

Wu, Qin, Xingqin Qi, Eddie Fuller & Cun-Quan Zhang (2013): „"follow the leader": a centrality guided clustering and its application to social network analysis“. *TheScientificWorldJournal*, Bd. 2013: S. 368568 (siehe S. 48).

Wu, Sitao & Tommy W.S. Chow (2004): „Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density“. *Pattern Recognition*, Bd. 37(2): S. 175–188 (siehe S. 51, 52).

Wu, Xindong, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand & Dan Steinberg (2008): „Top 10 algorithms in data mining“. *Knowledge and Information Systems*, Bd. 14(1): S. 1–37 (siehe S. 32, 59).

Xiao-Feng Wang & De-Shuang Huang (2009): „A Novel Density-Based Clustering Framework by Using Level Set Method“. *IEEE Transactions on Knowledge and Data Engineering*, Bd. 21(11): S. 1515–1531 (siehe S. 49).

Xu, Lu, Tommy W. S. Chow & Eden W. M. Ma (2015): „Topology-based clustering using polar self-organizing map“. *IEEE transactions on neural networks and learning systems*, Bd. 26(4): S. 798–807 (siehe S. 46, 47, 49, 52).

Yager, R. R. (2000): „Intelligent control of the hierarchical agglomerative clustering process“. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, Bd. 30(6): S. 835–845 (siehe S. 48).

Yao, Yuan, Lorenzo Rosasco & Andrea Caponnetto (2007): „On Early Stopping in Gradient Descent Learning“. *Constructive Approximation*, Bd. 26(2): S. 289–315 (siehe S. 80).

Yarbrough, I., Q. Sun, D. C. Reeves, K. Hackman, R. Bennett & D. S. Henshel (2015): „Visualizing building energy demand for building peak energy analysis“. *Energy and Buildings*, Bd. 91: S. 10–15 (siehe S. 24, 25).

Yen, G. G. & Haiming Lu (2000): „Hierarchical genetic algorithm based neural network design“. *2000 IEEE Symposium on Combinations of Evolutionary Computation and*

Neural Networks. Proceedings of the First IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks (Cat. No.00EX448). IEEE: S. 168–175 (siehe S. 56).

Yin, Yanhong, Shoshi Mizokami & Kohei Aikawa (2015): „Compact development and energy consumption: Scenario analysis of urban structures based on behavior simulation". *Applied Energy*, Bd. 159: S. 449–457 (siehe S. 40).

Yu, Zhun, Fariborz Haghighat, Benjamin C.M. Fung & Hiroshi Yoshino (2010): „A decision tree method for building energy demand modeling". *Energy and Buildings*, Bd. 42(10): S. 1637–1646 (siehe S. 59).

Zakaria, Z. & K. L. Lo (2009): „Two-stage fuzzy clustering approach for load profiling". *Universities Power Engineering Conference (UPEC), 2009 Proceedings of the 44th International*: S. 1–5 (siehe S. 53).

Zakaria, Zuhaina, K. L. Lo & Mohamad Hadi Sohod (2006): „Application of Fuzzy Clustering to Determine Electricity Consumers' Load Profiles". *2006 IEEE International Power and Energy Conference*: S. 99–103 (siehe S. 53).

Zalewski, W. (2006): „Application of fuzzy inference to electric load clustering". *2006 IEEE Power India Conference*: 5 pp (siehe S. 53).

Zhang, Guoqiang, B. Eddy Patuwo & Michael Y. Hu (1998): „Forecasting with artificial neural networks". *International Journal of Forecasting*, Bd. 14(1): S. 35–62 (siehe S. 32).

Zhang, Lipeng, Jianjun Xia, Jan Eric Thorsen, Oddgeir Gudmundsson, Hongwei Li & Svend Svendsen (2016): „Technical, economic and environmental investigation of using district heating to prepare domestic hot water in Chinese multi-storey buildings". *Energy*, Bd. 116: S. 281–292 (siehe S. 23, 24).

Zhang, Tian, Raghu Ramakrishnan & Miron Livny (1996): „BIRCH". *the 1996 ACM SIGMOD international conference*. Hrsg. von Widom, Jennifer: S. 103–114 (siehe S. 48).

Zhao, Hai Xiang & Frédéric Magoulès (2010): „Parallel Support Vector Machines Applied to the Prediction of Multiple Buildings Energy Consumption". *Journal of Algorithms & Computational Technology*, Bd. 4(2): S. 231–249 (siehe S. 32).

Zhao Yanchang & Song Junde (2001): „GDILC: a grid-based density-isoline clustering algorithm". *2001 International Conferences on Info-tech and Info-net*: S. 140–145 (siehe S. 50).

Zhen, Li, D. M. Lin, H. W. Shu, Shuang Jiang & Y. X. Zhu (2007): „District cooling and heating with seawater as heat source and sink in Dalian, China“. *Renewable Energy*, Bd. 32(15): S. 2603–2616 (siehe S. 19).

Zhou, Kaile, Chao Fu & Shanlin Yang (2016): „Big data driven smart energy management: From big data to big insights“. *Renewable and Sustainable Energy Reviews*, Bd. 56: S. 215–225 (siehe S. 16).

Abbildungsverzeichnis

1.1 Transformation urbaner thermischer Energiesysteme 4

2.1 Schematische Darstellung der Unterschiede zwischen physikalischen und datenbasierten Modellen zur gebäudespezifischen Energieanalyse in Anlehnung an Coakley et al., 2014 10
2.2 Vergleich der Kennlinien zum Gleichzeitigkeitsfaktor nach VDI 2072 und Sander in Abhängigkeit von der Abnehmeranzahl 19
2.3 Kennlinien zur Änderung des Gleichzeitigkeitsfaktors für Raumwärme und Trinkwarmwasser mit zunehmender Abnehmerzahl in Anlehung an Tol et al, 2012 21
2.4 Vergleich der Kennlinien zum Gleichzeitigkeitsfaktor für Raumwärme von Winter et al. und Tol et al. in Abhängigkeit von der Abnehmeranzahl . . . 21
2.5 Vergleich norm- und richtlinienbasierter Kennlinien zu Gleichzeitigkeitsfaktoren in Anlehnung an Frederiksen et al. 2014 23
2.6 Vergleich länderspezifischer Kennlinien zu Gleichzeitigkeitsfaktoren in Anlehnung an Zhang et al. 2016 24

3.1 Exemplarische Darstellung eines einschichtigen Perzeptron-Netzes in Anlehnung an Magoulès et al., 2016 33
3.2 Exemplarische Darstellung eines mehrschichtigen Perzeptron Modells in Anlehnung an Magoulès et al., 2016 33
3.3 Exemplarische Darstellung eines KNN mit radialer Basisfunktion in Anlehnung an Magoulès et al., 2016 34
3.4 Exemplarische Darstellung eines Rekurrenten Netzes in Anlehnung an Magoulès et al., 2016 34
3.5 Szenario-Typologien nach Börjeson et al., 2006 39
3.6 Visualisierung eines positiven und negativen Extremszenarios sowie eines Trendszenarios 40

4.1 Prinzip des partitionierenden Clusterings am Beispiel des k-Means Algorithmus' in Anlehnung an Han et al., 2012 47
4.2 Prinzip des hierarchischen Clusterings in Anlehnung an Han et al., 2012 . . 48
4.3 Prinzip des dichtebasierten Clusterings in Anlehnung an Ester et al., 1996 . 50
4.4 Prinzip des gitterbasierten Clusterings in Anlehnung an Han et al., 2012 . . 51
4.5 Prinzip der SOM von Teuvo Kohonen in Anlehnung an Enami et al., 2010 . 51

4.6 Prinzip der Klassifikation mit Entscheidungsbäumen in Anlehnung an Kotsiantis et al., 2007 . . . 55
4.7 Prinzip der Klassifikation mit SVM in Anlehnung an Han et al., 2012 . . . 56
4.8 Schematische Darstellung von Referenzvektoren mit den dazugehörigen Klassen im Rahmen von LVQ2 in Anlehnung an Kohonen, 1990a . . . 57
4.9 Prinzip der Ensemble Methoden in Anlehnung an Han et al., 2012 . . . 58

5.1 Datenanalyse mittels Boxplot in Anlehnung an David C. Hoaglin, 2003 . . . 69
5.2 Flussdiagramm zum Lernprozess der Prognosemodelle in Anlehnung an Magoulès et al., 2016 . . . 71
5.3 Architektur des NARX RNN mit einer verdeckten Schicht in Anlehnung an Haykin, op. 1999 und Koschwitz et al., 2018 . . . 73
5.4 Logistische Sigmoidfunktion . . . 74
5.5 Architektur des NARX RNN mit zwei verdeckten Schichten in Anlehnung an Haykin, op. 1999 . . . 75
5.6 Prinzip der Fehlerrückführung (Backpropagation) in einem Feedforward-Netzwerk in Anlehnung an Bishop, 2006 . . . 76
5.7 Darstellung relevanter Zusammenhänge zwischen Parametern zur Regularisierung . . . 81
5.8 Darstellung der Fehlerfunktion für die ε-Support Vector Machine Regression in Anlehnung an Drucker et al., 1997 und Koschwitz et al., 2018 . . . 82
5.9 Darstellung der Kernparameter für die ε-Support Vector Machine Regression in Anlehnung an Bishop, 2006 und Koschwitz et al., 2018 . . . 84
5.10 Kurvenanpassung mit einer Gauss-Funktion in Anlehnung an Koschwitz et al., 2018 . . . 87
5.11 Kurvenanpassung mit einer polynomialen Funktion dritten Grades in Anlehnung an Koschwitz et al., 2018 . . . 87
5.12 Schematische Darstellung der Entscheidungslogik für die gebäudespezifisch ausgewählte Prognosemethode auf Grundlage der MSE-Werte in der jeweiligen Modelltestphase . . . 89
5.13 Schematische Darstellung der Modellszenarien . . . 90
5.14 Schematische Darstellung der Interpolationsmethode zur Generierung zukünftiger Wetterdatensätze auf Basis der TRJ . . . 91
5.15 Schematische Darstellung der vereinfachten Lastprofilanpassung infolge von Gebäudesanierungen in Anlehnung an Koschwitz et al., 2019 . . . 92
5.16 Ablaufschema zur automatisierten Identifizierung und Quantifizierung von charakteristischen gebäudespezifischen Lastprofilen mittels SOM und LVQ . 94
5.17 Schematische Darstellung einer hexagonalen Anordnung der Neuronen im Referenzvektorraum . . . 96
5.18 Exemplarische Darstellung von N_c innerhalb der hexagonalen Struktur des Referenzvektorraums . . . 97

6.1 Schematische Darstellung der entwickelten Softwarearchitektur von *AMSA* . 102

6.2 Darstellung der Benutzeroberfläche für initiale Einstellungen bezüglich der Datenquellen nach Aufruf von *AMSA* 103
6.3 Benutzeroberfläche von *AMSA* 105
6.4 Darstellung des dynamisch veränderlichen Teils der Benutzeroberfläche in der Kategorie *Data* bei der Auswahl des kurzfristigen Prognosezeitraumes . 106
6.5 Darstellung des dynamisch veränderlichen Teils der Benutzeroberfläche in der Kategorie *Data* bei der Auswahl des mittelfristigen Prognosezeitraumes 107
6.6 Darstellung des dynamisch veränderlichen Teils der Benutzeroberfläche in der Kategorie *Data* bei der Auswahl des langfristigen Prognosezeitraumes . 107
6.7 Darstellung der verschiedenen Zustände der Baumstruktur auf der Haupt-GUI von *AMSA* 109
6.8 Darstellung des Symbols zur Kennzeichnung der Subgebäudegruppe mit dem ähnlichsten Lastprofil zu jenem der Gesamtgebäudegruppe 110
6.9 Exemplarische Darstellung der Datenvisualisierung und grafischen Ergebnisausgabe in *AMSA* 110
6.10 Exemplarische Darstellung des Aufbaus von lokalen Datencontainern zur Speicherung von gebäudespezifisch trainierten Prognosemodellen 113

7.1 GIS-basierte Darstellung der Gebäude des Forschungszentrums Jülich und der Baualtersklassen in Anlehnung an Spinnräker et al., 2018 118
7.2 Darstellung der Jahresdauerlinien zu den TRJ hinsichtlich der Außenlufttemperatur 123
7.3 Vergleich zwischen Datensätzen zur thermischen Leistung eines Gebäudes inklusive und exklusive der Eliminierung von globalen Ausreißern 124
7.4 Vergleich zwischen Datensätzen zur thermischen Leistung eines Gebäudes inklusive und exklusive der Eliminierung von lokalen Ausreißern 125
7.5 Beispiel für ein auf einen Mittelwert von 0 normiertes und auf eine Standardabweichung von 1 skaliertes Lastprofil eines Gebäudes 126
7.6 Exemplarische Darstellung eines gebäudebezogenen Kühlleistungsprofils in Anlehnung an Koschwitz et al., 2018 131
7.7 Exemplarische Darstellung kurzfristiger aggregierter Wärme- und Kühllastprognosen für das Demonstrationsquartier über 200 Stunden im Winter (Heizfall) bzw. Sommer (Kühlfall) 135
7.8 Exemplarische Darstellung von zwei gebäudespezifischen kurzfristigen Kühllastprognosen über 200 Stunden im Juni 136
7.9 Darstellung der geordneten aggregierten Dauerlinien zu Wärmelastprognosen für das Demonstrationsquartier in Abhängigkeit von zukünftigen Trend- und Extremwetterszenarien für die Jahre 2020, 2025 und 2030 in Anlehnung an Koschwitz et al., 2019 137
7.10 Darstellung der geordneten aggregierten Dauerlinien zu Kühllastprognosen für das Demonstrationsquartier in Abhängigkeit von Trend- und Extremwetterszenarien für die Jahre 2020, 2025 und 2030 in Anlehnung an Koschwitz et al., 2019 138

7.11 Lastprognosen und szenarienbasierte Lastmodifikationen infolge verschiedener Sanierungsreihenfolgen für die Jahre 2019, 2025 und 2050 in Anlehnung an Koschwitz et al., 2019 141
7.12 Szenariospezifische, prognostizierte Maximum-, Minimum-, Median- und Mittelwerte für den Wärmeleistungsbedarf auf Quartiersebene in Anlehnung an Koschwitz et al., 2019 142
7.13 Exemplarisches Ergebnis der Wärmeleistungsreduktionen infolge Gebäudesanierung in Anlehnung an Koschwitz et al., 2019 143
7.14 Vergleich der prognostizierten Fluktuationen von maximalen Wärmeleistungsspitzen in Bezug auf die Szenarien A bis C in Folge einzelner Gebäudesanierungen in Anlehnung an Koschwitz et al., 2019 144
7.15 Darstellung der SOM-Topologie und der dazugehörigen 49 Neuronen mit der Anzahl an zugewiesenen Eingangsvektoren auf Grundlage der Gruppierung von Kühllastprofilen des Sommerquartals 145
7.16 Darstellung der SOM-Topologie und der dazugehörigen 24 Neuronen mit der Anzahl an zugewiesenen Eingangsvektoren auf Grundlage der Gruppierung von Kühllastprofilen des Sommerquartals 145
7.17 Darstellung der SOM-Topologie und der dazugehörigen 6 Neuronen mit der Anzahl an zugewiesenen Eingangsvektoren auf Grundlage der Gruppierung von Kühllastprofilen des Sommerquartals 146
7.18 Darstellung der gruppenspezifischen Standardlastprofile auf Grundlage des Clusterings mit einem geringen Detailgrad 147
7.19 Exemplarischer Auszug aus den Ergebnissen der Clusterüberlagerungen mit zwei gruppierten Lastprofilen 149
7.20 Ergebnis der statistischen Analyse zur Entwicklung des Gleichzeitigkeitsfaktors für die Wärmeversorgung in Abhängigkeit von der Abnehmergruppengröße 150
7.21 Darstellung der Näherungsfunktion zur Approximation des Gleichzeitigkeitsfaktors für die Wärmeversorgung in Abhängigkeit von der Größe der Abnehmergruppe 151
7.22 Vergleich der Näherungsfunktionen zur Approximation des Gleichzeitigkeitsfaktors für die Wärmeversorgung in Abhängigkeit von der Größe der Abnehmergruppe und der gewählten Norminnentemperatur im Hinblick auf die Bestimmung der Nennleistungen 152
7.23 Vergleich der im Rahmen dieser Arbeit hergeleiteten Kennlinie für den Gleichzeitigkeitsfaktor in Abhängigkeit von der Abnehmeranzahl mit jenen auf Grundlage der Studien von Winter et al. und Tol et al. 152
7.24 Ergebnis der statistischen Analyse zur Entwicklung des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der Gruppengröße 154
7.25 Darstellung der Näherungsfunktion zur Approximation des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der Größe der Abnehmergruppe 154

7.26 Vergleich der Näherungsfunktionen zur Approximation des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der Größe der Abnehmergruppe und der gewählten Auslegungstemperatur im Hinblick auf die Bestimmung der Nennleistungen 155
7.27 Spektrum des Gleichzeitigkeitsfaktors auf Grundlage der 10 höchsten Wärmeleistungsspitzen im Kontext historischer Wärmeleistungsdaten aus den Jahren 2016 und 2017 sowie auf Basis der 50 höchsten Wärmeleistungsspitzen in den Wärmebedarfsprognosen bis 2025 156
7.28 Spektrum des Gleichzeitigkeitsfaktors auf Grundlage der 10 höchsten Kühlleistungsspitzen im Kontext historischer Kühlleistungsdaten aus den Jahren 2016 und 2017 sowie auf Basis der 50 höchsten Kühlleistungsspitzen in den Kühlbedarfsprognosen bis 2025 157
7.29 Darstellung eines exemplarischen Ergebnisses zur Identifizierung von Versorgungsinseln auf Grundlage der Gleichzeitigkeitsanalyse mit *AMSA* 158
7.30 Darstellung der Gebäudegruppierungen auf Grundlage von Wärmelastprognosen für das Jahr 2025 159
7.31 Darstellung der Gebäudegruppierungen auf Grundlage von Wärmelastprognosen für das Jahr 2025 unter Berücksichtigung der ökonomisch optimierten Sanierungsreihenfolge im Quartier 160
7.32 Gebäudegruppe (*LV*) mit dem größten kontinuierlichen Lastverschiebungspotenzial auf Grundlage von Wärmelastprognosen für das Jahr 2025 161
7.33 Gebäudegruppe (*KLV*) mit dem zweitgrößten kontinuierlichen Lastverschiebungspotenzial auf Grundlage von Wärmelastprognosen für das Jahr 2025 . 161
7.34 Prozentualer Anteil des gruppenspezifischen Wärmeleistungsbedarfs zum Zeitpunkt des Spitzenwärmeleistungsbedarfs auf Quartiersebene in Abhängigkeit vom Prognosezeithorizont und analysierter Leistungsspitzenzahl . . 162
7.35 Prozentualer Anteil des gruppenspezifischen Kühlleistungsbedarfs zum Zeitpunkt des Spitzenkühlleistungsbedarfs auf Quartiersebene in Abhängigkeit vom Prognosezeithorizont und der Leistungsspitzenzahl 163
7.36 Gruppenspezifische Anteile an Spitzenkühllasten auf Quartiersebene auf Basis von 72-Stunden-Prognosen und fünf untersuchten Leistungsspitzen . . 163
7.37 Exemplarische systemische Darstellung der Energiesystemvarianten infolge zukünftiger Lastprofilmodifikationen unter Berücksichtigung des szenarioabhängigen Lastverschiebungspotenzials 164
7.38 Exemplarische systemische Darstellung der Implikationen für die Dimensionierung zentraler Energieinfrastrukturen infolge der prädiktiven Lastszenarioanalysen und resultierender Gleichzeitigkeitsfaktoren 165

A.1 Schematische Darstellung der linearen Regressionsanalyse zur approximierten Ermittlung der gebäudespezifischen maximalen Wärmeleistung 212
A.2 Schematische Darstellung der linearen Regressionsanalyse zur approximierten Ermittlung der gebäudespezifischen Kühlnennleistung 213

B.1 Schematische Darstellung der Klassenstruktur von AMSA 220

D.1 Stündliche Heizlast eines Gebäudes in Abhängigkeit von der Außentemperatur mit einem witterungsabhängigen und witterungsunabhängigen Leistungsanteil bei einer Heizgrenztemperatur von circa 15°C 234
D.2 Stündliche nahezu ausschließlich witterungsabhängige Gebäudeheizlast . . . 234
D.3 Stündliche Heizlast eines Gebäudes in Abhängigkeit von der Außentemperatur mit einem witterungsabhängigen und witterungsunabhängigen Leistungsanteil bei einer Heizgrenztemperatur von circa 15°C mit Nacht- und Wochenendabsenkung (in Anlehnung an Frederiksen et al., 2014) 235
D.4 Ergebnis der statistischen Analyse zur Entwicklung des Gleichzeitigkeitsfaktors für die Wärmeversorgung in Abhängigkeit von der Abnehmergruppengröße bei einer gebäudebezogenen Norminnentemperatur von 22 °C 236
D.5 Ergebnis der statistischen Analyse zur Entwicklung des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der Abnehmergruppengröße bei einer gebäudebezogenen Auslegungstemperatur von 35 °C 236

E.1 Exemplarischer Auszug aus den Ergebnissen der saisonalen Heizlastprognosen mit NARX RNN und ε-SVM-R Modellen (Koschwitz et al., 2018) 238
E.2 Exemplarischer Auszug aus den Ergebnissen der saisonalen Kühllastprognosen mit NARX RNN und ε-SVM-R Modellen (Koschwitz et al., 2018) 238

Tabellenverzeichnis

2.1 Kennwerte für Gleichzeitigkeitsfaktoren in Abhängigkeit von dem Siedlungstyp nach Blesl, 2002 25

3.1 Zusammenfassung der Analyseergebnisse zur Eignung der Prognosemethodenkategorien für die Modellbildung in dieser Arbeit 43

4.1 Zusammenfassung der Analyseergebnisse zur Eignung der Methodenkategorien im Bereich des Clustering für die Modellbildung in dieser Arbeit . . . 61
4.2 Zusammenfassung der Analyseergebnisse zur Eignung der Methodenkategorien im Bereich der Klassifikation für die Modellbildung in dieser Arbeit . . 62

6.1 Übersicht zu den Auswahloptionen im Hauptfenster von *AMSA* 104

7.1 Anzahl der betrachteten Abnehmer im Wärme- und Kältenetz in Abhängigkeit von Gebäudenutzungsklassen in Anlehnung an Koschwitz et al., 2018 119
7.2 Auflistung der projektbezogenen Wetterparameter mit dem dazugehörigen individuellen Messzeitraum 120
7.3 Spezifische Wärmeverbrauchskennwerte in Abhängigkeit von Gebäudenutzungsklassen nach VDI 3807 in Anlehnung an Koschwitz et al., 2019 122
7.4 Eingangsparameter für die exemplarische Berechnung des Referenzwertes für den Wärmeverbrauch eines Gebäudekomplexes mit multipler Nutzung nach Sanierung in Anlehnung an Koschwitz et al., 2019 122
7.5 Auflistung der modellspezifischen MSE und MAE Medianwerte zu Prognosen des saisonalen gebäudeindividuellen Wärmeleistungsbedarfs in Anlehnung an Koschwitz et al., 2018 128
7.6 Auflistung der modellspezifischen MSE und MAE Medianwerte zu Prognosen des saisonalen gebäudeindividuellen Kühlleistungsbedarfs in Anlehnung an Koschwitz et al., 2018 129
7.7 Anzahl der minimalen MAE Werte saisonaler Wärmelastprognosen in Anlehnung an Koschwitz et al., 2018 130
7.8 Anzahl der minimalen MAE-Werte saisonaler Kühllastprognosen in Anlehnung an Koschwitz et al., 2018 130
7.9 Prozentuale Nutzung der einzelnen Prognosemodelle innerhalb des Ensembles auf Basis der Datengrundlage für kurzfristige Lastvorhersagen 132

7.10 Prozentuale Nutzung der einzelnen Prognosemodelle innerhalb des Ensembles auf Basis der Datengrundlage für mittel- und langfristige thermische Lastvorhersagen 132
7.11 Prozentuale Nutzung der Prognosemodelle auf Grundlage eines variierenden Grades der Eliminierung von Ausreißern im Trainingsdatensatz 133
7.12 Cluster für die Kälteversorgung auf Grundlage von zwei analysierten Jahren mit maximal 6 Clustern und mehr als einem Mitglied je Cluster 146
7.13 Cluster für die Kälteversorgung auf Grundlage von zwei analysierten Jahren mit maximal 24 Clustern und mehr als einem Mitglied je Cluster 148
7.14 Cluster für die Kälteversorgung auf Grundlage von zwei analysierten Jahren mit maximal 49 Clustern und mehr als einem Mitglied je Cluster 148

B.1 Auflistung und Beschreibung der Funktionen der Klasse *settingsGUI* von AMSA 221
B.2 Auflistung und Beschreibung der Funktionen der Klasse *gui* von AMSA . . 221
B.3 Auflistung und Beschreibung der Funktionen der Klasse *main* von AMSA . 222
B.4 Auflistung und Beschreibung der Funktionen der Klasse *forecastModels* von AMSA 223
B.5 Auflistung und Beschreibung der Funktionen der Klasse *profileClustering* von AMSA 223
B.6 Auflistung und Beschreibung der Funktionen der Klasse *profileClassification* von AMSA 223
B.7 Auflistung und Beschreibung der Funktionen der Klasse *webInterface* von AMSA 224
B.8 Auflistung und Beschreibung der Funktionen der Klasse *influxInterface* von AMSA 224
B.9 Auflistung und Beschreibung der Funktionen der Klasse *localInterface* von AMSA 224
B.10 Auflistung und Beschreibung der Funktionen der Klasse *dataPreprocessing* von AMSA 224
B.11 Auflistung der Komponenten des Kennzeichnungsschlüssels für Prognosezeitreihen in der Influx Datenbank 225

C.1 Auflistung der betrachteten Gebäudekomplexe mit Angabe der Nutzung und Versorgungsart (1/3) 228
C.2 Auflistung der betrachteten Gebäudekomplexe mit Angabe der Nutzung und Versorgungsart (2/3) 229
C.3 Auflistung der betrachteten Gebäudekomplexe mit Angabe der Nutzung und Versorgungsart (3/3) 230
C.4 Aufteilung des gebäudespezifischen Wärmebedarfs 231

A Methoden zur Datenanalyse

A.1 Statistische Kennwerte

Im Rahmen dieses Abschnitts werden die verschiedenen statistischen Kennwerte zur Charakterisierung der Zeitreihen eingeführt. Die Ermittlung des arithmetischen Mittelwertes einer Zeitreihe erfolgt nach Gleichung (A.1), wobei der Durchschnittswert für n Messwerte x berechnet wird (Han et al., 2012).

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{A.1}$$

Die Datenverteilung ist häufig nicht symmetrisch, sodass der Median, welcher den Datensatz in eine obere und untere Hälfte unterteilt, besser geeignet ist als das arithmetische Mittel (Han et al., 2012). Der Median x_{med} ist gemäß Gleichung (A.2) definiert (Fahrmeir et al., 2016).

$$x_{med} = \begin{cases} x_{\frac{n+1}{2}} & n \text{ ungerade} \\ \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & n \text{ gerade.} \end{cases} \tag{A.2}$$

Mit der Varianz und der Standardabweichung lässt sich die Abweichung der Datenpunkte von dem Mittelwert quantifizieren. Dabei nehmen die Varianz und die Standardabweichung jeweils einen geringeren Wert an je stärker sich die Datenpunkte dem Mittelwert annähern (Han et al., 2012). Die Standardabweichung σ ist ferner nach Gleichung (A.3) definiert (Han et al., 2012).

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} |x_i - \mu|^2} \tag{A.3}$$

Die Varianz lässt sich überdies mittels der Quadrierung der Standardabweichung ermitteln. Für die durchschnittliche Abweichung prognostizierter Werte y_i von gegebenen Zielwerten t_i lässt sich ferner der MAE nach Gleichung (A.4) als Fehlermaß anführen (Amasyali et al., 2018).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - t_i| \tag{A.4}$$

A.2 Korrelationsanalyse

Der Korrelationskoeffizient r zwischen zwei Attributen X_1 und X_2 ist gemäß Gleichung (A.5) zu ermitteln, wobei n Tupel gegeben sind und $x_{1,i}$ sowie $x_{2,i}$ repräsentative Werte von X_1 und X_2 darstellen (Han et al., 2012).

$$r\left(X_1,X_2\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_{1,i}-\mu_{X_1}}{\sigma_{X_1}}\right)\left(\frac{x_{2,i}-\mu_{X_2}}{\sigma_{X_2}}\right) \tag{A.5}$$

Nimmt r den Wert 0 an, besteht keine Korrelation zwischen den Attributen X_1 und X_2. Ein Wert für r, der kleiner als 0 ist, impliziert eine negative Korrelation zwischen beiden Attributen, wobei eines der beiden steigt und eines fällt. Eine positive Korrelation liegt bei einem Korrelationskoeffizienten vor, welcher größer als 0 ist. In diesem Fall steigen beide Attribute. (Han et al., 2012) Unter Berücksichtigung der Definition der Kovarianz nach Gleichung (A.6)

$$Cov\left(X_1,X_2\right) = \frac{1}{n}\sum_{i=1}^{n}\left(x_{1,i}-\mu_{X_1}\right)\left(x_{2,i}-\mu_{X_2}\right) \tag{A.6}$$

lässt sich Gleichung (A.5) für r wie folgt vereinfachen (Han et al., 2012):

$$r\left(X_1,X_2\right) = \frac{Cov\left(X_1,X_2\right)}{\sigma_{X_1}\sigma_{X_2}} \tag{A.7}$$

Die symmetrische Korrelationsmatrix R enthält ferner die Informationen bezüglich der Korrelation zwischen Attributen (Fahrmeir et al., 2016). Hinsichtlich der beliebigen Attribute X_1 und X_2 ergibt sich Gleichung (A.8).

$$R = \begin{pmatrix} r\left(X_1,X_1\right) & r\left(X_1,X_2\right) \\ r\left(X_2,X_1\right) & r\left(X_2,X_2\right) \end{pmatrix} \quad = \quad \begin{pmatrix} 1 & r\left(X_1,X_2\right) \\ r\left(X_2,X_1\right) & 1 \end{pmatrix} \tag{A.8}$$

Die Korrelationskoeffizienten auf der Hauptdiagonalen nehmen dabei aufgrund der Linearitätsbedingung stets den Wert 1 an.

A.3 Konventionelle Herleitung von Kennlinien für den Gleichzeitigkeitsfaktor

Im Kontext der Modelldemonstration in Kapitel 7 dient die Herleitung von Kennlinien für den Gleichzeitigkeitsfaktor für das Demonstrationsquartier der Evaluierung der in dieser Arbeit entwickelten Methodik sowie der Motivation für die Verwendung des autoadaptiven Modells.

Ermittlung von gebäudespezifischen maximalen Leistungsanforderungen

Die gebäudespezifischen maximalen Leistungsanforderungen lassen sich mittels linearer Regressionsanalysen abschätzen, indem die gemessene thermische Leistung in einem zu definierenden Zeitraum mit der simultan gemessenen Außentemperatur korreliert wird. Dieser Ansatz entspricht der vereinfachten Heizlastberechnung nach DIN EN 12831 Beiblatt 2 (DIN 12831 Beiblatt 2, 2012) und folgt der Annahme, dass der thermische Leistungsbedarf eines Gebäudes im Wesentlichen von der Außentemperatur abhängt. Vor diesem Hintergrund ist in Abbildung A.1 die lineare Regressionsanalyse mit den dazugehörigen Parametern zur Ermittlung der gebäudespezifischen maximalen Heizleistungen schematisch dargestellt.

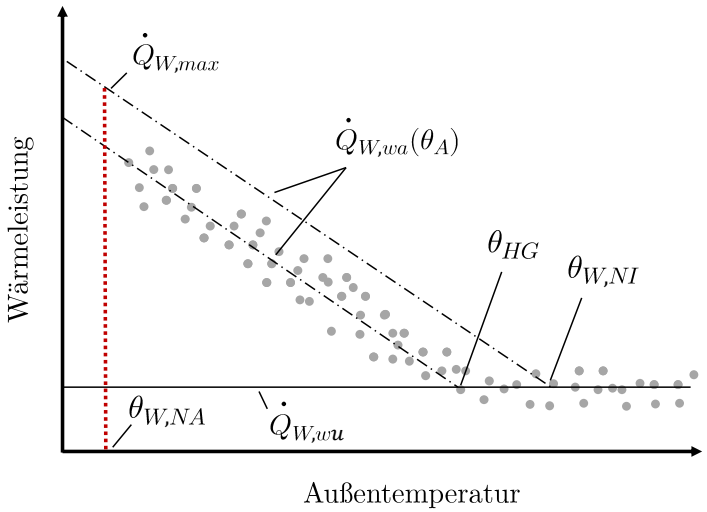


Abbildung A.1: Schematische Darstellung der linearen Regressionsanalyse zur approximierten Ermittlung der gebäudespezifischen maximalen Wärmeleistung

Sowohl für die Abbildung der witterungsunabhängigen Grundleistung $\dot{Q}_{W,wu}(\theta_A)$ als auch für die Approximation der witterungsabhängigen Wärmeleistung $\dot{Q}_{W,wu}$ des Gebäudes lässt sich jeweils eine lineare Funktion bestimmen, deren Parameter durch die dazugehörigen Messpunkte definiert werden. Der Schnittpunkt beider Regressionsgeraden repräsentiert

die Heizgrenze mit der dazugehörigen Außentemperatur θ_{HG}. Infolge der Extrapolation der Heizgeraden auf die Norminnentemperatur $\theta_{W,NI}$, lässt sich die erforderliche maximale Wärmeleistung $\dot{Q}_{W,max}$ zur Normaußentemperatur $\theta_{W,NA}$ für die Aufrechterhaltung von $\theta_{W,NI}$ im Gebäude bestimmen. Die Quantifizierung von $\theta_{W,NA}$ variiert ortsabhängig und ist daher individuell festzulegen.

Die Abschätzung der maximalen Kühlleistungsanforderungen eines Gebäudes basiert im Rahmen dieser Arbeit im Sinne der konventionellen Quantifizierung des Gleichzeitigkeitsfaktors ebenfalls auf der linearen Regression in umgekehrter Analogie. Abbildung A.2 veranschaulicht dazu die funktionalen Zusammenhänge und Parameter.

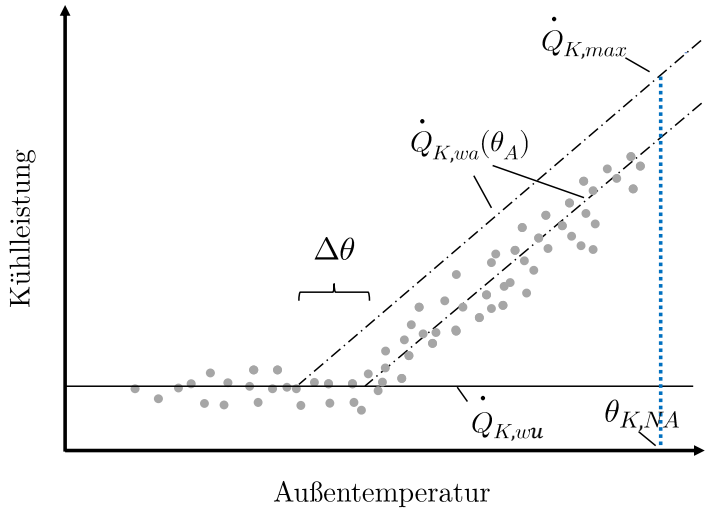


Abbildung A.2: Schematische Darstellung der linearen Regressionsanalyse zur approximierten Ermittlung der gebäudespezifischen Kühlnennleistung

Der Schnittpunkt zwischen witterungsabhängiger Kühlleistung $\dot{Q}_{K,wa}\,(\theta_A)$ und witterungsunabhängiger Kühlleistung $\dot{Q}_{K,wu}$ stellt die Kühlgrenztemperatur dar. Diese liegt häufig unter der Norminnentemperatur, sodass eine Extrapolation der Kühlgeraden auf die Norminnentemperatur nicht sinnvoll ist. Um dennoch einen Sicherheitsfaktor in der Berechnung der maximalen Kühlleistung zu berücksichtigen, erfolgt die Definition von $\Delta\theta = 5\text{K}$. Die Ableitung dieser Annahme basiert auf der Methodik zur Ableitung des maximalen Wärmebedarfs mittels linearer Regression. Dabei liegt häufig eine Heizgrenztemperatur von $15°C$ und eine anzusetzende Norminnentemperatur von 20 $°C$ mit einer resultierenden Differenz von fünf Kelvin vor. Folglich wird die Kühlgerade von der gebäudeindividuellen Kühlgrenztemperatur um fünf Kelvin nach links verschoben. Dadurch lässt sich in der Folge die maximal benötigte gebäudespezifische Kühlleistung $\dot{Q}_{K,max}$ bei Normaußentemperatur $\theta_{K,NA}$ näherungsweise bestimmen.

Dominanzanalyse

Um Unterdimensionierungen bei der Quantifizierung des Gleichzeitigkeitsfaktors infolge heterogener Nennleistungen innerhalb der Gebäudegruppe zu vermeiden, haben Winter et al. einen Algorithmus entwickelt, welcher das Dominierverhalten einzelner Abnehmer für die Quantifizierung des Gleichzeitigkeitsfaktors berücksichtigt (Winter et al., 2001; Gaderer, 2008). Dieser Algorithmus wird im Rahmen dieses Unterabschnitts vorgestellt. Eingangs erfolgt dabei die Definition eines Dominierfaktors d nach Gleichung (A.9), der sich aus dem Verhältnis zwischen der Summe dominierender Abnehmerleistungen $\dot{Q}_{sup}$ und der Summe aller dominierten Abnahmeleistungen $\dot{Q}_{inf}$ bestimmt (Gaderer, 2008).

$$d = \frac{\sum_{i=1}^{N} \dot{Q}_{i,sup}}{\sum_{i=1}^{N} \dot{Q}_{i,inf}} \tag{A.9}$$

Wie bereits erläutert, besteht die Gefahr einer Unterdimensionierung des Gleichzeitigkeitsfaktors in der Annahme vermeintlich homogener Abnahmeleistungen der Gebäude. Die Gebäude sind folglich der Höhe ihrer Nennleistungen entsprechend in Gruppen zusammen zu fassen. Somit ergibt sich der minimale Gleichzeitigkeitsfaktor ϕ_{min} aus jenem für sowohl dominierende Abnehmer n_{sup} als auch dominierte Abnehmer n_{inf} gemäß Gleichung (A.10) (Gaderer, 2008).

$$\phi_{min} = \phi\left(n_{inf} + n_{sup}\right) \tag{A.10}$$

Der maximale Gleichzeitigkeitsfaktor nach Gleichung (A.11) ergibt sich aus dem Verhältnis der Summe der jeweils um den Gleichzeitigkeitsfaktor korrigierten dominierenden und dominierten Abnahmeleistungen sowie der Summe aller Abnahmeleistungen.

$$\phi_{ges,max} = \frac{\phi\left(n_{inf}\right) \cdot \sum_{i=1}^{N} \dot{Q}_{i,inf} + \phi\left(n_{sup}\right) \cdot \sum_{i=1}^{N} \dot{Q}_{i,sup}}{\sum_{i=1}^{N} \dot{Q}_{i,inf} + \sum_{i=1}^{N} \dot{Q}_{i,sup}} \tag{A.11}$$

Durch Umformen der Gleichung (A.9) und Einsetzen in Gleichung (A.11) lässt sich Gleichung (A.12) formulieren (Gaderer, 2008).

$$\phi_{ges,max} = \frac{\phi\left(n_{inf}\right) + d \cdot \phi\left(n_{sup}\right)}{d+1} \tag{A.12}$$

Ferner hängt der Einfluss dominierender Abnehmer von der betrachteten Gruppengröße N ab, sodass ein minimaler Dominierfaktor d_{min} entsprechend Gleichung (A.13) festzulegen ist, wobei diese ausschließlich für $N > 2$ gilt (Gaderer, 2008).

$$d_{min} = \frac{0{,}1}{N-1} + 0{,}9 \tag{A.13}$$

Mit dieser Definition lässt sich d_{min} für ein Intervall zwischen 0 und 1 festlegen. Hinsichtlich des Vorgehens bei der Berechnung des Gleichzeitigkeitsfaktors unter Berücksichtigung heterogener Abnahmeleistungen wird im Anschluss an die Sortierung der Abnahmeleistungen nach ihrer Größe mit der Quantifizierung des Dominierfaktors des größten Abnehmers (d_n) nach Gleichung (A.14) begonnen (Gaderer, 2008).

$$d_n = \frac{\dot{Q}_n}{\sum_{i=1}^{n-1} \dot{Q}_i} \tag{A.14}$$

Als Dominierer werden jene Abnehmerleistungen definiert, für die $d_n > d_{min}$ gilt. Weiterhin ist ein maximaler Dominierfaktor d_{max} als obere Grenze festzulegen. Für letzteren lässt sich beispielsweise Gleichung (A.15) für $N > 2$ definieren (Gaderer, 2008).

$$d_{max} = \frac{9}{n-1} + 1 \tag{A.15}$$

Mittels dieses Algorithmus lassen sich für verschiedene Gruppengrößen unter Berücksichtigung von beliebigen Nennleistungsvariationen Gleichzeitigkeitsfaktoren ermitteln.

Herleitung von Kennlinien

Die Herleitung von Kennlinien, welche die funktionale Abhängigkeit zwischen dem Gleichzeitigkeitsfaktor und der Anzahl an Gebäuden als Energieabnehmern einer betrachteten Gruppe wiedergibt, bedingt die Zusammenstellung unterschiedlicher Gebäudegruppen von variierender Größe, für welche jeweils der Gleichzeitigkeitsfaktor bestimmt wird. Die maximale Gruppengröße wird auf $\frac{2}{3}$ der Grundgesamtheit begrenzt um eine ausreichende Variation in den Gruppenzusammensetzungen sicher stellen zu können (Winter et al., 2001). Je nach Gruppengröße ergeben sich, basierend auf den Prinzipien der Kombinatorik, eine sehr große Anzahl an möglichen Gruppenzusammensetzungen, wodurch der numerische Rechenaufwand sehr hoch ist. Um letzteren zu verringern werden Zufallszahlen verwendet, welche in verschiedensten Gebieten der Simulation von zufälligen Ereignissen Eingang finden. Eine Zufallszahl, die von einem Computer gezogen wird, beruht jedoch nicht auf

„echtem “ Zufall, sondern wird von dem hinterlegten Algorithmus beeinflusst, sodass sie als Pseudo-Zufallszahl zu betrachten ist (Im Sobol et al., 1999). Pseudo-Zufallszahlen lassen sich für festzulegende Intervalle mittels Zufallsgeneratoren, die in Softwarepaketen enthalten sind, generieren. Innerhalb eines definierten Intervalls sind sie jedoch nicht gleichmäßig verteilt, sondern gruppiert, getrennt durch Lücken. (Saltelli et al., 2008) Mittels Sobol' Sampling lässt sich dieser Effekt jedoch reduzieren (Burhenne, 2013). Daher findet dieser Sampling-Typ Anwendung. Je nach Gruppengröße werden ferner im Rahmen dieser Arbeit randomisiert 100 mögliche Zusammensetzungen mittels Zufallszahlen ermittelt und für jede Stichprobe der Gleichzeitigkeitsfaktor berechnet. Letzterer lässt sich anschließend in Abhängigkeit von der Gruppengröße in einem Diagramm auftragen. Um eine Unterdimensionierung des Gleichzeitigkeitsfaktors zu vermeiden, ist die Kennlinie dabei an den oberen Rand der Punktewolke anzupassen (Winter et al., 2001).

A.4 Ermittlung der Koordinaten zu den geometrischen Gebäudeschwerpunkten

Die gegebenen Gebäudekoordinaten beschreiben einen Polygonzug, der die Gebäudeflächen als Nettogrundflächen A definiert. Letztere werden auf Grundlage der x- und y-Koordinaten, ω_x bzw. ω_y, des Polygonzugs entsprechend Gleichung (A.16) ermittelt (Dankert et al., 2011).

$$A = \frac{1}{2} \sum_{i=1}^{N-1} \omega_{x,i} \cdot \omega_{y,i+1} - \omega_{x,i+1} \cdot \omega_{y,i} \tag{A.16}$$

Zur Ermittlung der Flächenschwerpunkte für die Darstellung der Gebäudeverortungen als Punktewolke in *AMSA* dienen ferner die statischen Momente nach den Gleichungen (A.17) und (A.18) (Dankert et al., 2011)

$$S_x = \frac{1}{6 \cdot A} \sum_{i=1}^{N-1} (\omega_{x,i} \cdot \omega_{y,i+1} - \omega_{x,i+1} \cdot \omega_{y,i}) \cdot (\omega_{x,i} + \omega_{x,i+1}) \tag{A.17}$$

$$S_y = \frac{1}{6 \cdot A} \sum_{i=1}^{N-1} (\omega_{x,i} \cdot \omega_{y,i+1} - \omega_{x,i+1} \cdot \omega_{y,i}) \cdot (\omega_{y,i} + \omega_{y,i+1}) \tag{A.18}$$

Die Schwerpunkte werden für alle Gebäudeteile einzeln ermittelt und anschließend für die Berechnung des Schwerpunkts eines Gebäudekomplexes flächengewichtet gemäß der

Gleichungen (A.19) und (A.20) addiert (Dankert et al., 2011).

$$S_{x,ges} = \frac{\sum S_{x,i} \cdot A_i}{\sum A_i} \tag{A.19}$$

$$S_{y,ges} = \frac{\sum S_{y,i} \cdot A_i}{\sum A_i} \tag{A.20}$$

B Software- und Datenstrukturen

B.1 UML-Diagramm

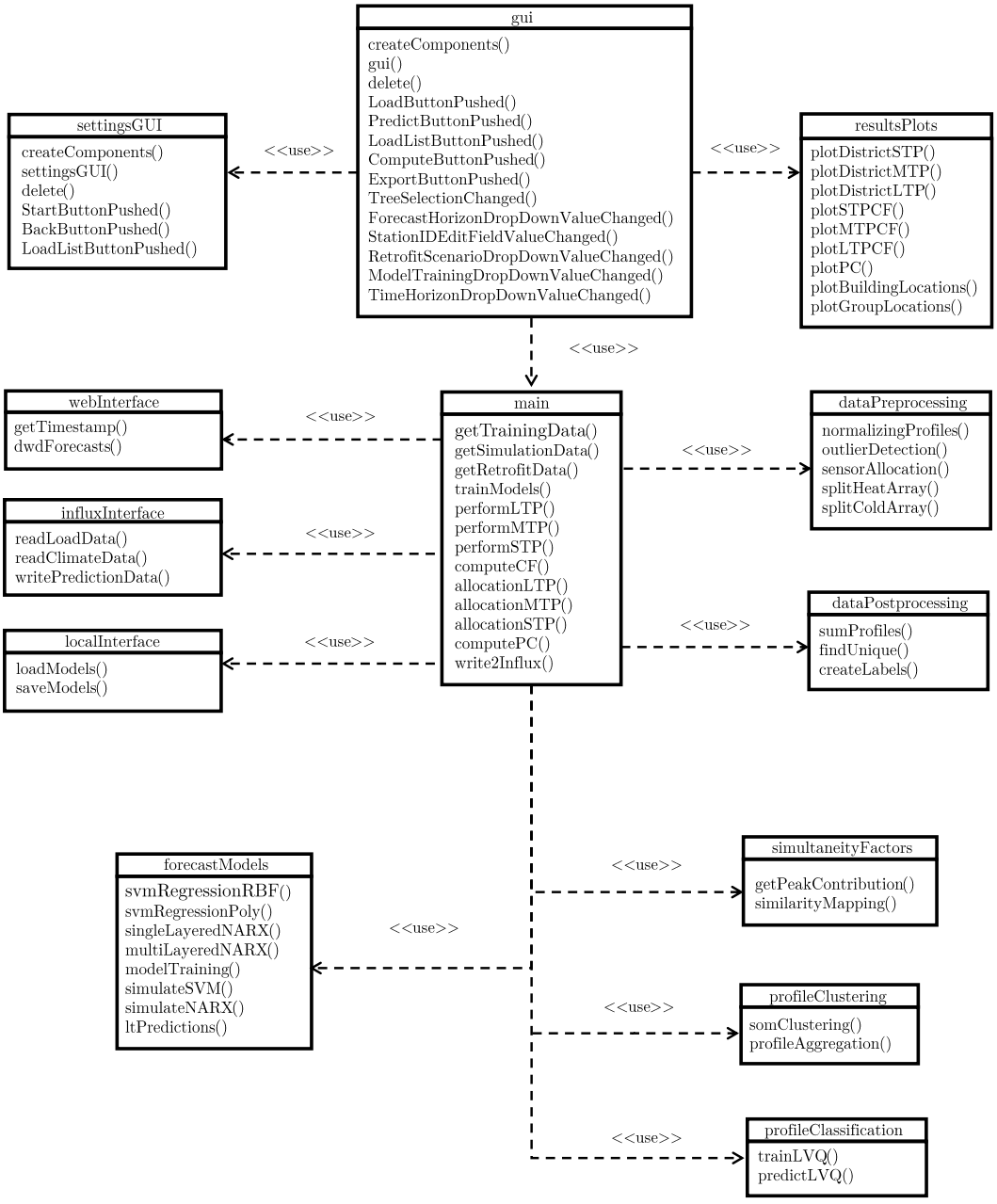


Abbildung B.1: Schematische Darstellung der Klassenstruktur von AMSA

B.2 Beschreibung der wichtigsten Klassen und Funktionen

Tabelle B.1: Auflistung und Beschreibung der Funktionen der Klasse *settingsGUI* von AMSA

Funktionsname	**Beschreibung**
createComponents()	Erzeugung der GUI-Elemente
settingsGUI()	Aufbau der grafischen Oberfläche
delete()	Schließen der grafischen Benutzeroberfläche
StartButtonPushed()	Aufruf der Haupt-GUI und Übergabe der Parameter; Sperren der Eingabefelder
BackButtonPushed()	Entsperren der Eingabefelder
LoadListButtonPushed()	Importieren der Excel-Dateien zu gebäudespezifischen Sensorbezeichnungen ín der Influx Datenbank (In Analogie dazu existieren Funktionen für den Import der weiteren Excel-Dateien)

Tabelle B.2: Auflistung und Beschreibung der Funktionen der Klasse *gui* von AMSA

Funktionsname	**Beschreibung**
createComponents()	Erzeugung der GUI-Elemente
gui()	Aufbau der grafischen Oberfläche
delete()	Schließen der grafischen Benutzeroberfläche
LoadButtonPushed()	Importieren der notwendigen Datengrundlage mittels der dazugehörigen Funktionen der Klasse *main*
PredictButtonPushed()	Prognose von gebäudespezifischen thermischen Lastprofillen mittels der dazugehörigen Funktionen der Klasse *main*
ComputeButtonPushed()	Quantifizierung von Gleichzeitigkeitseffekten mittels der dazugehörigen Funktionen der Klasse *main*
ExportButtonPushed()	Exportieren der Ergebnisse mittels der dazugehörigen Funktionen der Klasse *main*
TreeSelectionChanged()	Dynamische Anpassung der Baumstruktur in Abhängigkeit vom Berechnungsschritt; Aufruf der Funktionen der Klasse *resultsPlots*

Tabelle B.3: Auflistung und Beschreibung der Funktionen der Klasse *main* von AMSA

Funktionsname	**Beschreibung**
getTrainingData()	Importieren der Zeitreihen aus der Influx Datenbank; Importieren der generierten historischen Nutzungsprofile; Vorverarbeitung der Daten
getSimulationData()	Importieren des ausgewählten TRJ-Datensatzes für mittel- und langfristige bzw. Importieren von kurzfristigen Wettervorhersagen zur Generierung von kurzfristigen Prognosen; Importieren der generierten zukünftigen Nutzungsprofile
getRetrofitData()	Extrahierung von Parametern aus den importierten Sanierungsreihenfolgen
trainModels()	Trainieren des Prognose-Ensembles; Generierung der Modell-Container
performLTP()	Simulation der gebäudespezifisch ausgewählten Prognosemodelle zur langfristigen Prognose thermischer Lastprofile; bei Integration von Sanierungsreihenfolgen, Überprüfung, ob Gebäude saniert wird und gegebenenfalls entsprechende Lastgangmodifikationen; Identifikation von fehlgeschlagenen Lastprognosen
performMTP()	Simulation der gebäudespezifisch ausgewählten Prognosemodelle zur mittelfristigen Prognose thermischer Lastprofile; Identifikation von fehlgeschlagenen Lastprognose
performSTP()	Simulation der gebäudespezifisch ausgewählten Prognosemodelle zur kurzfristigen Prognose thermischer Lastprofile; Identifikation von fehlgeschlagenen Lastprognosen
computeCF()	Identifikation der größten Leistungsspitzen; Quantifizierung von zukünftigen und gegebenenfalls historischen Gleichzeitigkeitsfaktoren
allocationLTP()	Clustering der langfristig prognostizierten und gegebenenfalls historischen Lastprofile; Zuordnung des Lastprofils der Gesamtgruppe zu ähnlichstem Cluster im Rahmen einer Klassifikation
allocationMTP()	Clustering der mittelfristig prognostizierten und gegebenenfalls historischen Lastprofile; Zuordnung des Lastprofils der Gesamtgruppe zu ähnlichstem Cluster im Rahmen einer Klassifikation
allocationSTP()	Clustering der kurzfristig prognostizierten und gegebenenfalls historischen Lastprofile; Zuordnung des Lastprofils der Gesamtgruppe zu ähnlichstem Cluster im Rahmen einer Klassifikation
computePC()	Quantifizierung von Spitzenlastanteilen
write2Influx()	Exportieren der prognostizierten Lastprofile im Rahmen der Aktivierung der Schnittstelle zur Influx Datenbank

Tabelle B.4: Auflistung und Beschreibung der Funktionen der Klasse *forecastModels* von AMSA

Funktionsname	**Beschreibung**
svmRegressionRBF()	Konfiguration des SVM-R-Modells auf Grundlage eines RBF-Kernel
svmRegressionPoly()	Konfiguration des SVM-R-Modells auf Grundlage eines polynomialen Kernel dritten Grades
singleLayeredNARX()	Konfiguration des NARX RNN mit einer verdeckten Schicht
multiLayeredNARX()	Konfiguration des NARX RNN mit zwei verdeckten Schichten
modelTraining()	Trainieren der Prognosemodelle; Auswahl des gebäudespezifisch geeignetsten Prognosemodells aus dem Ensemble
simulateSVM()	Simulation von SVM-R-Modellen zur Lastprognose
simulateNARX()	Initialisierung und Simulation von NARX RNN zur Lastvorhersage
ltPredictions()	Auswahl der Simulationsfunktion in Abhängigkeit des gebäudespezifisch gewählten Prognosemodells

Tabelle B.5: Auflistung und Beschreibung der Funktionen der Klasse *profileClustering* von AMSA

Funktionsname	**Beschreibung**
somClustering()	Aufbau und Konfiguration der SOM
profileAggregation()	Identifizierung zeitraumübergreifender Cluster

Tabelle B.6: Auflistung und Beschreibung der Funktionen der Klasse *profileClassification* von AMSA

Funktionsname	**Beschreibung**
trainLVQ()	Konfiguration des LVQ-Modells
predictLVQ()	Simulation des kompetitiven neuronalen Netzwerkes zur Klassifikation von Lastprofilen

Tabelle B.7: Auflistung und Beschreibung der Funktionen der Klasse *webInterface* von AMSA

Funktionsname	Beschreibung
getTimestamp()	Ermittlung des Zeitstempels zum Zeitpunkt des Funktionsaufrufs
dwdForcasts()	Konfiguration und Herstellung der Verbindung zum DWD-Server; Importieren der aktuellen Wettervorhersagen des DWD; Datenvorverarbeitung für die Nutzung der Informationen in AMSA

Tabelle B.8: Auflistung und Beschreibung der Funktionen der Klasse *influxInterface* von AMSA

Funktionsname	Beschreibung
readLoadData()	Konfiguration und Herstellung der Verbindung zur Influx Datenbank; Importieren der gebäudebezogenen Lastgänge
readClimateData()	Konfiguration und Herstellung der Verbindung zur Influx Datenbank; Importieren der Wetterdatensätze
writePredictionData()	Konfiguration und Herstellung der Verbindung zur Influx Datenbank; Exportieren der gebäudebezogenen Lastgangprognosen in die Influx Datenbank

Tabelle B.9: Auflistung und Beschreibung der Funktionen der Klasse *localInterface* von AMSA

Funktionsname	Beschreibung
loadModels()	Laden der Modellcontainer für das Prognose-Ensemble
saveModels()	Speichern der Modellcontainer für das Prognose-Ensemble

Tabelle B.10: Auflistung und Beschreibung der Funktionen der Klasse *dataPreprocessing* von AMSA

Funktionsname	Beschreibung
normalizingProfiles()	Normalisierung der Lastprofile
outlierDetection()	Identifizierung und Eliminierung von Ausreißern in Datensätzen
sensorAllocation()	Zuordnung der gebäudespezifischen Sensornamen zu Gebäudekomplexen des Forschungszentrums Jülich

B.3 Datenformate und Kennzeichnungsschlüssel

Tabelle B.11: Auflistung der Komponenten des Kennzeichnungsschlüssels für Prognosezeitreihen in der Influx Datenbank

Element	**Option 1**	**Option 2**	**Option 3**
Betrachtungsebene	*District*	*Group*	*Building*
Name	*districtName*	*groupName*	*buildingName*
Energieart	*Heat*	*Cold*	-
Prognosehorizont	*STP*	*MTP*	*LTP*
Wetterszenario (optional)	*Trend*	*exSummer*	*exWinter*
Sanierungsszenario (optional)	*RF*	-	-

C Gebäudeinformationen

Tabelle C.1: Auflistung der betrachteten Gebäudekomplexe mit Angabe der Nutzung und Versorgungsart (1/3)

Gebäudenummer	Nutzungsart	Versorgungsart
1	Büro & Betrieb	Wärme & Kälte
2	Büro & Labor	Wärme & Kälte
3	Büro & Betrieb	Wärme & Kälte
4	Experimentierhalle	Wärme
5	Büro	Wärme & Kälte
6	Experimentierhalle	Wärme & Kälte
7	Büro	Wärme
8	Büro	Wärme
9	Büro	Wärme
10	Büro & Labor	Wärme
11	Labor	Wärme & Kälte
12	Labor	Wärme & Kälte
13	Labor	Wärme & Kälte
14	Versorgung	Wärme
15	Büro & Labor	Wärme & Kälte
16	Büro & Labor	Wärme & Kälte
17	Büro & Labor	Wärme & Kälte
18	Büro & Werkstatt	Wärme
19	Büro	Wärme
20	Büro+Werkstatt	Wärme
21	Labor	Wärme & Kälte
22	Experimentierhalle	Wärme & Kälte
23	Büro & Werkstatt	Wärme
24	Lager	Wärme
25	Büro & Labor	Wärme
26	Werkstatt & Lager	Wärme
27	Büro & Werkstatt	Wärme
28	Büro	Wärme
29	Sozial	Wärme & Kälte
30	Büro & Labor	Wärme
31	Büro	Wärme
32	Büro	Wärme
33	Büro & Labor	Wärme
34	Büro & Labor	Wärme
35	Büro	Wärme & Kälte
36	Büro & Lehre	Wärme
37	Büro & Labor	Wärme & Kälte
38	Büro	Wärme
39	Experimentierhalle	Wärme & Kälte
40	Büro & Labor	Wärme & Kälte

Tabelle C.2: Auflistung der betrachteten Gebäudekomplexe mit Angabe der Nutzung und Versorgungsart (2/3)

Gebäudenummer	Nutzungsart	Versorgungsart
41	Büro & Labor	Wärme & Kälte
42	Büro & Labor	Wärme & Kälte
43	Büro	Wärme
44	Büro	Wärme
45	Büro	Wärme
46	Büro & Labor	Wärme & Kälte
47	Büro & Labor	Wärme & Kälte
48	Labor	Wärme & Kälte
49	Labor	Wärme & Kälte
50	Büro & Labor	Wärme & Kälte
51	Experimentierhalle	Wärme & Kälte
52	Experimentierhalle	Wärme & Kälte
53	Labor	Wärme
54	Labor	Wärme
55	Betrieb	Wärme
56	Büro & Labor	Wärme & Kälte
57	Betrieb & Experimentierhalle	Wärme
58	Büro	Wärme
59	Experimentierhalle	Wärme & Kälte
60	Büro & Experimentierhalle	Wärme & Kälte
61	Büro & Labor	Wärme & Kälte
62	Büro	Wärme
63	Experimentierhalle	Wärme
64	Büro & Werkstatt	Wärme
65	Büro	Wärme
66	Büro & Werkstatt	Wärme
67	Garage	Wärme
68	Lager	Wärme
69	Lager	Wärme
70	Lager	Wärme
71	Entsorgung	Wärme
72	Büro	Wärme
73	Büro & Labor	Wärme
74	Lager	Wärme
75	Soziales	Wärme
76	Büro	Wärme
77	Büro	Wärme
78	Büro & Werkstatt	Wärme
79	Büro	Wärme
80	Büro & Labor	Wärme & Kälte

Tabelle C.3: Auflistung der betrachteten Gebäudekomplexe mit Angabe der Nutzung und Versorgungsart (3/3)

Gebäudenummer	Nutzungsart	Versorgungsart
81	Labor	Wärme & Kälte
82	Labor	Wärme & Kälte
83	Labor	Wärme & Kälte
84	Labor	Wärme & Kälte
85	Büro & Labor	Wärme & Kälte
86	Büro	Wärme & Kälte
87	Büro	Wärme & Kälte
88	Büro & Labor	Wärme & Kälte
89	Büro & Labor	Wärme & Kälte
90	Büro & Labor	Wärme & Kälte
91	Büro & Labor	Wärme & Kälte
92	Büro & Labor	Wärme & Kälte
93	Labor	Wärme & Kälte
94	Büro	Wärme & Kälte
95	Büro	Wärme & Kälte
96	Büro	Wärme & Kälte
97	Büro	Wärme
98	Büro & Labor	Wärme & Kälte
99	Labor	Wärme & Kälte
100	Soziales	Wärme
101	Soziales	Wärme

Tabelle C.4: Aufteilung des gebäudespezifischen Wärmebedarfs

Gebäudenummer	statischer Wärmebedarf [%]	dynamischer Wärmebedarf [%]
11	5,0	95,0
18	42,9	57,1
20	61,9	38,1
28	100,0	0,0
29	62,1	37,9
35	76,6	20,4
36	73,3	26,7
37	33,3	66,7
40	33,3	66,7
46	56,7	43,3
47	53,3	46,7
51	68,8	31,2
52	41,0	59,0
60	79,0	21,0
65	85,4	14,6
84	66,1	33,9
87	100,0	0,0
94	54,5	45,5
95	90,2	9,8
96	86,7	13,3
94	54,5	45,5

D Charakteristische Heizlasten

D.1 Thermische Leistungen in Abhängigkeit von der Außentemperatur

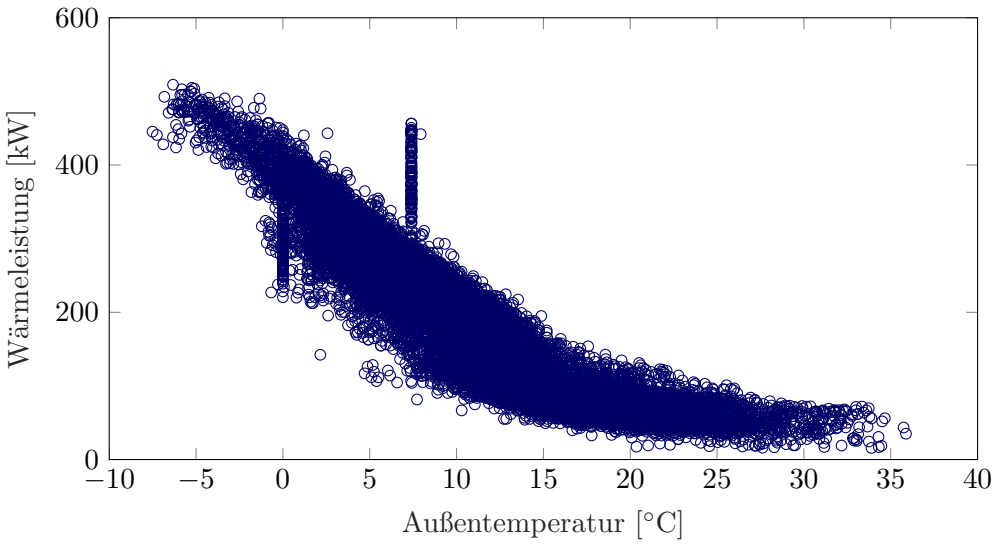


Abbildung D.1: Stündliche Heizlast eines Gebäudes in Abhängigkeit von der Außentemperatur mit einem witterungsabhängigen und witterungsunabhängigen Leistungsanteil bei einer Heizgrenztemperatur von circa 15°C

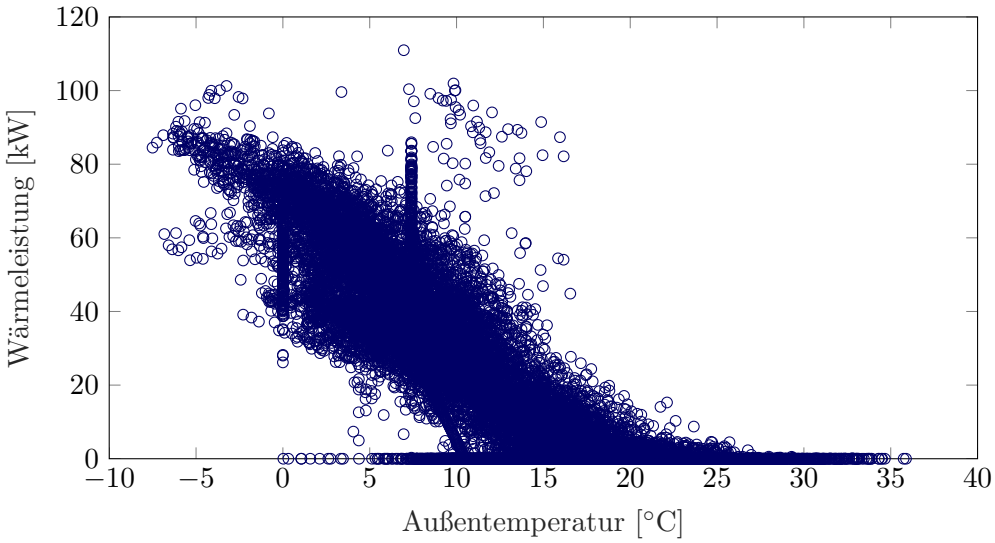


Abbildung D.2: Stündliche nahezu ausschließlich witterungsabhängige Gebäudeheizlast

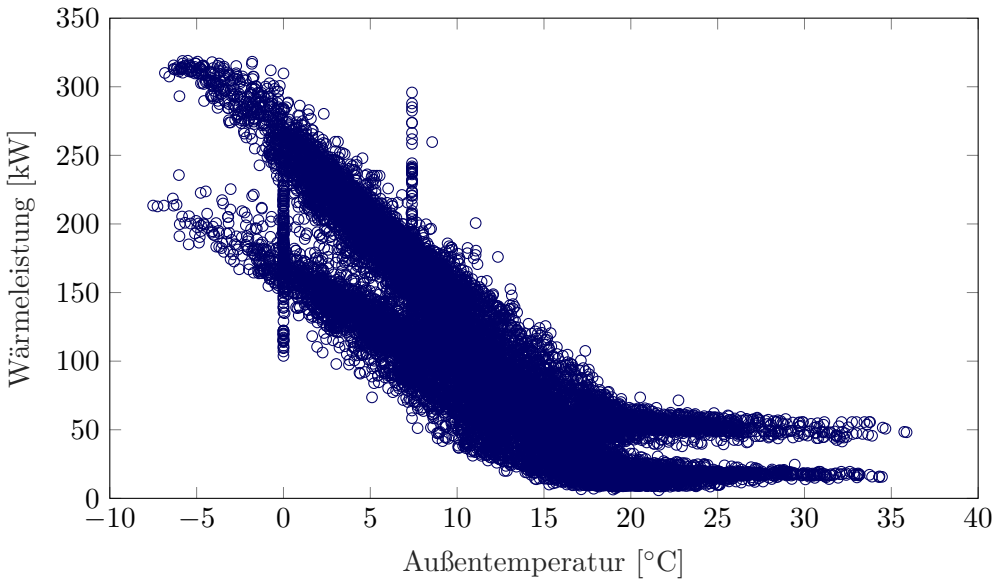


Abbildung D.3: Stündliche Heizlast eines Gebäudes in Abhängigkeit von der Außentemperatur mit einem witterungsabhängigen und witterungsunabhängigen Leistungsanteil bei einer Heizgrenztemperatur von circa 15°C mit Nacht- und Wochenendabsenkung (in Anlehnung an Frederiksen et al., 2014)

D.2 Ergänzende Ergebnisse der Gleichzeitigkeitsanalyse

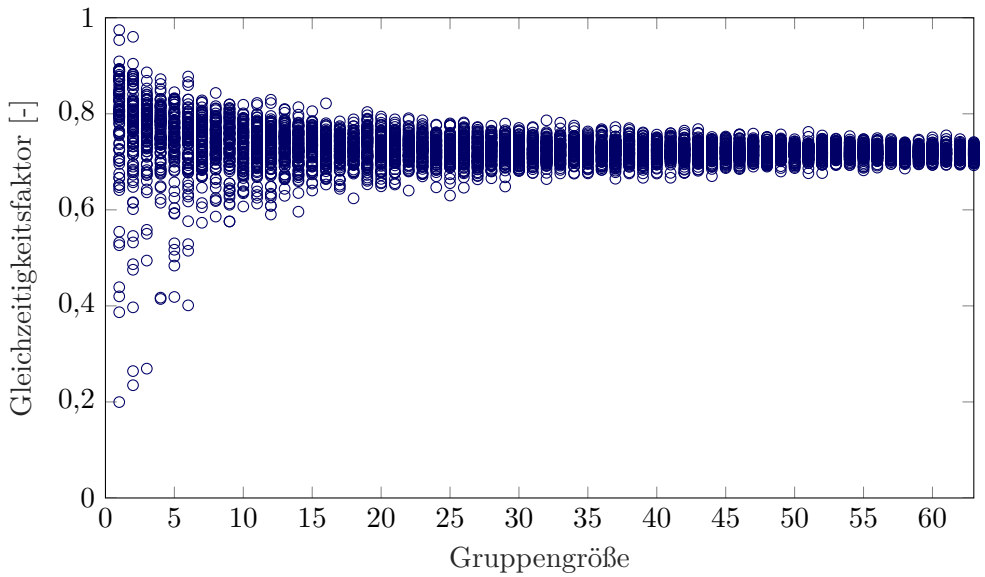


Abbildung D.4: Ergebnis der statistischen Analyse zur Entwicklung des Gleichzeitigkeitsfaktors für die Wärmeversorgung in Abhängigkeit von der Abnehmergruppengröße bei einer gebäudebezogenen Norminnentemperatur von 22 °C

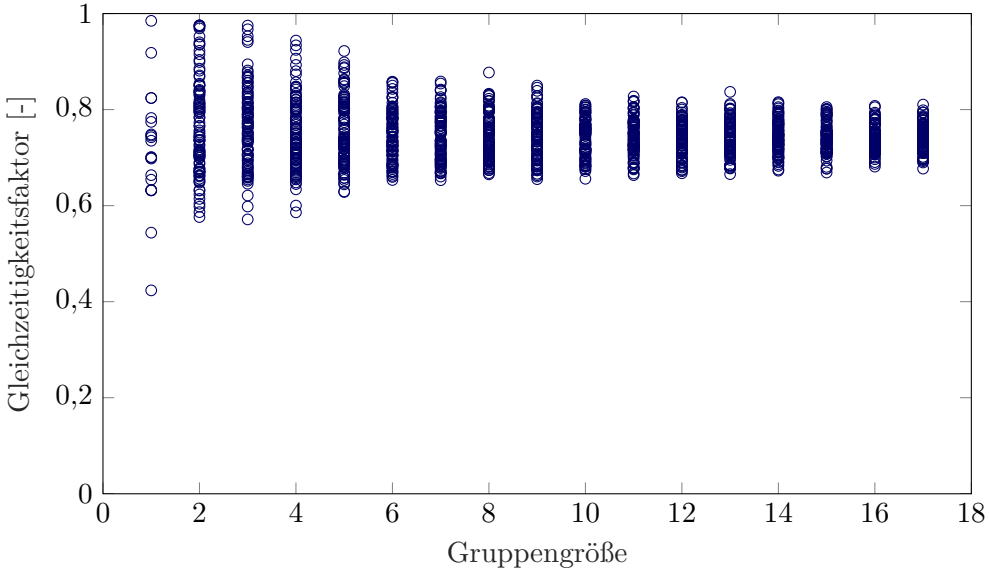


Abbildung D.5: Ergebnis der statistischen Analyse zur Entwicklung des Gleichzeitigkeitsfaktors für die Kälteversorgung in Abhängigkeit von der Abnehmergruppengröße bei einer gebäudebezogenen Auslegungstemperatur von 35 °C

E Ergänzende Prognoseergebnisse

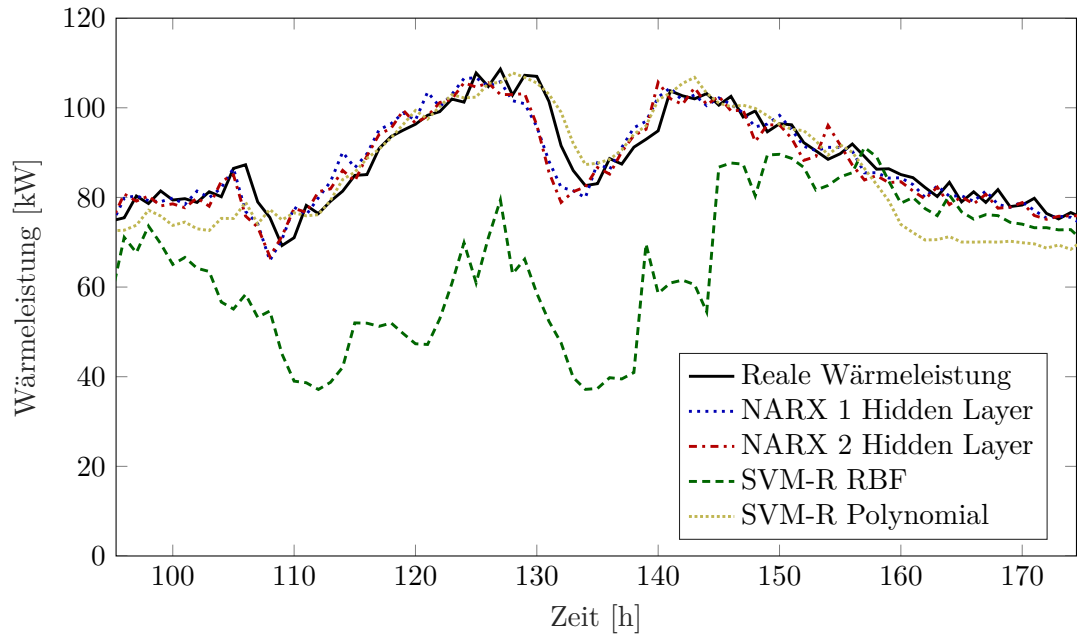


Abbildung E.1: Exemplarischer Auszug aus den Ergebnissen der saisonalen Heizlastprognosen mit NARX RNN und ε-SVM-R Modellen (Koschwitz et al., 2018)

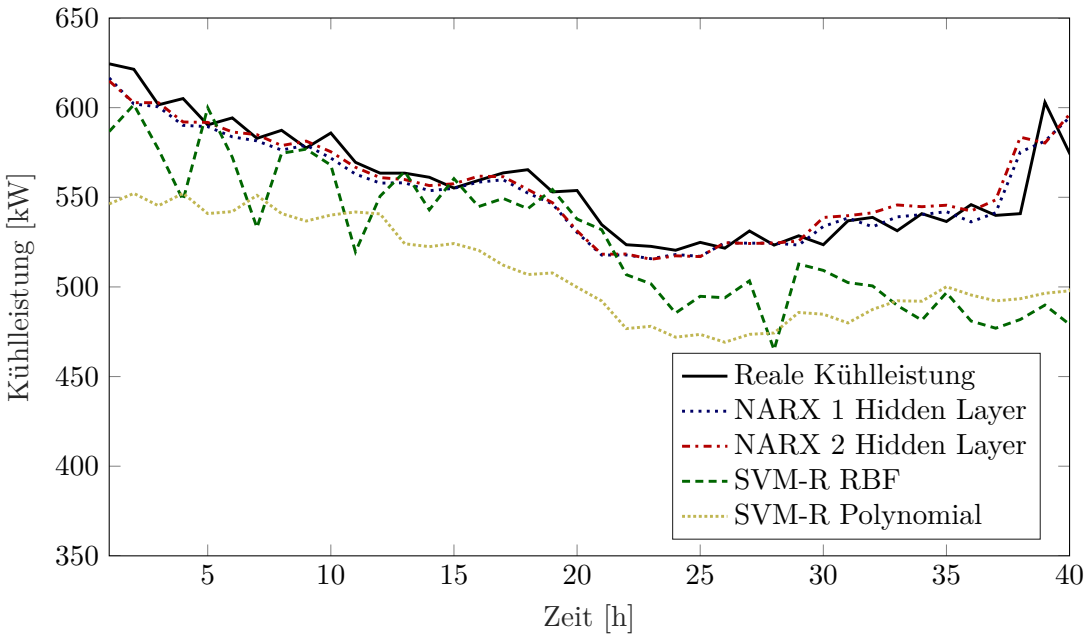


Abbildung E.2: Exemplarischer Auszug aus den Ergebnissen der saisonalen Kühllastprognosen mit NARX RNN und ε-SVM-R Modellen (Koschwitz et al., 2018)