



**TECHNISCHE  
UNIVERSITÄT  
DRESDEN**

---

Fakultät Bauingenieurwesen Institut für Bauinformatik

---

## **Analyse von Betriebsdaten Raumluftechnischer Anlagen**

Diplomarbeit von

Markus Bär

Dresden, den 31. Mai 2018



## **Aufgabenstellung für die Diplomarbeit**

Name: cand. Ing. Markus Bär

Vertiefung: Computational Engineering

---

**Thema: Analyse von Betriebsdaten Raumluftechnischer Anlagen**  
(Analysis of operating data of HVAC systems)

### **Zielsetzung:**

In vielen Gebäuden werden heute sogenannte Raumluftechnische Anlagen (RLT-Anlagen) eingesetzt, um die Raumluf zu konditionieren, d. h. beispielsweise die gewünschte Temperatur, Feuchte und CO<sub>2</sub>-Gehalt einzustellen. Die Anforderungen an die Konditionierung schwanken bei einem Gebäude mit dem Bedarf der Nutzer und den Wetterbedingungen, was eine Steuerung oder Regelung der RLT-Anlagen notwendig macht. In der Betriebspraxis zeigen sich an diesen Systemen oft Fehler, die sich durch erhöhten Energieverbrauch der Anlagen und/oder durch Verletzung der Nutzeranforderungen bemerkbar macht. Während letzteres aufgrund von Beschwerden der Nutzer meist bemerkt wird, bleiben Fehler, die lediglich erhöhten Energieverbrauch verursachen, oft längere Zeit unbemerkt, weil die Analyse der Betriebsdaten der Gebäudeleittechnik oft manuell durch den Nutzer erfolgen müsste, was sehr zeitaufwendig ist und vertieftes Wissen über RLT-Anlagen erfordert.

Am Fraunhofer-Institut für Integrierte Schaltungen / Institutsteil Entwurf Adaptiver Systeme (IIS-EAS) werden Ansätze verfolgt, diese Analysen mindestens teilweise zu automatisieren. Die Aufgabe besteht darin, einem Ansatz von Habib (2016) folgend verschiedene Betriebszustände einer RLT-Anlage automatisch zu definieren bzw. zu klassifizieren und zu untersuchen, ob und wie die gefundenen Zustände sinnvollen Betriebszuständen bzw. problematischen Zuständen zugeordnet werden können.

### **Arbeitsplan:**

1. Bewertung des Standes in Forschung und Technik auf dem Gebiet der automatischen Klassifikation des Betriebs von Systemen der technischen Gebäudeausrüstung, insbesondere von RLT-Anlagen
2. Übertragung des von Habib vorgestellten Verfahrens auf die Betriebszustandsklassifikation von RLT-Anlagen
3. Umsetzung
  - Implementierung der statistischen Verfahren in einer geeigneten Programmiersprache, z.B. MatLab oder Python
  - Anwendung der Verfahren auf Betriebsdaten von realen Anlagen. Die Daten werden seitens EAS zur Verfügung gestellt
4. Interpretation der Ergebnisse und Zuordnung der gefundenen Klassen von Betriebszuständen auf sinnvolle bzw. energetisch ungünstige Fälle



5. Zusammenfassung und Bewertung der Ergebnisse, Ausblick auf Möglichkeiten der Weiterentwicklung

Referenz:

Habib, U.; Hayat, K.; Zucker, G.: Complex building's energy system operation patterns analysis using bag of words representation with hierarchical clustering, Complex Adaptive Systems modelling, (2016) 4:8

Wiss. Betreuer TU Dresden: Prof. Dr.-Ing. Raimar Scherer  
Dipl.-Math. Tom Grille

Betriebl. Betreuer: Dr. Andreas Wilde

Beteiligtes Unternehmen: Fraunhofer Institut für Integrierte Schaltungen

ausgehändigt am: *18.01.2018*

einzureichen am: *18.05.2018*

  
Prof. Dr.-Ing. Raimar Scherer  
Verantwortlicher Hochschullehrer

## Inhaltsverzeichnis

Abbildungsverzeichnis .....	V
Tabellenverzeichnis .....	VIII
Formelverzeichnis.....	IX
Abkürzungsverzeichnis .....	X
<b>1 Einleitung .....</b>	<b>1</b>
1.1 Problemstellung der Arbeit .....	1
1.2 Ziele und Abgrenzung der Arbeit.....	1
1.3 Aufbau der Arbeit .....	2
<b>2 Theoretische Grundlagen.....</b>	<b>4</b>
2.1 Gebäudeleittechnik.....	4
2.1.1 Grundbegriffe .....	4
2.1.2 Erläuterung ausschlaggebender Anlagensystemkomponenten .....	5
2.1.3 Approximierte Wärmeleistung .....	7
2.2 Verwendete statistische Verfahren .....	7
2.2.1 Ablauf der Analyse .....	8
2.2.2 Data Preprocessing & Sanitation.....	9
2.2.3 Normierungsverfahren.....	10
2.2.4 Clusteranalyseverfahren .....	11
2.2.5 k-Means Clusteranalyse .....	11
2.2.6 Hierarchische Clusteranalyse .....	13
2.2.7 Evaluation der gefundenen Cluster mit der GAP-Statistic.....	14
2.2.8 Symbolic Aggregate Approximation.....	14
2.2.9 Bag-of-Words Representation .....	15
<b>3 Umsetzung .....</b>	<b>17</b>
3.1 Implementierung der verwendeten Algorithmen mit KNIME.....	17
3.2 Zu untersuchende Anlagensysteme .....	18
3.2.1 Täglicher Betriebszeitraum der raumluftechnischen Anlagen .....	19
3.2.2 Lüftungsanlage 06.....	19
3.2.3 Lüftungsanlage 03.....	23
3.3 Übertragung des Datenanalyseverfahrens auf raumluftechnische Anlagen ....	25
3.3.1 Unterscheidung von AN- & AUS-Zuständen .....	25

3.3.2	Umsetzung Symbolic Aggregate Approximation und Bag of Words Representation .....	26
3.3.3	Umsetzung Hierarchisches Clustering.....	27
3.3.4	Umsetzung der Cluster Evaluation der hierarchischen Clusteranalyse.....	27
3.4	Zu erwartende Betriebszustände.....	30
<b>4</b>	<b>Ergebnisse der Datenanalyse &amp; Diskussion.....</b>	<b>32</b>
4.1	Analyse und Auswertung von Lüftungsanlage 06 .....	32
4.1.1	Unterscheidung von AN- und AUS-Zyklen.....	32
4.1.2	Klassifizierung der Betriebszustände.....	35
4.1.3	Outlierproblematik .....	36
4.1.4	GAP-Statistic zu Cluster Evaluation .....	37
4.1.5	Einführung des distanzbasierten Schwellenwertes .....	37
4.1.6	Untersuchung der Auswirkung unterschiedlicher Chunk-Anzahlen.....	40
4.1.7	Zuordnung von Betriebszuständen zum hierarchischen Clustering I .....	42
4.1.8	HC I - Darstellung der Ergebnisse anhand ausschlaggebender Daten .....	44
4.1.9	Einführung der hierarchischen Clusteranalyse II .....	48
4.1.10	HC II - Darstellung der Ergebnisse anhand ausschlaggebender Daten .....	50
4.1.11	Darstellung der Auswirkung der Außentemperatur auf das Ergebnis der Clusteranalyse .....	54
4.2	Analyse und Auswertung von Lüftungsanlage 03 .....	57
4.2.1	Unterscheidung von AN- und AUS-Zyklen.....	57
4.2.2	Klassifizierung der Betriebszustände.....	59
4.2.3	Zuordnung von Betriebszuständen zum hierarchischen Clustering der Lüftungsanlage 03 .....	60
4.2.4	Darstellung der Ergebnisse anhand ausschlaggebender Daten .....	61
4.3	Vergleich der Ergebnisse von Lüftungsanlage 06 und Lüftungsanlage 03.....	65
<b>5</b>	<b>Schlussbetrachtung .....</b>	<b>68</b>
5.1	Zusammenfassung.....	68
5.2	Ergebnisse der Arbeit.....	69
5.3	Ausblick.....	71
	Literaturverzeichnis.....	72
	Anlagenverzeichnis.....	74
	Erklärung zur selbständigen Abfassung der Diplomarbeit .....	75

## Abbildungsverzeichnis

Abbildung 1: schematische Darstellung unterschiedlicher Luftarten.....	5
Abbildung 2: Darstellung eines Kreislaufverbundsystems zur Wärmerückgewinnung ...	6
Abbildung 3: Schrittweiser Iterationsprozess des k-Means Algorithmus.....	12
Abbildung 4: Beispielhafte Darstellung eines Dendrogramms.....	13
Abbildung 5: Piecewise Aggregate Approximation für zwei Zeitreihen ts1 & ts2 .....	14
Abbildung 6: Beispielhafte Illustration einer SAX Transformation anhand einer normierten Zeitreihe [Habib et al, 2016, S. 10, Abbildung 3] .....	15
Abbildung 7: Darstellung des Übergangs verschiedener Betriebszyklen in die Bag or Words Repräsentation und schließlich zum Dendrogram des hierarchischen Clustering [Habib et al., 2015, S 308, Abbildung 7].....	16
Abbildung 8: Beispielhafte Darstellung eines Workflows für das k-Means-Clusteranalyse .....	18
Abbildung 9: Anlagenschema "Lüftungsanlage 06".....	19
Abbildung 10: Anlagenschema Lüftungsanlage 03 .....	23
Abbildung 11: RLT06 Darstellung des Abluftvolumenstroms über den Jahresverlauf mit dem Ergebnis der Clusteranalyse in verschiedenen Farben .....	33
Abbildung 12: Darstellung von Outlier-Problematik im hierarchischen Clustering anhand RLT 06.....	36
Abbildung 13: Darstellung verschiedener Schwellenwerte für das Clustering .....	37
Abbildung 14: Cluster - Relationen bei HC I mit 120 über 60 Chunks bei einem Schwellenwert von 40 %.....	41
Abbildung 15: Dendrogramm von HC I bei 120 Chunks mit Kennzeichnung des 40 % Schwellenwerts und der Nummerierung der Cluster .....	43
Abbildung 16: Cluster Heatmap für HC I mit 120 Chunks und Schwellenwert von 40 % .....	44
Abbildung 17: HC I Clustering Ergebnis anhand der mittleren Außentemperatur je Zyklus .....	45
Abbildung 18: HC I Clustering Ergebnis anhand der mittleren LE-Ventilstellung je Zyklus .....	46
Abbildung 19: HC I Clustering Ergebnis anhand der mittleren Betriebsstufe des Luftkühlers je Zyklus.....	47
Abbildung 20: HC I Clustering Ergebnis anhand der mittleren WRG-Ventilstellung je Zyklus .....	48
Abbildung 21 : Dendrogramm Hierarchical Clustering II mit 120 Chunks und 40 % Schwellenwert .....	49

<b>Abbildung 22: Cluster Heatmap HC II - CH 120 - Thr.: 40%</b> .....	50
Abbildung 23: RLT06 - HC II - mittlere Außentemperatur je Zyklus.....	51
Abbildung 24: RLT06 - HC II - mittlere LE-Ventilstellung je Zyklus.....	52
Abbildung 25: RLT 06 - HC II - mittlere Betriebsstufe Luftkühler je Zyklus .....	53
Abbildung 26: RLT06 - HC II - mittlere WRG-Ventilstellung je Zyklus.....	53
Abbildung 27: Darstellung der Cluster-Relation von HC I zu HC II .....	54
Abbildung 28: Vergleich der Ergebnisse der Clusteranalysen HC I und HC II anhand der mittleren Außentemperatur je Zyklus .....	55
Abbildung 29: Vergleich der Ergebnisse der Clusteranalysen HC I und HC II anhand der mittleren LE-Ventilstellung je Zyklus .....	56
Abbildung 30: Vergleich der Ergebnisse der Clusteranalysen HC I und HC II anhand der mittleren Betriebsstufe des Luftkühlers je Zyklus .....	56
Abbildung 31: Vergleich der Ergebnisse der Clusteranalysen HC I und HC II anhand der mittleren WRG-Stellung je Zyklus .....	57
Abbildung 32: RLT03 - Dendrogramm .....	60
Abbildung 33: RLT03 Cluster Heatmap über den gesamten Betrachtungszeitraum ....	61
Abbildung 34: RLT03 - Ergebnis des Clustering anhand der mittleren Außentemperatur .....	62
Abbildung 35: RLT03 - Ergebnis des Clustering anhand der mittleren LE-Ventilstellung .....	63
Abbildung 36: RLT03 - Ergebnis des Clustering anhand des mittleren Kältezählerwertes .....	64
Abbildung 37: RLT03 - Ergebnis des Clustering anhand der mittleren WRG- Ventilstellung .....	65
Abbildung 38: Vergleich der Ergebnisse der automatisierten Betriebszustandsklassifikationen von Lüftungsanlage 06 HC I und Lüftungsanlage 03 nach der Entfernung fehlerhafter Betriebszyklen anhand der mittleren Außentemperatur je Betriebszyklus .....	65
Abbildung 39: Vergleich der Ergebnisse der automatisierten Betriebszustandsklassifikationen von Lüftungsanlage 06 HC I und Lüftungsanlage 03 nach der Entfernung fehlerhafter Betriebszyklen anhand der mittleren LE- Ventilstellung je Betriebszyklus.....	66
Abbildung 40: Vergleich der Ergebnisse der automatisierten Betriebszustandsklassifikationen von Lüftungsanlage 06 HC I und Lüftungsanlage 03 nach der Entfernung fehlerhafter Betriebszyklen anhand der mittleren Betriebsstufe Luftkühler je Betriebszyklus.....	66
Abbildung 41: Vergleich der Ergebnisse der automatisierten Betriebszustandsklassifikationen von Lüftungsanlage 06 HC I und Lüftungsanlage	

03 nach der Entfernung fehlerhafter Betriebszyklen anhand der mittleren WRG- je  
Betriebszyklus ..... 67

## Tabellenverzeichnis

Tabelle 1: Beispielhafte Darstellung von AN-Zyklen in Form von SAX-Symbolen (in Anlehnung an [Habib et al., 2015, S. 306, Tabelle 2]) .....	8
Tabelle 2: Aufzählung und Erläuterung verwendeter Datenpunkte der Anlage 06.....	22
Tabelle 3: Aufzählung und Erläuterung verwendeter Datenpunkte der Anlage 03.....	24
Tabelle 4: Erläuterung der Kontrollwerte der Evaluationstabelle .....	30
Tabelle 5: zu erwartende Betriebszustände der raumluftechnischen Anlagensysteme .....	31
Tabelle 6: RLT06 k-Means Szenario 1 Vergleich der Ergebnisse mit verschiedenen Normierungsmethoden .....	34
Tabelle 7: RLT06 k-Means Szenario 1 Vergleich der Ergebnisse mit verschiedenen Normierungsmethoden .....	35
Tabelle 8: Verwendete Datenpunkte für die hierarchische Clusteranalyse I bei RLT06 .....	36
Tabelle 9: RLT06 Ausschnitt aus der Evaluationstabelle zur Darstellung von Cluster 2 und Cluster 3 für HC I bei 120 Chunks und einem Schwellenwert von 40 %.....	38
Tabelle 10: RLT06 Evaluationstabelle für HC I mit 120 Chunks & Schwellenwert 30 % für die Cluster 22 bis 25.....	39
Tabelle 11: maximale euklidische Distanz und Gesamtanzahl gefundener Cluster für verschiedene Chunk-Anzahlen bei Datenpunktauswahl HC I und einem Schwellenwert von 40 %.....	40
Tabelle 12: Übereinstimmende Relationen aus Clusteranalysen mit unterschiedlichen Chunk-Anzahlen .....	42
Tabelle 13: HC I - Darstellung gefundener Cluster mit ausschlaggebenden Evaluationswerten und Zuordnung des zu erwartenden Anlagenverhaltens. ....	43
Tabelle 14: RLT06 – HC II – Datenpunktauswahl .....	48
Tabelle 15: HC II - Darstellung gefundener Cluster mit ausschlaggebenden Evaluationswerten und Zuordnung des zu erwartenden Anlagenverhaltens .....	50
Tabelle 16: Identische Cluster aus Vergleich HC I und HC II .....	55
Tabelle 17: RLT03 k-Means-Clustering Ergebnisse mit verschiedenen Normierungsmethoden .....	58
Tabelle 18: RLT03 – HC I – Datenpunktauswahl .....	59
Tabelle 19: Ergebnis der Betriebszustandsklassifikation von RLT03.....	61

**Formelverzeichnis**

- (1) Wärmeleistung  $Q$
- (2) Approximation der Wärmeleistung  $P_w$
- (3) z-Score-Normierungsmethode
- (4) MINMAX-Normierungsmethode
- (5) Quadratischer Fehler  $J(C_k)$  bei Cluster  $C_k$
- (6) Quadratischer Fehler  $J(C)$  bei  $K$  Clustern

## Abkürzungsverzeichnis

AB	Abluft
AU	Außenluft
BoWR	Bag of Words Representation
CoP	Coefficient of Performance
FO	Fortluft
KVS	Kreislaufverbundsystem
LE	Luftheritzer
LK	Luftkühler
PAA	Piecewise Aggregate Approximation
RLT	Raumluftechnische Anlage
RLT03	Lüftungsanlage 03
RLT06	Lüftungsanlage 06
SAX	Symbolic Aggregate Approximation
WRG	Wärmerückgewinnung
ZU	Zuluft

# 1 Einleitung

## 1.1 Problemstellung der Arbeit

Ein effizienter Gebäudebetrieb ist heutzutage wichtiger denn je, wofür sowohl ökologische als auch ökonomische Faktoren sprechen. Der Energieverbrauch in Deutschland wird zu einem nicht unerheblichen Teil durch die Nutzung von Gebäuden verursacht, wodurch nicht unerhebliche Energiekosten zum einen und ein nicht zu verkennender ökologischer Fußabdruck zum anderen entstehen. Der effiziente Betrieb von gebäudeenergie-technischen Anlagen ist hierbei ein guter Ansatzpunkt, um durch ein optimiertes Anlagenverhalten Energieverbräuche zu minimieren und somit die entstehenden Kosten als auch die Belastung der Umwelt gering zu halten. Zudem soll durch einen korrekten Anlagenbetrieb ein bestmöglicher Nutzerkomfort ermöglicht werden.

Da es sich bei Anlagen der Gebäudeleittechnik vornehmlich um komplexe adaptive Systeme handelt, ist die korrekte Nutzung selbiger nicht trivial. [Habib et al., 2016, S. 4] Die notwendigen Analysen des Anlagenverhaltens entsprechend dem aktuellen Stand von Forschung und Technik sind derzeit mit einem nicht unerheblichen händischen Arbeitsaufwand verbunden. [Habib et al., 2016, S. 2] Zudem wird dafür auch ein profundes Verständnis der jeweiligen Vorgänge in der Anlage vorausgesetzt. Im Rahmen der industrienahe Forschung des Fraunhofer-Instituts für Integrierte-Schaltungen – Instituts-Teil Entwurf adaptiver Systeme in Dresden werden verschiedene Lösungsansätze entwickelt, um die Anlagensysteme automatisiert zu analysieren und zu optimieren.

## 1.2 Ziele und Abgrenzung der Arbeit

Im Rahmen der vorliegenden Diplomarbeit soll der von [Habib et al., 2016] entworfene Ansatz zur automatischen Erkennung von Betriebsmustern von Gebäudeenergiesystemen auf dessen Übertragbarkeit auf raumluftechnische Anlagen (RLT) untersucht werden. Das grundlegende Ziel des von Habib vorgestellten Verfahrens besteht darin, Analysealgorithmen bereitzustellen, welche automatisch verschiedene Betriebsmuster von Energiesystemen in Gebäuden klassifizieren. Die Evaluation der Ergebnisse der Datenanalyse nach [Habib et al., 2016] erfolgt unter Einbezug professioneller Expertise. Betrachtet man die aktuelle Vorgehensweise bei der Analyse von Gebäudebetriebsdaten lässt sich feststellen, dass das Verständnis von komplexen adaptiven Systemen [Habib et al., 2016, S. 3] wie der Gebäudeleittechnik einen nicht unerheblichen Arbeitsaufwand und ein entsprechendes technisches Verständnis voraussetzt. Außerdem besteht durch die Menge an Daten immer die Gefahr, dass bestimmte Muster in den Daten übersehen werden können. [Habib et al., 2016, S. 2]

Hierfür liefern Habib et al. einen vielversprechenden Ansatz, der eine weitgehend automatische Analyse von Gebäudebetriebszuständen ermöglicht und dabei mit einer minimalen Expertise bezüglich der Gebäudeleittechnik auskommt. Dies geschieht unter dem Einsatz verschiedener Algorithmen, die entsprechend des Forschungsberichts von Habib et al. per se ohne ein Verständnis der Anlagentechnik auskommen. [Habib et al., 2016, S. 9]

Ziel der Diplomarbeit ist es, das von Habib et al. gezeigte Verfahren auf raumluftechnische Anlagen zu übertragen und somit Betriebszustände zunächst automatisiert zu klassifizieren. Neben der Klassifizierung soll aufgezeigt werden, welchen Einfluss für die Algorithmen gewählte Parameter auf die Klassifikation der Betriebszustände haben. Die Betriebszustandsklassifikation soll anhand geeigneter Darstellung der Eingangsdaten verifiziert werden. Ferner soll untersucht werden, ob sich der Untersuchungsansatz auf Anlagensysteme, welche ein Fehlverhalten aufweisen, übertragen lässt. Dazu soll der Vergleich der zu untersuchenden Anlagensysteme anhand der Ergebnisse der Datenanalyse vorgenommen werden.

### 1.3 Aufbau der Arbeit

Im Folgenden soll dem Leser ein Überblick über die theoretischen Grundlagen der Arbeit vermittelt werden. Dabei wird zunächst das zum Verständnis der Arbeit notwendige Hintergrundwissen bezüglich der Gebäudeenergie-technik vermittelt. Im zweiten Teil des Kapitels soll die Grundidee des von Habib vorgestellten Ansatzes erläutert werden. Zudem sollen die einzelnen, zur Implementierung verwendeten Algorithmen vorgestellt werden. Im dritten Kapitel ist die Umsetzung der automatisierten Klassifikation von Betriebszuständen raumluftechnischer Anlagen aufgeführt. Zunächst wird das im Rahmen der Arbeit verwendete statistische Analysetool, der Konstanz Information Miner, vorgestellt. Im nächsten Schritt werden die für die Datenanalyse gewählten Anlagensysteme beschrieben. Neben der Darstellung des Anlagenschemas werden die Zusammenhänge der einzelnen Anlagenkomponenten erläutert. Außerdem werden die in der Untersuchung der Anlage verwendeten Sensoren aufgeführt und erklärt. Zudem wird in Kapitel 3 dargestellt, inwiefern der von [Habib et al., 2016] verwendete Analyseansatz auf raumluftechnische Anlagen übertragen wird. Dabei wird auf die exakte Verwendung der notwendigen Analysealgorithmen eingegangen und erläutert, wie die gefundenen Betriebszustände anhand eigens bestimmter Evaluationswerte beurteilt werden können.

Im letzten Teil von Kapitel drei soll aufgezeigt werden, welche Betriebszustände bei den im Rahmen der Diplomarbeit untersuchten Lüftungsanlagen zu erwarten sind.

In Kapitel vier sollen die Ergebnisse der Datenanalyse aufgeführt und diskutiert werden. Dabei wird zuerst Lüftungsanlage 06 (RLT06) analysiert. Anhand dieser Anlage werden die unterschiedlichen Parameter der Analyse hinsichtlich ihrer Auswirkung auf das Ergebnis diskutiert. Es wird zudem gezeigt, welche Parameter letztendlich genutzt wurden, um die Betriebszustände der Lüftungsanlage zu klassifizieren. Die gefundenen Betriebs-szenarien werden durch geeignete Darstellung der Eingangsdaten verifiziert.

Bei der darauffolgenden Untersuchung des zweiten Anlagensystems Lüftungsanlage 03 (RLT03) werden die bei der ersten Lüftungsanlage bewährten Parameter verwendet werden. Im Vorfeld der Datenanalyse ist die Anlage durch einen fehlerhaften Betrieb auffällig geworden. Deswegen wird sie verwendet, um zu untersuchen, inwiefern die Analysealgorithmen für eine Anlage mit fehlerhaftem Betrieb genutzt werden können. Die Ergebnisse werden wiederum unter Darstellung geeigneter Eingangsdaten bestätigt. Den Abschluss der Arbeit bilden eine inhaltliche Zusammenfassung und die Darstellung der Ergebnisse der Arbeit. Letztere beinhaltet neben den gewonnenen Erkenntnissen eine kritische Beurteilung der Universalität des gewählten Verfahrens und eine Kritik an der Umsetzbarkeit der Implementierung unter Verwendung des Konstanz Information Miner.

In Kapitel fünf wird die Schlussbetrachtung der angefertigten Diplomarbeit dargestellt. Dabei wird eine Zusammenfassung der Arbeit gegeben, die Ergebnisse retrospektiv dargestellt und auf Möglichkeiten zur Weiterentwicklung hingewiesen.

## 2 Theoretische Grundlagen

### 2.1 Gebäudeleittechnik

#### 2.1.1 Grundbegriffe

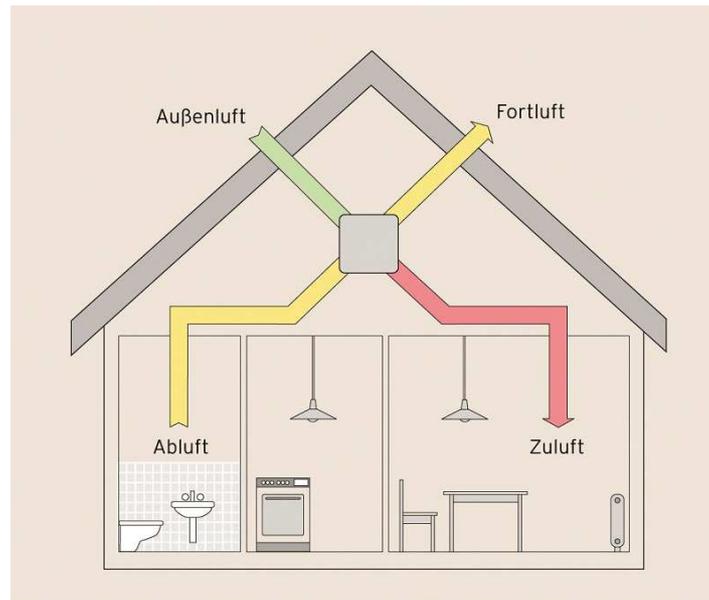
Im Folgenden sollen grundlegende Begriffe der Gebäudeenergie-technik in Bezug auf raumluftechnische Anlagen erläutert werden.

Lüftungsanlagen sind essentielle Bestandteile vieler Gebäude. Ihre Nutzung verfolgt eine Reihe von Zielstellungen, welche für den Nutzerkomfort des Gebäudes essentiell sind. Zum einen sollen sie einer Verunreinigung der Luft durch die Zufuhr von Außenluft (AU) entgegenwirken. Zudem sollen sie Raumlasten verringern und ein Mindestmaß an Raumdurchströmung garantieren oder für den im jeweiligen Gebäude erforderlichen Schutzdruck gewährleisten. Außerdem unterstützen sie den Erhalt der Gebäudesubstanz. [Schmidt et al., 2011, S. 76 - 78]

Generell kann beim Betrieb von Systemen der Gebäudeleittechnik zwischen einem Heiz- und einem Kühlbetrieb unterschieden werden. Beiden Betriebsszenarien liegen unterschiedliche Regelstrategien zugrunde, welche dazu dienen, den Anforderungen an die klimatischen Bedingungen im Gebäude gerecht zu werden. Zudem muss bei der Analyse der Systeme in Betracht gezogen werden, dass diese entsprechend der Nutzungszeiten des jeweiligen Gebäudes nur dann betrieben werden, wenn ein Klimatisierungsbedarf besteht.

Neben unterschiedlichen Anlagenbetriebsarten können auch verschiedene Luftarten unterschieden werden. In der Gebäudetechnik kann der Begriff Zuluft (ZU) auf unterschiedlich verstanden werden. Zum einen bezieht sich der Begriff auf die Luft, welche aus der Aufbereitungsanlage austritt und vor Eintritt in den zu klimatisierenden Raum das Kanalnetz der raumluftechnischen Anlage durchqueren muss. Die zweite Definition der Zuluft ist die der Luft, welche direkt in den zu klimatisierenden Raum einströmt. Eine Unterscheidung ist notwendig, weil sich die charakteristischen Eigenschaften der Luft beim Durchströmen des Kanalnetzes ändern können. [Schmidt et al., 2011, S. 74 - 75] Im Folgenden wird die Zuluft als die Luft verstanden werden, welche direkt in den zu klimatisierenden Raum einströmt. An den erwähnten Lufttyp werden hohe Anforderungen gestellt, da er im direkten Zusammenhang mit der Raumlufqualität und dem Empfinden des Nutzers steht.

Bei Abluft (AB) handelt es sich um die vom Raum abströmende, verbrauchte Luft. Die Abluft wird dabei oftmals einer Reinigung durch Filteranlagen unterzogen, um die Schadstoffemissionen zu minimieren. Sobald die Abluft das Gebäude verlässt spricht man von Fortluft (FO). Diese ist im Gebäudebetrieb nicht mehr nutzbar und alle enthaltenen thermischen Energiepotentiale gehen verloren. [Eichmann, 1998, S. 4]



**Abbildung 1: schematische Darstellung unterschiedlicher Luftarten<sup>1</sup>**

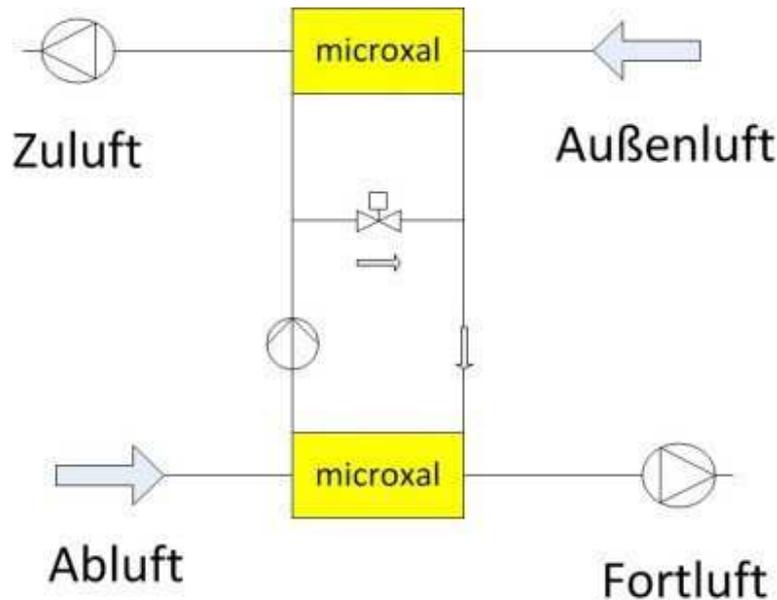
### 2.1.2 Erläuterung ausschlaggebender Anlagensystemkomponenten

Ventilatoren sind die charakterisierenden Bauteile raumluftechnischer Anlagen. Die im Rahmen dieser Arbeit verwendeten Untersuchungsbeispiele verfügen über einen Zuluft- und einen Abluftventilator. Diese Anlagenkomponenten erzeugen den Luftvolumenstrom der notwendig ist, um das Gebäude zu versorgen. Zudem werden sie dazu verwendet, die durch die Beschaffenheit des Kanalnetzes entstehenden Druckverluste zu überwinden. [Schmidt et al., 2011, S. 231]

Bei allen untersuchten Anlagen handelt es sich bzgl. der Art der Wärmerückgewinnungsanlage (WRG) um Kreislaufverbundsysteme (KVS). Diese bestehen aus zwei Wärmeübertragern, welche sich im Zuluftkanal und im Abluftkanal der raumluftechnischen Anlage befinden. Diese sind durch Rohrleitungen miteinander verbunden. Die darin befindliche Wärmeträgerflüssigkeit, ein Gemisch aus Wasser und Ethylenglycol, ermöglicht bei geöffnetem Mischventil den Wärmeenergieaustausch zwischen den unterschiedlichen Volumenströmen. Vorteilhaft am benannten Wärmerückgewinnungssystem ist, dass der Zuluftvolumenstrom und der Abluftvolumenstrom nicht vermischt werden. Durch das Ausbleiben einer direkten Verbindung des Abluft- und des Zuluftkanals können in beiden Bereichen zudem unterschiedliche Volumenströme angesetzt werden. [Schmidt et al., 2011, S 362 - 365]

In Abbildung 2 ist der schematische Aufbau eines Kreislaufverbundsystems dargestellt. Dabei sind die beiden Wärmetauscher in gelber Farbe gekennzeichnet. Wie zu erkennen ist, verfügen die Rohrleitungen zur Verbindung der Wärmetauscher neben einem Mischventil noch über eine Pumpe, welche die Zirkulation der Trägerflüssigkeit ermöglicht.

<sup>1</sup> <https://www.fibav.de/wp-content/uploads/2018/02/w%C3%A4rmer%C3%BCckgewinnung.jpg>, Stand 25.04.2018



**Abbildung 2: Darstellung eines Kreislaufverbundsystems zur Wärmerückgewinnung<sup>2</sup>**

Die WRG kann dazu genutzt werden, um die einströmende Außenluft zu kühlen, falls die Abluft kälter ist, als die Außenluft. Umgekehrt kann die einströmende Luft erwärmt werden, wenn die Abluft wärmer ist, als die Außenluft. Luftheritzer (LE) werden dazu verwendet, die in das Gebäude einströmende Außenluft zu erwärmen. Genauso, wie Luftkühler (LK) werden sie zu den Wärmeüberträgern zugeordnet. Letztere dienen der Abkühlung der einströmenden Luft, bevor diese den Räumen des Gebäudes zugeführt wird. Bei beiden Analgensystemkomponenten wird die Temperatur eines Mediums auf den Zuluftstrom übertragen. [Eichmann, 1998, S 32 - 36]

Jede der essentiellen in der Anlage verbauten Komponenten wird über Sensoren überwacht. Im Rahmen der hier vorliegenden Arbeit werden diese auch als Datenpunkte bezeichnet werden. Diese werden verwendet um in der Gebäudesystemtechnik unterschiedliche Momentanzustände zu messen. Diese Messungen werden dafür benötigt, die Anlage während des Betriebs zu überwachen und optimal einzustellen. Zudem werden diese benötigt um eventuelles Fehlverhalten von Anlagen erkennen und beheben zu können. Dennoch können die von Sensoren aufgezeichneten Messwerte in ihren Werten verfälscht sein, was verschiedene Ursachen, wie Temperatureinwirkungen, Feuchtigkeit oder Störspannungen haben kann. [Schmid et al., 2005, S117]

Die Messung von beim Anlagenbetrieb ausschlaggebender Kennwerte steht im Zusammenhang mit der Steuerung der raumluftechnischen Anlage. Bei der Steuerung von Anlagensystemen wird während Verhalten des Systems während des Betriebs nach bestimmten Zielvorgaben verändert. Bei der Regelung von Anlagen wird ebenfalls direkt in den Anlagenbetrieb eingegriffen, um eine gewünschte Auswirkung auf die klimatischen Bedingungen im Gebäude zu erhalten. Im Gegensatz zur Steuerungstechnik, wird dabei unter Verwendung von Sensoren die Auswirkung der durchgeführten Regemaßnahmen festgestellt und daraufhin eine eventuell nötige Neuregelung vorgenommen.

<sup>2</sup> <https://www.baunetzwissen.de/imgs/1/3/2/1/4/7/5/5e6b3ec8c50bcc5d.jpg>, Stand 30.05.2018

[Eichmann, 1998, S. 133] Die korrekte Regelung von Anlagensystemen bietet zahlreiche Vorteile. Durch die Beschränkung der Nutzungszeit auf die notwendigen Zeiträume kann beim Gebäudebetrieb Energie eingespart werden. [Schmid et al., 2005, S117]

### 2.1.3 Approximierte Wärmeleistung

Um ein Kriterium für den von der Lüftungsanlage erbrachten Kühl- oder Heizaufwand bereitzustellen, wird unter Verwendung einzelner vorhandener Messwerte ein Beurteilungswert geschaffen. Dabei handelt es sich um eine Approximation der Wärmeleistung der Lüftungsanlage.

Die Berechnung der Wärmeleistung ist in Formel (1) dargestellt. Dabei beschreibt  $m$  die Masse der durchströmenden Luft, mit  $c_v$  ist die spezifische Wärmekapazität der Luft aufgeführt.  $\Delta T$  ist der gemessene Temperaturunterschied in Kelvin und  $t$  das Zeitintervall, in dem die Wärmeleistung betrachtet wird.

$$Q = \dot{m} * c_v * \Delta T \quad (1)$$

Die Approximation der beschriebenen Wärmeleistung kann anhand von Formel (2) nachvollzogen werden. Es wurde angenommen, dass sich der gewählte Zuluftvolumenstrom  $V_{Zuluft}$  dazu eignet, die Heizleistung des Systems zu approximieren.

$$P_W = \dot{V}_{Zuluft} * \rho_{Luft} * c_v * (T_{Zuluft} - T_{Außenluft}) \quad (2)$$

Als  $\rho_{Luft}$  wird die Dichte der einströmenden Luft bezeichnet. Zur Vereinfachung und aufgrund der ständigen Änderung der Dichte der Luft je nach der Lufttemperatur wurde die Dichte von trockener Luft bei 20 °C verwendet. Der Wert liegt bei  $\rho_{Luft} = 1,204 \text{ kg} / \text{m}^3$ . Zudem wird die spezifische Wärmekapazität der Luft mit  $c_v = 1,005 \text{ kJ} / (\text{K} * \text{kg})$  für trockene Luft für die Berechnung der approximierten Heizleistung angenommen. Das Zeitintervall  $t$  das für die Berechnung gewählt wurde liegt bei 900 Sekunden und entspricht dem Zeitintervall der verwendeten Eingangsdaten. Die Differenz der Zulufttemperatur  $T_{Zuluft}$  und der Außenlufttemperatur  $T_{Außenluft}$  wurde gewählt, da ein negativer Wert darauf schließen lässt, dass die Anlage gekühlt hat. Wenn umgekehrt die Außenlufttemperatur geringer ist, als die Zulufttemperatur wird anhand eines positiven Wertes ersichtlich, dass eine Heizleistung erbracht wurde. Bei der Approximation muss darauf hingewiesen werden, dass in Rahmen von Formel (2) die Wärmeenergiegewinne, welche durch die WRG erzielt werden können vernachlässigt wurden.

## 2.2 Verwendete statistische Verfahren

Im folgenden Abschnitt soll der in dem von Habib et al. verfolgte Ablauf der Datenanalyse erläutert und die verwendeten Analysealgorithmen dargestellt werden. In diesem Kapitel soll weiterhin der Aufbau der von Habib et al. entworfenen Analysemethodik zur auto-

matisierten Betriebszustandserkennung vorgestellt werden. Daher sind sämtliche in diesem Kapitel enthaltenen Informationen, sofern nicht anders gekennzeichnet, aus [Habib et al, 2016] entnommen.

### 2.2.1 Ablauf der Analyse

In einem ersten Bearbeitungsschritt sollen die Daten durch Data Preprocessing für die Untersuchung vorbereitet werden. Dies wird im Rahmen der Arbeit von Habib als gegeben vorausgesetzt und nicht genauer erläutert. Im nächsten Schritt werden die Daten des Referenzsystems unter Verwendung eines k-Means-Algorithmus in AN- und AUS-Zustände unterschieden. Dabei werden alle aufeinanderfolgenden Zeitpunkte desselben Zustands zu einem Zyklus zusammengefasst. Im weiteren Verlauf der Datenanalyse werden nur AN-Zyklen betrachtet, weil diese die ausschlaggebenden Informationen über das Anlagenverhalten beinhalten. [Habib et al, 2016, S. 9] Im nächsten Schritt werden die Daten aller AN-Zyklen durch eine „Symbolic Aggregate Approximation“ (SAX) in eine symbolische Repräsentation gebracht. Anschließend werden die entstandenen Symbolreihen in eine „Bag of Words Representation“ (BoWR) umgewandelt. Die Anzahl des Auftretens jedes Symbols im jeweiligen Zyklus wird erfasst. In der folgenden Tabelle ist ein mögliches Resultat einer BoWR beispielhaft dargestellt.

Zyklus Nummer	SAX-Symbole								
	A	B	C	D	E	F	G	H	...
<b>C1</b>	0	5	4	3	2	1	0	0	...
<b>C2</b>	1	9	7	4	5	3	1	5	...
...	...	...	...	...	...	...	...	...	...
<b>C<sub>N</sub></b>	...	...	...	...	...	...	...	...	...

**Tabelle 1: Beispielhafte Darstellung von AN-Zyklen in Form von SAX-Symbolen (in Anlehnung an [Habib et al., 2015, S. 306, Tabelle 2])**

Um Zyklen unterschiedlicher Länge vergleichen zu können, wird die Häufigkeit jedes Symbols zudem durch die Anzahl aller im Betriebszyklus auftretenden Symbole geteilt. Die somit entstandene Dokumentation der relativen Häufigkeiten der Symbole eines Zyklus kann als Berechnungsgrundlage für die hierarchische Clusteranalyse dienen. Dabei werden die Ähnlichkeiten zwischen einzelnen Zyklen durch den Vergleich der ermittelten relativen Häufigkeiten jedes Symbols bestimmt.

Durch die hierarchische Clusteranalyse sollen die einzelnen Betriebszyklen unterschiedlichen Gruppierungen zugeordnet werden. Nachdem man durch die GAP-Statistic die ideale Clusterzahl für den betreffenden Datensatz bestimmt hat, werden die einzelnen Zyklen untersucht. Dafür wurden durch Habib et al. für die wichtigsten Cluster zum einen die durchschnittlichen Zykluszeiten und zum anderen ein Koeffizient zur Beurteilung der

Leistung des Systems gebildet. Dieser Coefficient of Performance(CoP) ist im Wesentlichen vom Aufbau des verwendeten Anlagensystems abhängig, wobei dessen Bildung Expertenwissen voraussetzt. [Habib et al, 2016, S.15 - 16]

Nach der Aussage von [Habib et al., 2016] kann das gezeigte Vorgehen bei der Untersuchung von Gebäudebetriebsdaten dabei helfen, Gebäudebetriebsmuster zu erkennen und zu separieren. Im Anschluss werden sie durch einen Leistungskoeffizienten und die durchschnittliche Dauer der jeweiligen Zyklen beurteilt. Dies kann vor allem dazu genutzt werden, dem Experten bei der Untersuchung des jeweiligen Anlagenverhaltens einen Überblick über auftretende Muster im Gebäudebetrieb zu ermöglichen, bevor dieser sich mit der exakten Bewertung des auftretenden Anlagenverhaltens beschäftigt. Nachdem die grundlegende Idee des von Habib et al. dargestellte Datenanalysemodells erläutert wurde, werden die theoretischen Voraussetzungen für die Untersuchung vertieft aufgezeigt.

### 2.2.2 Data Preprocessing & Sanitation

Da im Gebäudebetrieb eine große Menge an Daten anfällt, die oft heterogener Natur ist, müssen die zu analysierenden Daten erst durch eine „Data Sanitation“ vorbereitet werden. Dies ist ein essentieller Vorbereitungsschritt für die Verwendung der Daten in allen folgenden Analyseverfahren. Laut Zucker et al. bezeichnet „Data Sanitation“ den Prozess der Korrektur der zu untersuchenden Daten mit dem Ziel, die Datenqualität zu erhöhen. [Zucker et al., 2015, S. 12776] Dies umfasst die Rekonstruktion von fehlenden und unplausiblen Daten und das Durchführen von Tests, ob die von den Sensoren aufgenommenen Daten physikalisch sinnvoll sind und in einem akzeptablen, durch die jeweilige Anlagentechnik bedingten Wertebereich liegen. [Zucker et al., 2015, S. 12777]

Bei der Überwachung von Anlagensystemen werden Daten oftmals nur als Nebenprodukt der Steuerung der Anlage vor Ort aufgezeichnet. Die Daten werden dabei oft unvollständig oder in großen zeitlichen Abständen aufgezeichnet. Auch können beispielsweise Sensorausfälle auftreten und unbemerkt bleiben. Deswegen kann nicht von einer optimalen Datenqualität ausgegangen werden. [Zucker et al., 2015, S. 12779] Zudem sind die Daten bezüglich des Typs, der Einheit oder der Ermittlungszeitpunkte je nach Hersteller der Komponenten der Regelungstechnik abhängig. Aufgrund der Herstellerabhängigkeit existiert bisher noch kein standardisiertes Datenformat mit entsprechenden Meta-Informationen. [Zucker et al., 2015, S. 12779] Um die verfügbaren Daten dennoch nutzen zu können, müssen sie entsprechend angepasst werden, um die folgenden Eigenschaften qualitativ hochwertiger Daten erfüllen zu können: [Zucker et al., 2015, S. 12777]

- Zugänglichkeit
- Glaubwürdigkeit/Nachvollziehbarkeit
- Vollständigkeit
- konsistente Aufzeichnung
- Interpretierbarkeit

Folglich muss darauf geachtet werden, irrelevante und redundante Datenmengen zu erkennen und frühzeitig zu entfernen. Gerade beim Arbeiten mit Zeitreihen (engl. *timeseries*), ist es unabdingbar, die Daten der einzelnen Sensoraufzeichnungen auf eine einheitliche Zeitbasis zu bringen. Die Aufzeichnungen haben somit einen einheitlichen Startzeitpunkt und alle folgenden Aufzeichnungen werden in identischen Zeitabständen getätigt. Zudem müssen sogenannte „Outlier“, also Ausreißer aus dem plausiblen Verlauf der Zeitreihe entweder automatisch oder händisch erkannt und ggf. entfernt werden. Diese können im weiteren Verlauf der Datenanalyse eine Verzerrung der Ergebnisse bedingen. [Zucker et al., 2015]

Sowohl durch das Entfernen von Ausreißern, als auch fehlerhafte Sensoraufzeichnungen kann es zu sogenannten „Gaps“, also Lücken in den Zeitreihenaufzeichnungen kommen, die dementsprechend behandelt werden müssen. Die jeweiligen Lücken müssen nach ihrer Beschaffenheit und Größe beurteilt und dementsprechend behoben werden. Bei kleinen Lücken bietet sich eine lineare Interpolation an, bei größeren Lücken allerdings muss überlegt werden, ob man auf den jeweiligen Zeitraum, oder den kompletten Datenpunkt hinsichtlich der Analyse verzichtet. [Zucker et al., 2015, S. 12781]

Nachdem eine erfolgreiche Datenvorbereitung durchgeführt worden ist, kann ein breites Spektrum an Analyse- und Manipulationsverfahren auf den bestehenden Datensatz angewandt werden. In den folgenden Abschnitten sollen die im Rahmen dieser Arbeit genutzten Verfahren theoretisch erläutert.

### 2.2.3 Normierungsverfahren

Im Rahmen des von Habib et al. dargestellten Untersuchungsansatzes wurden die Eingangsdaten der Datenpunkte vor der Nutzung zur Datenanalyse unter Verwendung des Z-score Normierungsverfahrens aufbereitet. [Habib et al, 2016, S. 10] Die hier vorgestellten Normierungsverfahren wurden verwendet, da sie im Rahmen der Implementierung mit KNIME zum festen Bestandteil der internen Modulbibliothek gehören.

Durch die Anwendung einer z-Score Normierungsmethode werden die einzelnen Features so standardisiert, dass sie eine Standardnormalverteilung annehmen. Dafür wird folgende Formel verwendet:

$$Z = \frac{x_i - \mu}{\sigma} \quad (3)$$

Hierbei wird der Mittelwert  $\mu$  des jeweiligen Features von jedem Wert  $x_i$  des Features subtrahiert und die Differenz durch die Standardabweichung  $\sigma$  dividiert.

Bei der Verwendung der MINMAX-Normierungsmethode werden die Daten der einzelnen Features nicht auf eine Standardnormalverteilung zurückgeführt. Der Datensatz wird hingegen auf einen festen Wertebereich skaliert, welcher im Normalfall zwischen 0 und 1 liegt. (Formel 4)

$$x' = \frac{x - \min_S}{\max_S - \min_S} \quad (4)$$

Bei  $x'$  handelt es sich um den neuen Wert. Unter  $\min_S$  und  $\max_S$  versteht man das Minimum und das Maximum des Vektors  $S$  des jeweiligen Datenpunktes. Bei der dritten, im Rahmen dieser Arbeit verwendeten Normierungsmethode handelt es sich um das Decimal Scaling. Dabei werden das Maximum und das Minimum des jeweiligen Datenpunktes ermittelt und so oft durch Zehn geteilt, bis sich alle Werte im gleichen Bereich  $[-1, 1]$  befinden.

### 2.2.4 Clusteranalyseverfahren

Bei der Clusteranalyse handelt es sich um eine

„deskriptive Methode der multivariaten Statistik zur Strukturierung der beobachteten Elemente durch Bildung in sich möglichst homogener und untereinander möglichst unähnlicher Gruppen oder Cluster. Die Clusteranalyse erfolgt durch Quantifizierung der Ähnlichkeit zwischen zwei (Mengen von) Elementen mittels Ähnlichkeits- oder Distanzmaßen und anschließender Clusterbildung durch geeignete Algorithmen. Hierbei werden bes. Verfahren mit Austauschalgorithmen (partitionierende Verfahren) und hierarchische Verfahren unterschieden, die sukzessive gegebene Cluster zusammenfassen (agglomerativ) oder aufteilen (divisiv).“<sup>3</sup>

### 2.2.5 k-Means Clusteranalyse

Der im Rahmen dieser Diplomarbeit verwendete Clusteranalyse-Algorithmus wird als k-Means-Algorithmus bezeichnet. Er wurde genutzt um Betriebszuständen der Anlagentechnik von Ruhezuständen zu unterscheiden. Ein Betriebszyklus ist dabei eine Abfolge von aufgezeichneten Werten, die den gleichen Status aufweisen. Die prinzipielle Eignung des Algorithmus für die Unterscheidung wurde bereits bestätigt. [Lloyd, S., 1982] k-Means ist eine Methode zum Finden von Clustern in Daten basierend auf dem Kriterium des quadratischen Fehlers [Jain, A., 2010]. Der Algorithmus findet k-Partitionen, sodass der quadratische Fehler zwischen dem Mittelwert  $\mu_k$  des Clusters und den einzelnen Elementen des Clusters minimiert wird. Der quadratische Fehler  $J$  eines Clusters  $C_k$  ist wie folgt definiert [Zucker et al., 2015, S. 12780]:

$$J(C_k) = \sum_{x_i \in C_k}^n \|x_i - \mu_k\|^2 \quad (5)$$

Die Berechnung wird für alle Elemente  $x_i$  des jeweiligen Clusters durchgeführt. Dabei ist  $n$  die maximale Anzahl der Elemente innerhalb des jeweiligen Clusters  $C_k$ , welcher den Mittelwert  $\mu_k$  besitzt. Das Ziel des Algorithmus ist es, die Summe der quadratischen Fehler über die Menge aller Cluster  $C = \{C_k, k = 1, \dots, K\}$  entsprechend der folgenden Formel zu minimieren [Zucker et al., 2015]:

---

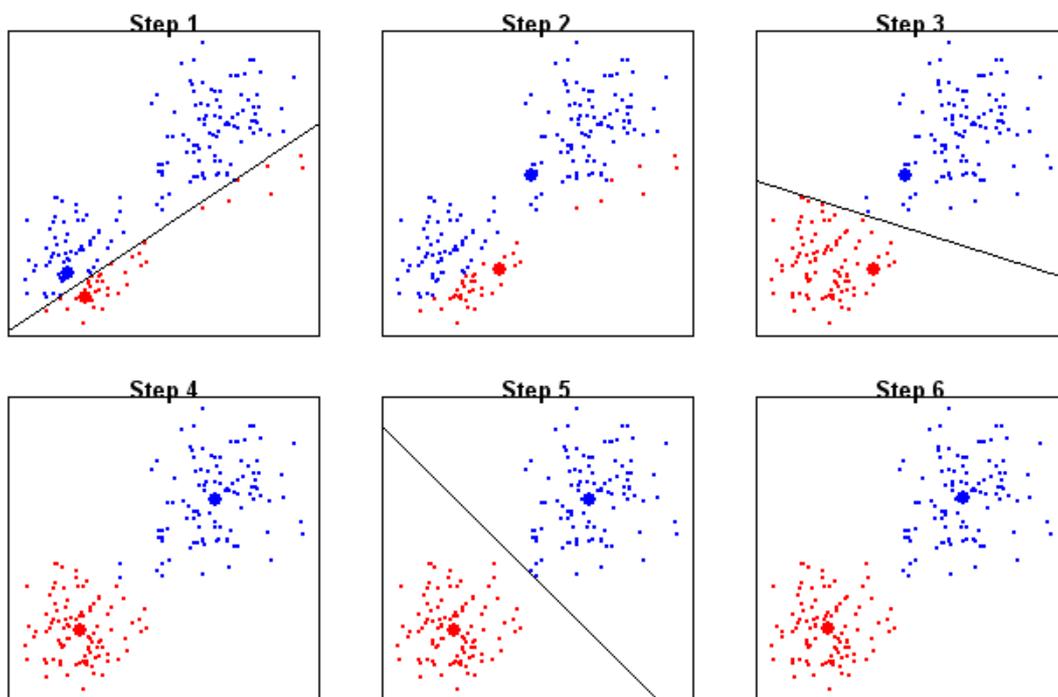
<sup>3</sup> <https://wirtschaftslexikon.gabler.de/definition/clusteranalyse-28221>, Stand 27.05.2018

$$J(C) = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (6)$$

Der Algorithmus geht dabei wie folgt vor:

1. Zufälliges Generieren der Start-Partitionen der K-Cluster
2. Zuweisen der Elemente  $x_i$  zum jeweilig nächstgelegenen Cluster-Zentrum
3. Berechnung der neuen Position der Cluster-Zentren als Mittelpunkte der ausgewählten Daten.
4. Schritt zwei und drei werden solange wiederholt, bis sämtliche Punkte einem Cluster zugeordnet sind

Anhand von Abbildung 3 kann der oben beschriebene Iterationsprozess des k-Means Algorithmus nachvollzogen werden.



**Abbildung 3: Schrittweiser Iterationsprozess des k-Means Algorithmus<sup>4</sup>**

Der Algorithmus erfordert per se keinen Konfigurationsaufwand. Nach der Clusteranalyse durch den Algorithmus muss lediglich manuell zwischen AN-Zuständen und AUS-Zuständen unterschieden werden. [Zucker et al., 2015, S. 12780] Dennoch sollte die Gesamtanzahl der zu bildenden Cluster  $K$  dabei im Voraus in Bezug auf die entsprechende Anwendung festgelegt werden. Bei der Unterscheidung von AN- und AUS-Zuständen wird  $K = 2$  angesetzt. Als Distanzmaß für die Unterscheidung der verschiedenen Elemente wird die euklidische Distanz verwendet. [Habib et al., 2016]

<sup>4</sup> [https://cdn-images-1.medium.com/max/800/0\\*rrzG3LyOnAvOepbJ.png](https://cdn-images-1.medium.com/max/800/0*rrzG3LyOnAvOepbJ.png), Stand 26.05.2018

## 2.2.6 Hierarchische Clusteranalyse

Im nächsten Schritt, dem hierarchischen Clustering, geht es darum, die Ähnlichkeit zwischen den einzelnen Zyklen zu untersuchen. Dafür werden die euklidischen Distanzen aller Zyklen zueinander bestimmt, durch die hierarchische Clusteranalyse gruppiert und in einem Dendrogramm dargestellt [Vesanto, J. et al, 2000]. Das Dendrogramm zeigt die Hierarchie der gebildeten Cluster. Auf jeder Hierarchiestufe des Dendrogramms werden zuvor getrennte Gruppierungen zu einer neuen, übergeordneten Gruppe zusammengefasst. [Habib et al, 2016, S. 11]. Laut Habib et al hat die Average Methode bei der hierarchischen Clusteranalyse die besten Ergebnisse erzeugt. [Habib et al, 2016, S. 13] Deswegen wird sie auch im Rahmen der hier vorgestellten Analysen raumluftechnischer Anlagen genutzt werden.

Die Ergebnisse der hierarchischen Clusteranalyse werden durch ein Dendrogramm dargestellt. Dabei handelt es sich um ein Baumdiagramm, anhand dessen man die Clusterbildung bei hierarchischen Clusteranalysen nachvollziehen kann. Vorteilig an der Illustrationsweise ist, dass die Distanzen einzelner Gruppierungen die zu einer neuen Gruppierung zusammengefasst werden, anhand der senkrechten Linien nachvollziehbar sind. In Abbildung 4 ist ein Dendrogramm beispielhaft dargestellt.<sup>5</sup>

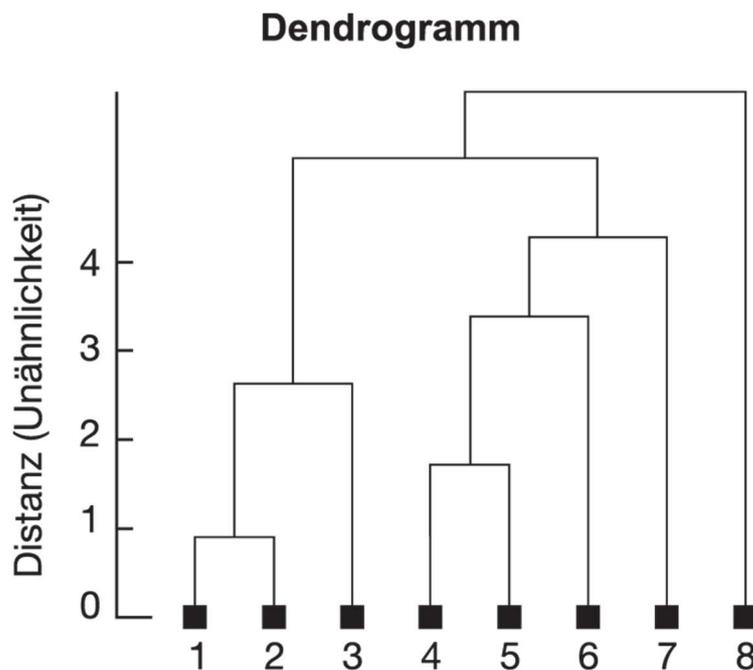


Abbildung 4: Beispielhafte Darstellung eines Dendrogramms<sup>6</sup>

<sup>5</sup> <https://wirtschaftslexikon.gabler.de/definition/dendrogramm-31008#references>, Stand 24.05.2018

<sup>6</sup> <https://wirtschaftslexikon.gabler.de/sites/default/files/styles/keyword/public/images/wilex/73528.png?itok=KI07mdMF>, Stand 26.05.2018

### 2.2.7 Evaluation der gefundenen Cluster mit der GAP-Statistic

Bei der Verwendung der hierarchischen Clusteranalyse ist es unbedingt erforderlich, das entstandene Dendrogramm hinsichtlich der idealen Clusteranzahl zu evaluieren. Eine dafür verwendbare Evaluationsmethode ist der Einsatz der GAP-Statistics-Methode. Dabei wird die Veränderung der Streuung der Elemente eines Clusters mit einer geeigneten Referenznullverteilung verglichen. Resultierend wird der GAP-Wert, also die Differenz eines aus der Referenznullverteilung errechneten Erwartungswertes und einem Streuungsmaß für den Eingangsdatensatz gebildet. Die ideale Anzahl an Clustern kann durch diese Methode für die Anzahl an Clustern abgelesen werden, für die der Wert ein Maximum annimmt. [Tibshirani et al., 2001] Bei der im Rahmen dieser wissenschaftlichen Arbeit durchgeführten Datenanalyse konnte die GAP-Statistic-Methode zwar umgesetzt werden, hat aber zu keinem nennenswerten Resultat geführt.

### 2.2.8 Symbolic Aggregate Approximation

Im Folgenden soll die Transformation von Daten unter Verwendung der Symbolic Aggregate Approximation (SAX) beschrieben werden. Bei diesem Verfahren werden die Daten nach einer geeigneten Normierung zunächst in Zeitintervalle mit identischer Länge zusammengefasst und der Mittelwert des jeweiligen Intervalls gebildet, was man als „Piecewise Aggregate Approximation“ (PAA) bezeichnet. [Lin et al., 2007]

Der prinzipielle Vorgang der PAA wird in der folgenden Abbildung 5 ersichtlich. Dieser erste Teil der SAX-Transformation soll im Rahmen der Arbeit allerdings nur aufgeführt werden.

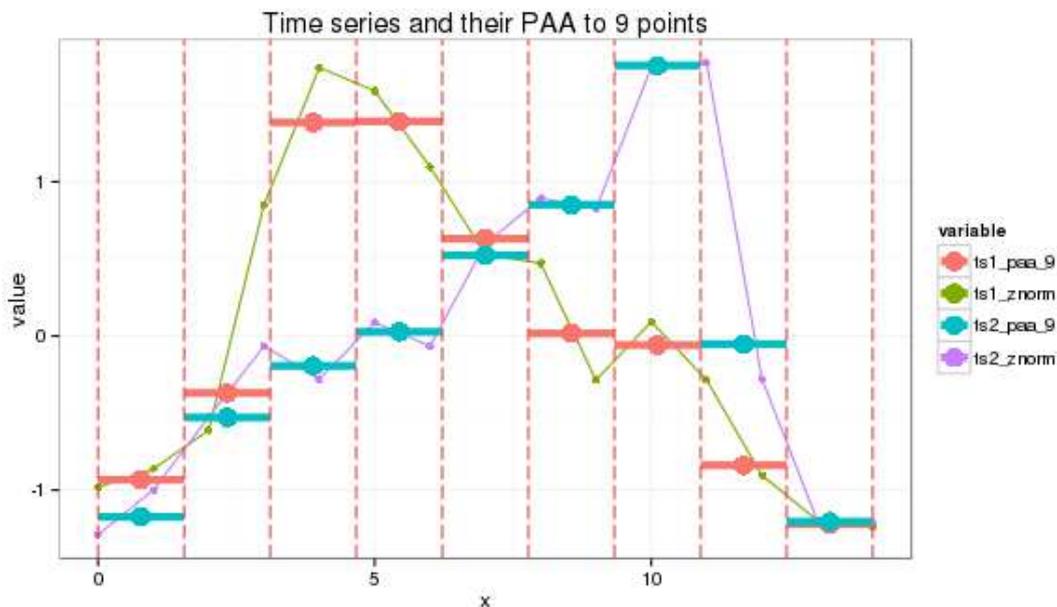
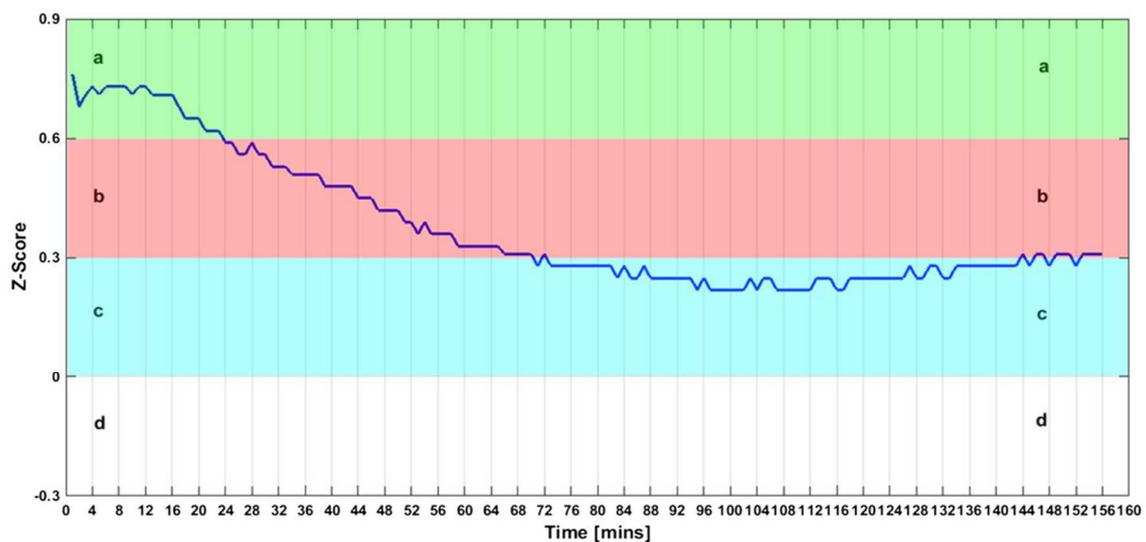


Abbildung 5: Piecewise Aggregate Approximation für zwei Zeitreihen ts1 & ts2<sup>7</sup>

<sup>7</sup> [https://jmotif.github.io/sax-vsm\\_site/morea/assets/sax\\_paa1.png](https://jmotif.github.io/sax-vsm_site/morea/assets/sax_paa1.png), Stand 24.04.2018

In einem zweiten Schritt werden die Daten in nicht überlappende Wertebereiche gleicher Größe unterteilt, denen ein eindeutiges Symbol zugewiesen wird. Diese Wertebereiche werden im Folgenden als „Chunks“ bezeichnet. Im Verlauf jeder Zeitreihe jedes AN-Zyklus werden also anstatt zeit- und wertkontinuierlicher numerischer Daten zeit- und wertdiskrete Symbole erzeugt. [Habib et al., 2015]

In Abbildung 6 ist das prinzipielle Resultat einer SAX Transformation beispielhaft illustriert. Im Bildausschnitt sind zudem vier unterschiedliche Wertebereiche in unterschiedlichen Farben dargestellt. Im Bildausschnitt ist die Anzahl dieser Wertebereiche, die auch Chunks genannt werden mit  $M = 4$  gewählt. In jedem der dargestellten Wertebereiche ist das jeweilig zuzuordnende Symbol dargestellt. Die Länge der Zeitintervalle ist mit  $P = 4$  min angegeben. Außerdem sind die zugehörigen Symbole innerhalb der Bereiche eingetragen. [Habib et al, 2016]



aaaaaabbbbbbbbbbbbccccccccccccccccccccbb

**Abbildung 6: Beispielhafte Illustration einer SAX Transformation anhand einer normierten Zeitreihe [Habib et al, 2016, S. 10, Abbildung 3]**

Die bei [Habib et al., 2016, S 10] verwendeten Formeln wurden im Rahmen dieser Arbeit auf identische Art und Weise verwendet und werden aus diesem Grund an dieser Stelle nicht weiter aufgeführt.

## 2.2.9 Bag-of-Words Representation

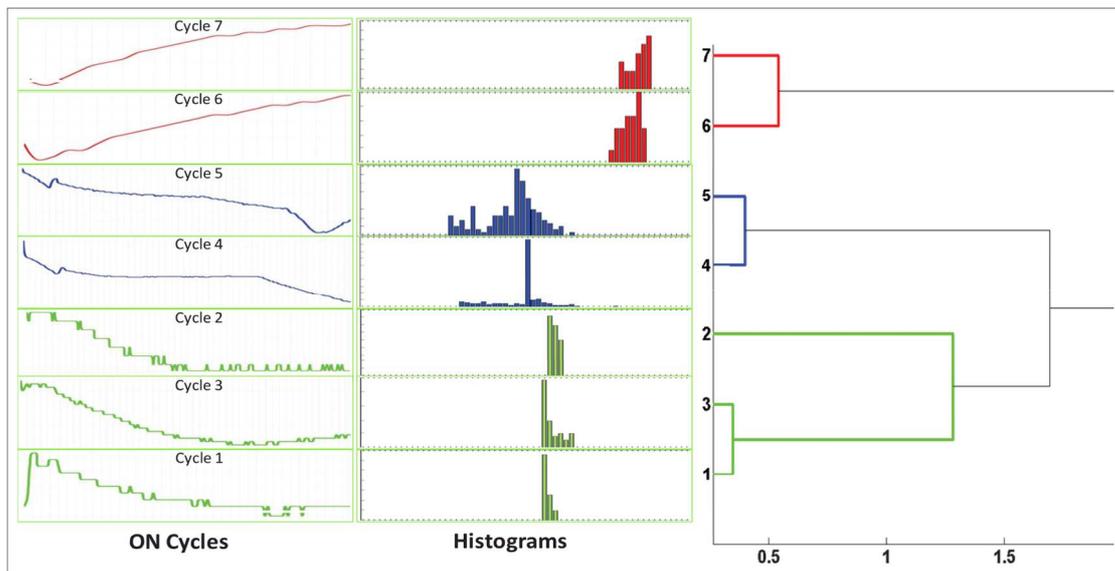
Durch das im vorherigen Schritt durchgeführte „Chunking“ als Teil der SAX-Transformation, also die Unterteilung der Timeseries in wertdiskrete Partitionen verfügt man nun über eine Menge von Symbolen, die in jedem ON-Zyklus für jeden Sensor eine unterschiedliche Häufigkeit besitzen. Um das Verhalten des Zyklus zu veranschaulichen, werden die Daten jedes Sensors in eine Bag of Words Repräsentation (BoWR)

umgewandelt. Sie wird verwendet, um Betriebszyklen unterschiedlicher Dauer vergleichbar zu machen. [Habib et al, 2016, S. 10 - 11].

Für sämtliche Sensoren die später für das hierarchische Clustering verwendet werden sollen, werden die im Zeitintervall jedes Betriebszyklus gesammelten Symbole zu einer BoWR zusammengefasst.

Die Anzahl unterschiedlicher Symbole ergibt sich dabei anhand der Anzahl der bei der SAX-Transformation gewählten Chunks  $M$ . Für jedes dieser individuellen Symbole wird nun die Anzahl des Vorkommens im zu betrachtenden Sensor des zu betrachtenden Zyklus gezählt. Um auch Betriebszyklen mit unterschiedlicher Dauer vergleichen zu können, wird die Häufigkeit des Vorkommens der einzelnen Symbole für jeden Sensor eines Zyklus durch die Gesamtanzahl der Elemente im Zyklus geteilt, also die relative Häufigkeit gebildet. [Habib et al., 2016, S. 11]. Die bei der Bestimmung der BoWR verwendeten Formeln wurden aus [Habib et al., 2016, S. 11] übernommen und für die Implementierung verwendet.

In Abbildung 7 sind mehrere Betriebszyklen in der linken Spalte beispielhaft dargestellt. Die sich aus ihnen ergebenden Bag of Words Repräsentationen sind in der Mittleren Spalte als Histogramme illustriert. Anhand der Histogramme ist die Verteilung der Symbole erkennbar. Auf deren Grundlage wird die Ähnlichkeit der Betriebszyklen durch die hierarchische Clusteranalyse bestimmt. Die Darstellung der Ähnlichkeiten kann anhand des in der rechten Spalte gezeigten Dendrogramms nachvollzogen werden.



**Abbildung 7: Darstellung des Übergangs verschiedener Betriebszyklen in die Bag of Words Repräsentation und schließlich zum Dendrogramm des hierarchischen Clustering [Habib et al., 2015, S 308, Abbildung 7]**

Im Vorangegangenen wurden die der Datenanalyse zugrundeliegenden Bestandteile theoretisch erläutert. Hierbei wurde ein Überblick über die einzelnen Aspekte der Gebäudeleittechnik gegeben, des Weiteren wurden die verwendeten statistischen Verfahren beschrieben. Dies dient als Voraussetzung für die im folgenden Kapitel erläuterte Umsetzung der Datenanalyse.

### 3 Umsetzung

Gegenstand dieses Kapitels ist die Übertragung des von Habib et al. verwendeten Verfahrens zur Klassifizierung von Betriebszuständen auf verschiedene raumlufotechnische Anlagen. Dabei wird zunächst auf die im Rahmen der Arbeit entstandene Implementierung der Datenanalyse mit KNIME eingegangen. Anschließend werden die zur Verifizierung des Forschungsansatzes analysierten raumlufotechnischen Anlagen vorgestellt. Daraufhin wird die Übertragung des Datenanalysevorgangs beschrieben. Zudem wird das zu erwartende Anlagenverhalten näher erläutert.

#### 3.1 Implementierung der verwendeten Algorithmen mit KNIME

Beim „Konstanz Information Miner“ (KNIME) handelt es sich um ein graphisches Tool für die interaktive Analyse von Daten. Die Software ist als Open Source Platform erhältlich.<sup>8</sup> Im Rahmen der Nutzung wird eine Auswahl aus über 2000 verschiedenen Modulen getroffen. Diese – nachfolgend auch als Nodes bezeichneten – Knoten werden mit einander zu sog. Workflows verbunden und den Anforderungen des Nutzers entsprechend konfiguriert.

Neben der Verwendung der programminternen Modulstruktur bietet sich auch die Möglichkeit, Programmcode aus anderen Programmiersprachen in den KNIME-Workflow aufzunehmen, also verschiedene andere Programmiersprachen in KNIME zu verwenden. Außerdem ist eine Anbindung von Big Data-Analyseprogrammen wie Apache Hadoop oder Spark möglich. Zudem können Verfahren des maschinellen Lernens und des Deep Learning abgebildet werden.<sup>9</sup> In Abbildung 8 ist ein Workflow der k-Means Clusteranalyse zu Veranschaulichungszwecken dargestellt.

---

<sup>8</sup> <https://www.knime.com/knime-open-source-story>, Stand 27.02.2018

<sup>9</sup> <https://www.knime.com/software>, Stand 27.02.2018

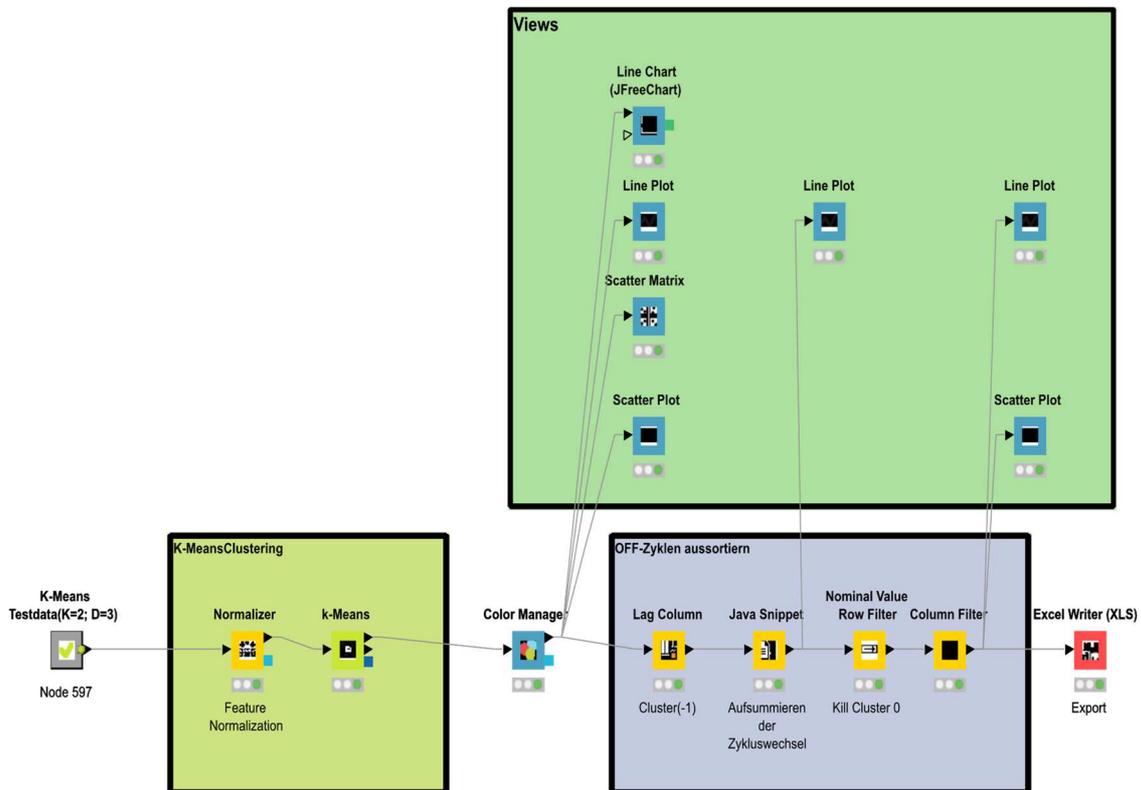


Abbildung 8: Beispielhafte Darstellung eines Workflows für das k-Means-Clusteranalyse

### 3.2 Zu untersuchende Anlagensysteme

Exemplarisch wurden mehrere Anlagensysteme aus einer Industriekooperation des Fraunhofer Instituts zur Untersuchung der jeweiligen Betriebsdaten ausgewählt. Bei dem zu untersuchenden Gebäudekomplex handelt es sich um eine größere Liegenschaft, in welcher insgesamt 15 raumlufttechnische Anlagen in Betrieb sind. Da diese für unterschiedliche Bereiche des Gebäudekomplexes vorgesehen wurden, welche den Anspruch auf individualisierte Klimatisierung erheben, agieren sie weitestgehend autonom und können somit separat untersucht werden. Stellvertretend für die Gesamtheit der vorhandenen Gebäudeenergietechnik werden nachfolgend die Daten der Anlagen Lüftungsanlage 06 und Lüftungsanlage 03 untersucht. Beide raumlufttechnischen Anlagen sind sowohl in ihrem Aufbau, als auch bezüglich der Klimatisierungsanforderungen weitestgehend identisch. Dennoch weisen beide Systeme ein unterschiedliches Betriebsverhalten auf. Das Hintergrundwissen zur Beurteilung der grundlegenden Funktion der einzelnen Anlagenkomponenten und deren Zusammenwirken konnte im Rahmen des Anstellungsverhältnisses am Fraunhofer-Institut für Integrierte-Schaltungen – Institutsteil Entwurf adaptiver Systeme (Dresden) vor Beginn der Diplomarbeit erworben werden und soll im Folgenden erläutert werden.

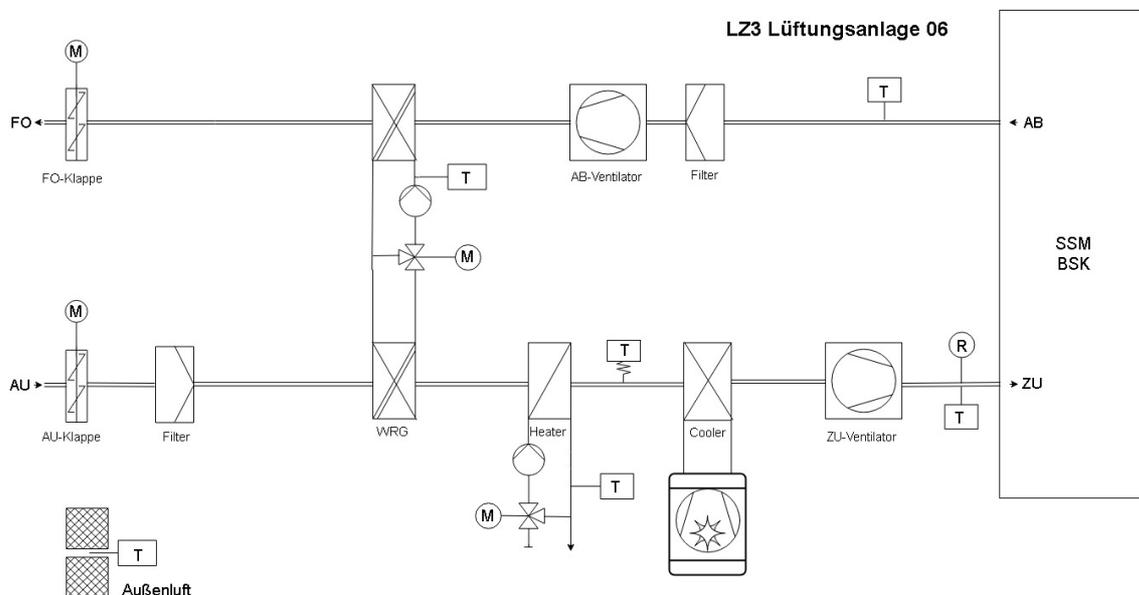
### 3.2.1 Täglicher Betriebszeitraum der raumluftechnischen Anlagen

Bei der zu untersuchenden Liegenschaft handelt es sich um ein mehrstöckiges Einkaufszentrum. Es wird folglich an 6 Tagen in der Woche von 9:30 Uhr bis 20:00 Uhr von Besuchern genutzt. Um das Gebäude schon im Vorfeld entsprechend klimatisiert zu haben sowie aufgrund diverser Arbeitsleistungen des Personals der einzelnen Geschäfte sowie des Facility Managements, beginnt man mit dem Anlagenbetrieb schon deutlich früher.

Dadurch werden beide zu untersuchenden Anlagen an Öffnungstagen von 7 bis 19 Uhr betrieben. Der tägliche Startzeitpunkt der Anlage variiert geringfügig, aufgrund unterschiedlicher Anforderungen, welche sich aus den klimatischen Bedingungen ergeben. Aufgrund der Funktion des Gebäudes als Speichermasse besteht in der letzten Öffnungsstunde kein Klimatisierungsbedarf, weswegen der Betrieb in der Regel schon gegen 19 Uhr endet. Diese Annahme spiegelt sich auch in einer Untersuchung der Volumenströme der täglichen Betriebszyklen wider und wird somit als korrekt angesehen.

### 3.2.2 Lüftungsanlage 06

Bei der Untersuchung unterschiedlicher Anlagen wird zunächst mit der Lüftungsanlage 06 (RLT 06) begonnen. Diese setzt sich aus einem Zuluftkanal und einem Abluftkanal zusammen, die durch die Wärmerückgewinnungsanlage (WRG) miteinander verbunden sind. Der genaue Aufbau der Anlage kann in Abbildung 9 nachvollzogen werden.



**Abbildung 9: Anlagenschema "Lüftungsanlage 06"**

Die genannte WRG kann zum einen genutzt werden, um bei niedrigen Temperaturen einen Teil der erzeugten Wärmeenergie aus dem Abluftkanal in den Zuluftkanal rückzuführen. Umgekehrt kann bei hohen Außentemperaturen ein Teil der in der Außenluft enthaltenen Energie über die WRG abgebaut werden, was den Energieverbrauch durch

andere Kühlmaßnahmen verringert. Zur Überwachung der Funktionsweise der Anlagenkomponente werden zum einen die Stellung des Regelungsventils und die Temperatur der in der WRG befindlichen Klimatisierungsflüssigkeit, sowie der Betriebsstatus der WRG-Pumpe in regelmäßigen zeitlichen Abständen über Sensortechnik dokumentiert. Zudem verfügt die Wärmerückgewinnung über einen kumulativen Stundenzähler. Neben der WRG nehmen ein Luffterhitzer und ein Luftkühler Einfluss auf die klimatischen Bedingungen der in das Gebäude einzubringenden Zuluft. Die Kontrolle der korrekten Regelung kann hierbei über den Temperatursensor im Fernwärmekreislauf und über die Stellung des zugehörigen Ventils erfolgen. Auch der Betriebszustand der Pumpe im Vorlauf des Luffterhitzers wird durch einen Datenpunkt dokumentiert. Zudem verfügt dieser über einen Betriebsstundenzähler.

Der lokale Luftkühler wird über eine zentrale Kältemaschine mit der erforderlichen Kühlleistung versorgt. Dabei stehen vier unterschiedliche Betriebsmodi zur Verfügung, welche sich in Ihrer Intensität unterscheiden. Im Rahmen der Betriebsdatenanalyse stehen hierbei Sensorwerte von Betriebsstundenzählern sowie Betriebssensoren des Luftkühlers zur Verfügung. Die binären Sensorwerte geben Auskunft, ob das Kühlaggregat zum jeweiligen Zeitpunkt in Bezug auf die Lüftungsanlage in Benutzung war. Zwischen dem Luffterhitzer und dem Luftkühler befindet sich außerdem ein Temperaturwächter. Dieser dient dazu, beim Anlagenbetrieb ein Warnsignal auszusenden, insofern eine bestimmte Mindesttemperatur unterschritten wird. Dieser Datenpunkt wurde im Rahmen der Datenanalyse vernachlässigt.

Um den für den Anlagenbetrieb erforderlichen Volumenstrom zu gewährleisten, befinden sich ein Abluft- und ein Zuluftventilator in der Anlage. Diese übermitteln über verbaute Sensortechnik Informationen über den Volumenstrom in  $\text{m}^3 / \text{h}$  und die zu beobachtende Frequenz. Außerdem werden Stundenzählerwerte mit Informationen zur jeweiligen Anlagentechnik kumulativ ermittelt. Zudem wird in beiden ein binäres Signal übergeben, welches Auskunft darüber gibt, ob der Ventilator zum jeweiligen Zeitpunkt betrieben wurde. Der Lüftungskreislauf verfügt des Weiteren über eine Fortluft- und eine Außenluftklappe, welche bei Bedarf dazu genutzt werden können, um die Anlagen gegen äußere Einflüsse abzuschotten. Dies hilft zudem, die Energieverluste des Gebäudes im Ruhezustand der Anlage zu minimieren. Bei korrekter Einstellung können Luftklappen Auskunft darüber geben, ob die Anlage im Betrieb ist. Um den Eintrag von Schadstoffen oder Partikeln aus der Außenluft in das Gebäude zu minimieren, bzw. die Schadstoffemission resultierend aus dem Gebäudebetrieb möglichst gering zu halten, werden an beiden Abschnitten der Lüftungsanlage Luftfilter verwendet. Diese sind im Rahmen der Luftqualität zwar unverzichtbar, werden jedoch bei der hier durchgeführten energetischen Bewertung der Anlage vernachlässigt.

In Tabelle 2 sind sämtliche bei der Analyse des Anlagensystems verwendeten Datenpunkte aufgeführt. Zudem wurde eine Erläuterung der einzelnen Datenpunkte bereitgestellt. Dabei wurde eine Vorauswahl der Daten getroffen, um defekte Sensoren nicht aufzulisten. Zudem wurden einige der Datenpunkte bereits in eine für die Analyse verwendbare Form gebracht.

Da die Betriebsstundenzähler der verwendeten Lüftungsanlagen als kumulative Aufzeichnungen vorliegen, wurde im Vorfeld die zeitliche Ableitung dieser gebildet. Sie stellt

einen Index für die Betriebsintensität der jeweiligen Anlage dar und wird im Folgenden als Betriebsstufe bezeichnet werden. Außerdem wurden die in der Anlage verbauten Wärmemengenzähler aufgrund der Beschaffenheit der Daten nach der Zeit abgeleitet, weil es sich um einen kumulativen Sensor handelt. Da die Aktualisierung des Sensors nur zu bestimmten Tageszeitpunkten stattgefunden hat, wurden Mittelwerte der Ableitung für jeden Tag gebildet.

<b>Name des Datenpunkts</b>	<b>Erläuterung</b>
<b>RLT06 Ablufttemperatur</b>	Temperatursensor im Abluftkanal
<b>RLT06 Abluftvolumenstrom</b>	Messung des Luftvolumenstroms im Abluftkanal; gemessen am Abluftventilator
<b>RLT06 Frequenzumrichter Abluftventilator</b>	Binäres Betriebssignal des Frequenzumrichters des Abluftventilators
<b>RLT06 Außentemperatur</b>	Temperatursensor für die Außentemperatur
<b>RLT06 LE-Rücklauftemperatur</b>	Temperatursensor für die Rücklauftemperatur des Lufferhitzers
<b>RLT06 LE-Ventilstellung</b>	Angabe der prozentualen Öffnung des Ventils im Vorlauf des Lufferhitzers
<b>RLT06 Betriebssensor WRG-Pumpe</b>	Binäres Betriebssignal der Pumpe der Wärmerückgewinnungsanlage; gibt an, ob die Pumpe in Betrieb ist
<b>RLT06 WRG-Temperatur</b>	Temperatursensor der Wärmerückgewinnungsanlage
<b>RLT06 WRG-Ventilstellung</b>	Angabe der prozentualen Öffnung des Ventils der Wärmerückgewinnungsanlage
<b>RLT06 Zulufttemperatur</b>	Temperatursensor für die Zulufttemperatur
<b>RLT06 Zuluftvolumenstrom</b>	Messung des Luftvolumenstroms im Zuluftkanal; gemessen am Zuluftventilator
<b>RLT06 Frequenzumrichter Zuluftventilator</b>	Binäres Betriebssignal des Frequenzumrichters des Zuluftventilators
<b>RLT06 Betriebssensor Kältemaschine</b>	Binäres Betriebssignal des Luftkühlers
<b>RLT06 Betriebssensor Kältemaschine Stufe 1</b>	Binäres Betriebssignal des Luftkühlers auf Stufe 1
<b>RLT06 Betriebssensor Kältemaschine Stufe 2</b>	Binäres Betriebssignal des Luftkühlers auf Stufe 2
<b>RLT06 Betriebssensor Kältemaschine Stufe 3</b>	Binäres Betriebssignal des Luftkühlers auf Stufe 3
<b>RLT06 Betriebssensor Kältemaschine Stufe 4</b>	Binäres Betriebssignal des Luftkühlers auf Stufe 4
<b>RLT06 Betriebsstufe Abluftventilator</b>	Zeitableitung des Betriebsstundenzählers des Abluftventilators

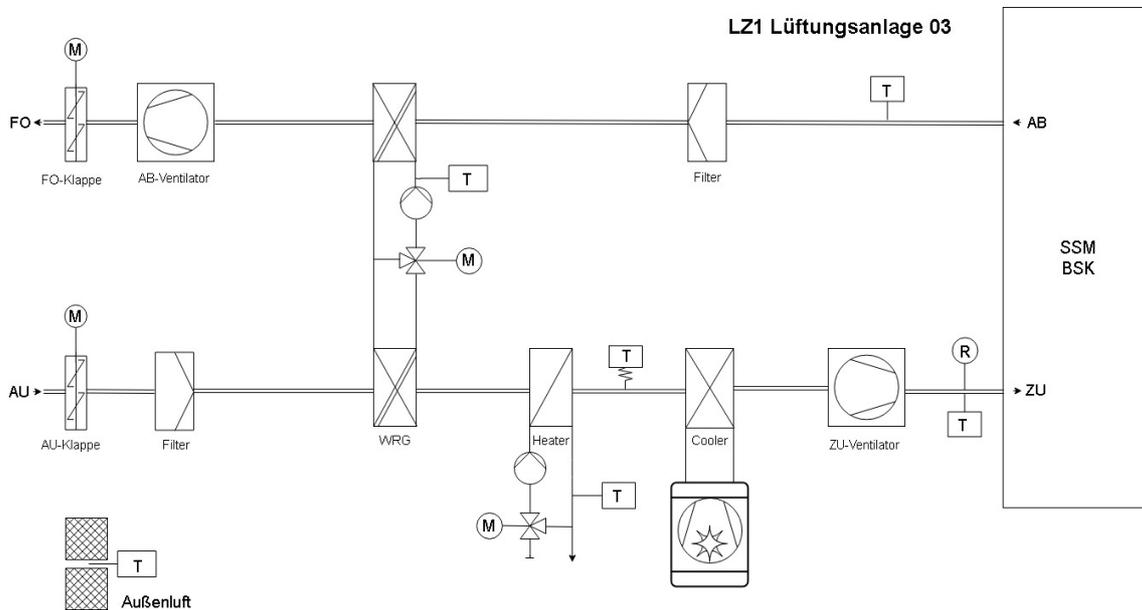
<b>RLT06 Betriebsstufe LE-Pumpe</b>	Zeitableitung des Betriebsstundenzählers der LE-Pumpe
<b>RLT06 Betriebsstufe WRG-Pumpe</b>	Zeitableitung des Betriebsstundenzählers der WRG-Pumpe
<b>RLT06 Betriebsstufe Zuluftventilator</b>	Zeitableitung des Betriebsstundenzählers des Zuluftventilators
<b>RLT06 Betriebsstufe Kältemaschine Stufe 1</b>	Zeitableitung des Betriebsstundenzählers der Kältemaschine für Stufe 1
<b>RLT06 Betriebsstufe Kältemaschine Stufe 2</b>	Zeitableitung des Betriebsstundenzählers der Kältemaschine für Stufe 2
<b>RLT06 Betriebsstufe Kältemaschine Stufe 3</b>	Zeitableitung des Betriebsstundenzählers der Kältemaschine für Stufe 3
<b>RLT06 Betriebsstufe Luftkühler gesamt</b>	Zeitableitung des Betriebsstundenzählers des Luftkühlers
<b>RLT06 Tagesmittelwert Wärmemengenzähler statische Heizung 1</b>	Zeitableitung des Wärmemengenzählers der statischen Heizanlage 1; zudem als Tagesmittelwert bestimmt;
<b>RLT06 Tagesmittelwert Wärmemengenzähler statische Heizung 0</b>	Zeitableitung des Wärmemengenzählers der statischen Heizanlage 0; zudem als Tagesmittelwert bestimmt;
<b>RLT06 Tagesmittelwert Wärmemengenzähler Lüftungsanlage</b>	Zeitableitung des Wärmemengenzählers der Lüftungsanlage zudem als Tagesmittelwert bestimmt;

**Tabelle 2: Aufzählung und Erläuterung verwendeter Datenpunkte der Anlage 06**

Der Betrachtungszeitraum der Datenanalyse bei RLT06 beschränkt sich auf sämtliche Werte zwischen dem 03.03.2017 10.45 Uhr und dem 05.11.2017 23:45 Uhr. Im ersten Quartal des Jahres 2017 sind die Daten aufgrund von Sensorfehlern unbrauchbar gewesen, weswegen eine Untersuchung des Anlagenverhaltens zum Anfang des Jahres nicht möglich war. Zudem lässt sich feststellen, dass das Gebäude in der Regel von Montag bis Samstag genutzt wurde. Zusätzlich dazu wurde die Anlage an zwei Sonntagen, dem 19.03 und dem 28.05 betrieben. Es wird angenommen, dass es sich um verkaufsoffene Sonntage handelt. Die Anlage wurde am 14.04.2017, am 17.04.2017, am 01.05.2017 und am 31.10.2017 nicht betrieben, was auf verschiedene Feiertage zurückzuführen ist.

### 3.2.3 Lüftungsanlage 03

Das zweite zu untersuchende System, Lüftungsanlage 03 (RLT03) gleicht in seinem Aufbau im Wesentlichen dem von Anlage 06. Die Anlage setzt sich ebenfalls aus einem Zuluft- und einem Abluftkanal zusammen. Diese sind durch eine WRG verbunden.



**Abbildung 10: Anlagenschema Lüftungsanlage 03**

Der Zuluftkanal besteht aus einer Außenluftklappe, einem Luftfilter, einem WRG Modul, einem Luftherhitzer, einem Luftkühler und einem Zuluftventilator. Der Abluftkanal ist aus einem Luftfilter, einem Abluftventilator, dem anderen WRG-Modul und der Fortluftklappe aufgebaut. Die genaue Anordnung der einzelnen Komponenten kann in Abbildung 10 entnommen werden. Für die Datenanalyse sind die Ventilatoren, die WRG, der LE und der Kühler die wichtigsten Komponenten der Anlage. An diesen befinden sich auch die für die Analyse ausschlaggebenden Datenpunkte. Einziger Unterschied im Aufbau von Lüftungsanlage 03 im Vergleich zu Lüftungsanlage 06 liegt in der Anordnung des Abluftventilators, der direkt vor der Fortluftklappe angeordnet ist. Dadurch kann die beim Betrieb des Ventilators entstehende Rotationswärme nicht durch die WRG rückgeführt werden. Umgekehrt wird dadurch eine effizientere Nutzung der Wärmerückgewinnung im Kühlbetrieb möglich.

Für die Datenanalyse von RLT03 wurde ebenfalls eine Auswahl ausschlaggebender Datenpunkte getroffen. Defekte Sensoren wurden nicht in die Datenanalyse mit einbezogen. Die Stundenzähler und Wärmemengenzähler der Anlage wurden entsprechend der in Kapitel 3.2.2 gegebenen Erläuterung vorbereitet. Die Liste der für die Klassifizierung von Betriebszuständen verwendeten Datenpunkte, sowie die Erläuterung dieser können in Tabelle 3 nachvollzogen werden.

<b>Name des Datenpunkts</b>	<b>Erläuterung</b>
<b>RLT03 Ablufttemperatur</b>	Temperatursensor im Abluftkanal
<b>RLT03 Abluftvolumenstrom</b>	Messung des Luftvolumenstroms im Abluftkanal; gemessen am Abluftventilator
<b>RLT03 Frequenzumrichter Abluftventilator</b>	Binäres Betriebssignal des Frequenzumrichters des Abluftventilators
<b>RLT03 Außentemperatur</b>	Temperatursensor für die Außentemperatur
<b>RLT03 LE-Rücklauftemperatur</b>	Temperatursensor für die Rücklauftemperatur des Luftherhitzers
<b>RLT03 LE-Ventilstellung</b>	Angabe der prozentualen Öffnung des Ventils im Vorlauf des Luftherhitzers
<b>RLT03 WRG-Temperatur</b>	Temperatursensor der Wärmerückgewinnungsanlage
<b>RLT03 WRG-Ventilstellung</b>	Angabe der prozentualen Öffnung des Ventils der Wärmerückgewinnungsanlage
<b>RLT03 Zulufttemperatur</b>	Temperatursensor für die Zulufttemperatur
<b>RLT03 Zuluftvolumenstrom</b>	Messung des Luftvolumenstroms im Zuluftkanal; gemessen am Zuluftventilator
<b>RLT03 Betriebsstufe Abluftventilator</b>	Zeitableitung des Betriebsstundenzählers des Abluftventilators
<b>RLT03 Betriebsstufe LE-Pumpe</b>	Zeitableitung des Betriebsstundenzählers der LE-Pumpe
<b>RLT03 Betriebsstufe WRG-Pumpe</b>	Zeitableitung des Betriebsstundenzählers der WRG-Pumpe
<b>RLT03 Betriebsstufe Zuluftventilator</b>	Zeitableitung des Betriebsstundenzählers des Zuluftventilators
<b>RLT03 Frequenzumrichter Zuluftventilator</b>	Binäres Betriebssignal des Frequenzumrichters des Zuluftventilators
<b>RLT03 Betriebsstufe Kältemaschine Stufe 1</b>	Zeitableitung des Betriebsstundenzählers der Kältemaschine für Stufe 1
<b>RLT03 Betriebsstufe Kältemaschine Stufe 2</b>	Zeitableitung des Betriebsstundenzählers der Kältemaschine für Stufe 2
<b>RLT03 Betriebsstufe Kältemaschine Stufe 3</b>	Zeitableitung des Betriebsstundenzählers der Kältemaschine für Stufe 3
<b>RLT03 Betriebsstufe Kältemaschine Stufe 4</b>	Zeitableitung des Betriebsstundenzählers der Kältemaschine für Stufe 4
<b>RLT03 Betriebsstufe Luftkühler gesamt</b>	Zeitableitung des Betriebsstundenzählers des Luftkühlers
<b>RLT03 Tagesmittelwert Wärmemengenzähler statische Heizung 0</b>	Ableitung des Wärmemengenzählers der statischen Heizanlage 0; zudem als Tagesmittelwert bestimmt;
<b>RLT03 Tagesmittelwert Wärmemengenzähler Lüftungsanlage</b>	Ableitung des Wärmemengenzählers der Lüftungsanlage zudem als Tagesmittelwert bestimmt;

**Tabelle 3: Aufzählung und Erläuterung verwendeter Datenpunkte der Anlage 03**

Bei der Datenanalyse von RLT03 werden sämtliche Werte zwischen dem 01.01.2017 00:00 Uhr und dem 05.11.2017 23:45 Uhr verwendet. Im Gegensatz zu RLT06, hat RLT03 keine längeren Sensorausfälle vorzuweisen. Wie schon bei Lüftungsanlage 06

lässt sich feststellen, dass das Gebäude in der Regel von Montag bis Samstag und zusätzlich am 19.03 und am 28.05 betrieben wurde. Es wird angenommen, dass es sich um verkaufsoffene Sonntage handelt. Die Anlage wurde am 14.04.2017, am 17.04.2017, am 01.05.2017 und am 31.10.2017 nicht betrieben, was auf verschiedene Feiertage zurückzuführen ist. Bei der Untersuchung der Lüftungsanlage wurde festgestellt, dass neben den eigentlichen Betriebszyklen ein Anlagenbetrieb zwischen 2:00 Uhr und 4:00 Uhr zu erkennen war.

### **3.3 Übertragung des Datenanalyseverfahrens auf raumluftechnische Anlagen**

In diesem Kapitel soll aufgezeigt werden, wie die verwendeten Algorithmen genutzt wurden, um zu den Ergebnissen der Datenanalyse zu gelangen. Dabei soll deutlich werden, in welchem Ausmaß die Implementierung dem Vorbild von [Habib et al., 2016] folgt und wo eigene Annahmen getroffen oder Änderungen vorgenommen wurden, um die Betriebszustandsklassifikation durchzuführen. Die hier vorliegende Datenanalyse weist eine Vielzahl unterschiedlicher Parameter auf, die das Ergebnis der Untersuchung signifikant beeinflussen können. Aus diesem Grund wurden die Parameter zunächst anhand von Lüftungsanlage 06 bestimmt, um diese im Anschluss auch auf Lüftungsanlage 03 anzuwenden.

Im Rahmen der Vorbereitung der Daten wurden diese zunächst auf eine einheitliche Zeitbasis gebracht. Dabei wurde ein Zeitintervall von 15 min gewählt. Zudem wurden zahlreiche Sensoren hinsichtlich ihrer Verwendbarkeit und Relevanz beurteilt und gegebenenfalls im Rahmen der Analyse vernachlässigt. So wurden beispielsweise Sensoren vernachlässigt, die den CO<sub>2</sub>-Gehalt der Abluft messen, da sie ein veritables Kriterium der Luftqualität darstellen, jedoch für die energetische Beurteilung der Gebäudeenergie-technik eine untergeordnete Rolle spielen. Datenpunkte, wie beispielsweise die Fortluftklappen der Anlage hätten ein adäquates Kriterium für den Betrieb der Anlage darstellen können, mussten aber aufgrund fehlerhafter Sensoraufzeichnungen vernachlässigt werden. Bei den Stundenzählern war es erforderlich, die Differenz zum jeweilig vorherigen Wert zu bilden. Somit kann die Betriebsdauer einzelner Komponenten der Anlage in der Analyse verwendet werden. Dies wäre mit der kumulierten Betriebsstundenzahl nicht möglich. Sensoren, bei denen der größte Teil der Aufzeichnungen fehlerhaft war, wurden in der Datenanalyse nicht berücksichtigt. Dies hat sich bei Lüftungsanlage 06 für den Zeitraum zwischen Januar bis März bewahrheitet, weswegen dieser bei der Analyse des Anlagensystems vernachlässigt werden musste.

#### **3.3.1 Unterscheidung von AN- & AUS-Zuständen**

Die Unterscheidung von Betriebszustand und Ruhezustand der Lüftungsanlage ist die Grundvoraussetzung für die spätere Klassifikation der Betriebszustände. Um diese Unterscheidung durchzuführen wurde der k-Means-Algorithmus verwendet. Da die Anlage in diesem Schritt in zwei unterschiedliche Zustände unterschieden werden sollte, wurde  $k = 2$  gewählt.

Das Ergebnis des für die Unterscheidung verwendeten k-Means-Algorithmus ist von verschiedenen Faktoren abhängig. Die korrekte Normierung der Eingangsdaten des Algorithmus ist ein essentieller Einflussfaktor für die korrekte Unterscheidung der Betriebszyklen. Es wurden drei verschiedene Normierungsverfahren genutzt und deren Auswirkungen auf das Ergebnis der Datenanalyse untersucht. Zudem kann die Auswahl verschiedener Sensoren das Ergebnis nachhaltig beeinflussen. Es wurden verschiedene Datenpunktauswahlen getestet und die Ergebnisse dargestellt. Um die Ergebnisse der Unterscheidung von AN- & AUS-Zuständen zu verifizieren, wurde ein Kontrollwert aus den Daten der Betriebssensoren der Frequenzumrichter des Zuluft- und des Abluftventilators gebildet. Es wurde angenommen, dass sich der Einsatz der Ventilatoren als entscheidendes Kriterium eignet, um zu beurteilen, ob die Anlage in Betrieb war.

Nachdem der Anlagenbetriebszustand korrekt ermittelt wurde, können aufeinanderfolgende Werte des gleichen Status zu AN- oder AUS-Zyklen zusammengefasst werden. Die Menge der gefundenen AN-Zyklen wird im nächsten Schritt durch die Symbolic Aggregate Approximation weiterverarbeitet.

### **3.3.2 Umsetzung Symbolic Aggregate Approximation und Bag of Words Representation**

Im Rahmen der Symbolic Aggregate Approximation werden Daten normalerweise in Zeitperioden unterteilt und der Mittelwert über diese gebildet. Da die Eingangsdaten bereits einen zeitlichen Abstand von 15 min aufwiesen, wurde bei der Datenanalyse auf diesen Schritt verzichtet. Dennoch könnte dieser Schritt das Ergebnis der Datenanalyse beeinflussen und sollte in anderen Fällen beachtet werden. Im zweiten Schritt der SAX-Transformation wurden die Zeitreihen in diskrete Wertintervalle unterteilt und jedem dieser Werte ein individuelles Symbol zugeordnet. Diese Wertintervalle werden auch Chunks genannt und sind einer der Parameter, deren Veränderung im Ergebnisteil der Arbeit exemplarisch dargestellt wird.

Diese Symbole werden beim Übergang in die BoWR aufsummiert und die zeitliche Reihenfolge somit gänzlich vernachlässigt. Durch die relativen Häufigkeiten des Vorkommens der Symbole in jedem Zyklus können diese verglichen werden. Essentiell ist dabei, dass die Grenzen der Intervalle für den gesamten Betrachtungszeitraum identisch festgelegt werden müssen, weil ansonsten eine Vergleichbarkeit nicht gegeben ist.

Auch ist es laut [Lin et al., 2007] notwendig, die Daten unter Verwendung der Z-Score-Normierungsmethode zu standardisieren. Nimmt man nun an, dass zwei Beispielwerte sich sehr ähnlich sind, würden diese zu unterschiedlichen Intervallen zugeordnet werden, insofern man sehr kleine Intervalle wählt. Wählt man größere Intervalle, so würden die genannten Beispielwerte womöglich im gleichen Intervall liegen, also als identisch angesehen werden. Aus diesem Grund ist die Anzahl der gewählten Wertintervalle ein zu beachtender Faktor. Im Folgenden werden verschiedene Anzahlen an Chunks gewählt, um die Unterschiede bei einem sonst identischen Datensatz aufzuzeigen. Anschließend soll die Umsetzung des hierarchischen Clustering beschrieben werden.

### 3.3.3 Umsetzung Hierarchisches Clustering

Das zu Klassifikation der Betriebszustände verwendete hierarchische Clustering ist der komplexeste Schritt der Analyse. Die eingehenden Daten entspringen der im Vorherigen erwähnten Bag of Words Representation. Anhand der relativen Häufigkeiten zwischen jeweilig gleichen Symbolen kann die Differenz gebildet werden. Daraus können die paarweisen euklidischen Distanzen zwischen allen Betriebszyklen gebildet werden. Die dabei entstehende Distanzmatrix wird bei der Clusteranalyse verwendet, um in Form eines Dendrogramms die Ähnlichkeiten der Zyklen abzubilden. Diese Ähnlichkeiten werden dabei zum einen durch den Linkage Type beeinflusst. Dieser wurde im Rahmen der Untersuchung nach dem Vorbild von [Habib, 2016, S13] als Average Methode verwendet.

Auch die Auswahl der Datenpunkte kann das Ergebnis deutlich beeinflussen. Aus diesem Grund wurde bei der Untersuchung darauf geachtet, dass alle für den Betrieb ausschlaggebenden Anlagenkomponenten im Datensatz vertreten sind. Zudem wurde die Außentemperatur als hauptsächliche Ursache für jeglichen Klimatisierungsbedarf bei der Datenanalyse berücksichtigt. Diese Auswahl kann ohne das Verständnis der zugrundeliegenden Gebäudeenergiesysteme nicht getätigt werden und muss somit als Expertenwissen verstanden werden.

Bei der Verwendung der hierarchischen Clusteranalyse muss stets entschieden werden, wie man das Analyseergebnis partitioniert. Dazu muss die Anzahl sinnvoller Cluster bestimmt werden. Aus diesem Grund wurde bei [Habib et al., 2016] die GAP-Statistic zur Festlegung der besten Clusterzahl verwendet. Diese hat trotz korrekter Implementierung zu keiner erfolgreichen Evaluation der Cluster geführt. Aus diesem Grund wurde ein anderer Ansatz zur Festlegung der Idealen Clusterzahl gewählt. Die Festlegung der Cluster-Anzahl beim hierarchischen Clustering entscheidet, auf welcher Höhe das resultierende Dendrogramm abgeschnitten wird.

Die Relation zwischen den einzelnen Werten des Datensatzes bleibt unabhängig von der Clusterzahl jedoch gleich. Das bedeutet, dass ein Ergebnis mit zu wenigen Clustern lediglich zu ungenau, nicht aber falsch ist. Aus diesem Grund wurde eine Grenze für das Abschneiden des Dendrogramms festgelegt, die sich an einem Prozentsatz der maximalen Distanz der Elemente der Datenanalyse orientiert. Der beste Schwellenwert wurde anhand von Lüftungsanlage 06 empirisch bestimmt und im Anschluss auch für Lüftungsanlage 03 verwendet. Dies ist ein Bestandteil der Analyse und kann in Kapitel 4 nachvollzogen werden.

### 3.3.4 Umsetzung der Cluster Evaluation der hierarchischen Clusteranalyse

Um die bei der Datenanalyse gefundenen Gruppierungen von Gebäudebetriebszyklen zu bewerten, wurden verschiedene Bewertungskriterien festgelegt. Jedes dieser Kriterien wurde für jeden Betriebszyklus berechnet. Danach wurde der Mittelwert für jedes Kriterium im Cluster gebildet. Cluster können somit bewertet und verglichen werden.

Bei [Habib et al., 2016, S 16] wurde der Coefficient of Performance verwendet um die Betriebszyklen der gebäudetechnischen Anlage, die sich nun in unterschiedlichen Clustern befinden nach ihrer Betriebsqualität zu bewerten. Dieser Gedanke wurde aufgefasst

und durch eigens für Lüftungsanlagen definierte Maßzahlen umgesetzt. Anhand von Tabelle 4 können die alle verwendeten Kontrollkriterien für den Anlagenbetrieb nachvollzogen werden.

<b>Evaluationswert</b>	<b>Aussage</b>	<b>Deutung</b>
<b>Anz_Zyklen_je_Cluster</b>	Gibt an, wie viele Zyklen in jedem Cluster enthalten sind	Indikator für die Mächtigkeit des gefundenen Clusters
<b>Häufigster Monat (nominal)</b>	Gibt an, in welchem Monat die meisten Zyklen des jeweiligen Clusters liegen	Lässt den Monatsschwerpunkt des Clusters erkennen
<b>Monatszahl (Durchschnitt)</b>	Der Monat, in dem sich der Zyklus befindet wird als Integer extrahiert und dann über alle Zyklen des jeweiligen Clusters der Mittelwert errechnet.	In Verbindung mit dem nominal häufigsten Monat lässt sich erkennen, welche Tendenz beim Vorkommen der Zyklen besteht.  z.B. wenn der Juni der nominal häufigste Monat ist und der Monatsdurchschnitt bei 6,7 liegt, kann man erkennen, dass der nächstgrößere Teil der Zyklen des Clusters in den späteren Monaten des Jahres liegen muss.
<b>Zyklusdauer[h] (Mean) &amp; Zyklusdauer[h] (Standard deviation)</b>	Zeigt die durchschnittliche Dauer aller Zyklen bzw. die Standardabweichung im Cluster an.	Eine Zyklusdauer, die stark von der erwarteten Gebäudenutzungsdauer abweicht kann auf einen unsachgemäßen Anlagenbetrieb oder ein Fehlverhalten selbiger hinweisen.  Durch die Standardabweichung kann einfach erkannt werden, wenn sich unterschiedlich lange Betriebszyklen im gleichen Cluster befinden.
<b>Mittlere Heizleistung [kW] (Mean) &amp;</b>	Für die mittlere Heizleistung, die für jeden Zyklus bestimmt wurde,	Da bei der Berechnung die Außenlufttemperatur von der Zulufttemperatur abgezogen

<b>Mittlere Heizleistung [kW] (Standard deviation)</b>	wird sowohl der Mittelwert für alle Zyklen des jeweiligen Clusters, als auch die Standardabweichung dessen gebildet.	<p>wurde, kennzeichnet ein negatives Ergebnis den Kühlbedarf des Gebäudes und ein positives Ergebnis umgekehrt den Heizbedarf.</p> <p>Deutlich negative Mittelwerte weisen also auf den Kühlbetrieb der Anlage hin, deutlich positive Mittelwerte umgekehrt auf Heizbetrieb hin. Ist der Mittelwert in der Nähe des Nullpunktes zu finden, handelt es sich entweder um einen Cluster, dessen enthaltene Zyklen einen Übergangsbereich zwischen Heizbetrieb und Kühlbetrieb, oder ein unsauberes Clustering.</p> <p>Die Standardabweichung hilft dabei die Ähnlichkeit der einzelnen Zyklen im Cluster zu erkennen und somit die Authentizität der gefundenen Cluster nachvollziehen zu können.</p>
<b>Durchschnittlicher WRG-Betriebsanteil _im_Cluster</b>	Dieser Wert gibt das Verhältnis von durchschnittlicher WRG-Betriebsstundenzahl zu durchschnittlicher Zykllänge an.	Durch diesen Wert kann schnell sichtbar gemacht werden, zu welchem Anteil die WRG im den Zyklen des jeweiligen Clusters im Betrieb war, was Aufschluss über das Anlagenverhalten geben kann.
<b>Betriebsstunden WRG (Mean) &amp; Betriebsstunden WRG (Standard deviation)</b>	Zeigt an, wie viele Stunden die WRG in allen Zyklen des jeweiligen Clusters im Durchschnitt gelaufen ist.	
<b>Mittlere Aussentemperatur (Mean) &amp;</b>	Gibt den Mittelwert, bzw. die Standardabweichung der mittleren	Da die Außentemperatur unmittelbaren Einfluss auf die Betriebsart von Lüftungsanlagen hat, kann der Wert dafür

<b>Mittlere Aussentemperatur (Standard deviation)</b>	Außentemperaturen aller Zyklen des jeweiligen Clusters wieder.	genutzt werden, die gefundenen Cluster zu verifizieren.
<b>Durchschnittliche Startstunde der Zyklen je Cluster</b>	Gibt den durchschnittlichen Anfangszeitpunkt aller Zyklen des jeweiligen Clusters an.	Der Wert kann Aufschluss darüber liefern, wie einheitlich Anlagen betrieben wurden, bzw. ob es im Tagesverlauf zu einem großen, oder mehreren kleinen Betriebszyklen kommt.
<b>Intra_Cluster_Dist/ Number_pairwise_distances</b>	Nachdem die Betriebszyklen in die Symbolrepräsentation des BoWR umgewandelt wurden, können zwischen allen Zyklen eines Clusters die paarweisen Euklidischen Distanzen berechnet werden.  Dieser Wert gibt die durchschnittliche Euklidische Distanz wieder.	Der Wert zeigt die Streuung aller Zyklen im Cluster an. Je weiter ein Cluster gestreut ist, desto größer der Wert.

**Tabelle 4: Erläuterung der Kontrollwerte der Evaluationstabelle**

Die nach dem Vorbild der Tabelle erstellten Werte, werden in Kapitel 4 bei der Beurteilung der Ergebnisse der Datenanalyse Verwendung finden. Dabei werden die für die Beurteilung ausschlaggebenden Werte in Form von Tabellen im Text dargestellt. Die unter Nutzung der Tabelle zu erkennenden Betriebszustände sind vielseitig und von der Beschaffenheit der jeweiligen Anlage abhängig. Im folgenden Kapitel soll das zu erwartende Anlagenverhalten erläutert werden. Dabei werden die gängigen Betriebsmodi anhand einzelner Komponenten der Anlagensysteme definiert und benannt.

### 3.4 Zu erwartende Betriebszustände

Im Folgenden sollen die zu erwartenden Betriebszustände der zu untersuchenden Anlagen benannt und anhand verschiedener, für die Klassifikation ausschlaggebender Komponenten differenziert werden. Zudem wird auf die zu erwartenden Betriebszeiten der zu untersuchenden Anlagensysteme eingegangen. Diese ergeben sich aus dem Nutzungsverhalten der Liegenschaft.

Der korrekte Betrieb raumlufttechnischer Anlagen ist von einer Vielzahl an Faktoren, wie klimatischen Bedingungen, Nutzerverhalten und Betrieb von elektronischen Geräten abhängig. Deswegen wird der jeweilige Aufbau der Anlage den Anforderungen an die Klimatisierung, sowie örtlichen Gegebenheiten, wie beispielsweise der Gebäudegeometrie

angepasst. In der folgenden Tabelle werden die über den Jahresverlauf auftretenden Zustände aufgeführt und die Zustände der jeweiligen Komponenten während des jeweiligen Betriebszustandes dargestellt.

Die in Tabelle 5 gezeigten Anlagenzustände setzen voraus, dass die Anlage fehlerfrei betrieben wird. Beim Eintreten eines Heizfalls ist das „Heizen mit WRG“ der hauptsächlich zu erwartende Betriebsfall. Die Wärme der Abluft wird dazu genutzt, die einströmende Außenluft zusätzlich zu erhitzen. Das Heizen ohne den Einsatz der Wärmerückgewinnung wäre nur dann sinnvoll, wenn die Temperatur der Abluft kälter ist, als die der Außenluft, was unter Umständen zu Beginn eines Betriebszyklus sinnvoll sein kann, aber keinen Normalfall im Anlagenbetrieb ausmacht. Einen Kühlvorgang durch eine WRG zu unterstützen ist bei sehr hohen Außenlufttemperaturen sinnvoll. Nimmt man an, dass die Abluft eine geringere Temperatur aufweist als die einströmende Außenluft, so kann die Anlage durch Wärmerückgewinnung in ihrer Funktion unterstützt werden. Das Kühlen ohne Wärmerückgewinnung ist der wahrscheinlichere Kühl-Fall, da durch den oftmals zu geringen Temperaturunterschied von Abluft und Außenluft die unterstützende Kühlfunktion der WRG nicht zum Tragen kommt.

Anlagenteile	Anlage aus	Heizen mit WRG	Kühlen mit WRG	Kühlen ohne WRG
Luftkühler	aus	aus	an	an
LE	aus	an	aus	aus
WRG	aus	an	an	aus
ZU/AB - Ventilator	aus	an	an	an

**Tabelle 5: zu erwartende Betriebszustände der raumluftechnischen Anlagensysteme**

Neben den in der Tabelle erwähnten zu erwartenden Anlagenbetriebszuständen kann es im Bereich gemäßiger Temperaturen dazu kommen, dass das Gebäude in den frühen Morgenstunden geheizt und zur Mittagszeit gekühlt wird. Da ein Betriebszyklus meist eine Dauer von ca. 12 h aufweist, ist zu vermuten, dass unter den genannten Voraussetzungen Zyklen auftreten können, bei denen sowohl Heizbetrieb als auch Kühlbetrieb festgestellt wird. Diese Zyklen und die daraus resultierenden Cluster werden im Folgenden als Übergangszyklen bzw. Übergangcluster bezeichnet werden.

## 4 Ergebnisse der Datenanalyse & Diskussion

In diesem Kapitel sollen die Ergebnisse der durchgeführten Datenanalyse dargestellt werden.

### 4.1 Analyse und Auswertung von Lüftungsanlage 06

Im diesem Abschnitt werden die Ergebnisse der Analyse von RLT 06 vorgestellt und gedeutet. Die Untersuchung der Anlage wurde dafür genutzt, die in Kapitel 3.3 benannten Parameter hinsichtlich ihrer Auswirkung auf das Ergebnis der Datenanalyse zu testen.

#### 4.1.1 Unterscheidung von AN- und AUS-Zyklen

Im ersten Schritt der Klassifikation von Betriebszuständen müssen die Daten durch ein k-Means-Clustering in AN- und AUS-Zyklen unterteilt werden. Ziel ist es, die Einsetzbarkeit des verwendeten Algorithmus zu verdeutlichen und die Auswirkungen von Datenpunktauswahl und Normierungsmethode darzustellen.

Im Rahmen der Arbeit wurde ein Kriterium zur Unterscheidung von AN- und AUS-Zuständen unter Verwendung von Anlagenkenntnis definiert. Dieser Kontrollwert wurde aus den Betriebssensoren der Frequenzumrichter von Zuluftventilator und Abluftventilator gebildet. Sobald die Ventilatoren in Betrieb sind, wird im Kontrollwert eine Eins übergeben, andernfalls eine Null. Im Rahmen der Diplomarbeit wurde angenommen, dass man bei dem Betrieb der Ventilatoren davon ausgehen kann, dass das gesamte System in Betrieb ist. Dafür ist es nötig, den Aufzeichnungen der jeweiligen Datenpunkte zu vertrauen. Anhand der erzeugten Kontrollwerte wurden verschiedene k-Means-Szenarien durchgeführt. Dabei wurden zum einen verschiedene Normierungsverfahren angewandt und zum anderen unterschiedliche Sensorauswahlen getätigt.

#### Szenario 1

Grundgedanke des ersten Szenarios war, alle Datenpunkte an das k-Means-Clustering zu übergeben. Dadurch soll verifiziert werden, dass eine exakte Auswahl der Datenpunkte für ein erfolgreiches k-Means-Clustering nicht notwendig ist und somit auch kein Expertenwissen voraussetzt. Im ersten Untersuchungsszenario wurden folgende Datenpunkte verwendet:

- RLT06 Ablufttemperatur
- RLT06 Abluftvolumenstrom
- RLT06 Außentemperatur
- RLT06 LE-Rücklauftemperatur
- RLT06 LE-Ventilstellung
- RLT06 Betriebssensor WRG-Pumpe
- RLT06 WRG-Temperatur
- RLT06 WRG-Ventilstellung
- RLT06 Zulufttemperatur
- RLT06 Zuluftvolumenstrom

- RLT06 Betriebssensor Kältemaschine
- RLT06 Betriebssensor Kältemaschine Stufe 1
- RLT06 Betriebssensor Kältemaschine Stufe 2
- RLT06 Betriebssensor Kältemaschine Stufe 3
- RLT06 Betriebssensor Kältemaschine Stufe 4
- RLT06 Betriebsstufe Abluftventilator
- RLT06 Betriebsstufe LE-Pumpe
- RLT06 Betriebsstufe WRG-Pumpe
- RLT06 Betriebsstufe Zuluftventilator
- RLT06 Betriebsstufe Kältemaschine Stufe 1
- RLT06 Betriebsstufe Kältemaschine Stufe 2
- RLT06 Betriebsstufe Kältemaschine Stufe 3
- RLT06 Betriebsstufe Luftkühler gesamt
- RLT06 Tagesmittelwert Wärmemengenzähler statische Heizung 1
- RLT06 Tagesmittelwert Wärmemengenzähler statische Heizung 0
- RLT06 Tagesmittelwert Wärmemengenzähler Lüftungsanlage

In der Datenanalyse wurde der folgende Plot zur Veranschaulichung des Clustering-Ergebnisses verwendet. Dabei ist der Abluftvolumenstrom in allen Datenpunkten über den gesamten Jahresverlauf zu erkennen. Der dabei in Rot dargestellte „AN-Cluster“ entspricht der Menge aller AN-Zustände und der schwarze „AUS-Cluster“ der Menge aller AUS-Zustände. Im Bereich zwischen 18.000 m<sup>3</sup> / h und 19.000 m<sup>3</sup> / h sind vereinzelt schwarze Punkte zu erkennen. Aus diesem farblichen Unterschied lässt sich ein anteiliger Fehler im Clustering aufzeigen. Schwarze Punkte, die Teil des Aus-Cluster sind, sollten nur bei einem Luftvolumenstrom von 0 m<sup>3</sup> / h auftreten. Das Resultat kann in Abbildung 11 nachvollzogen werden.

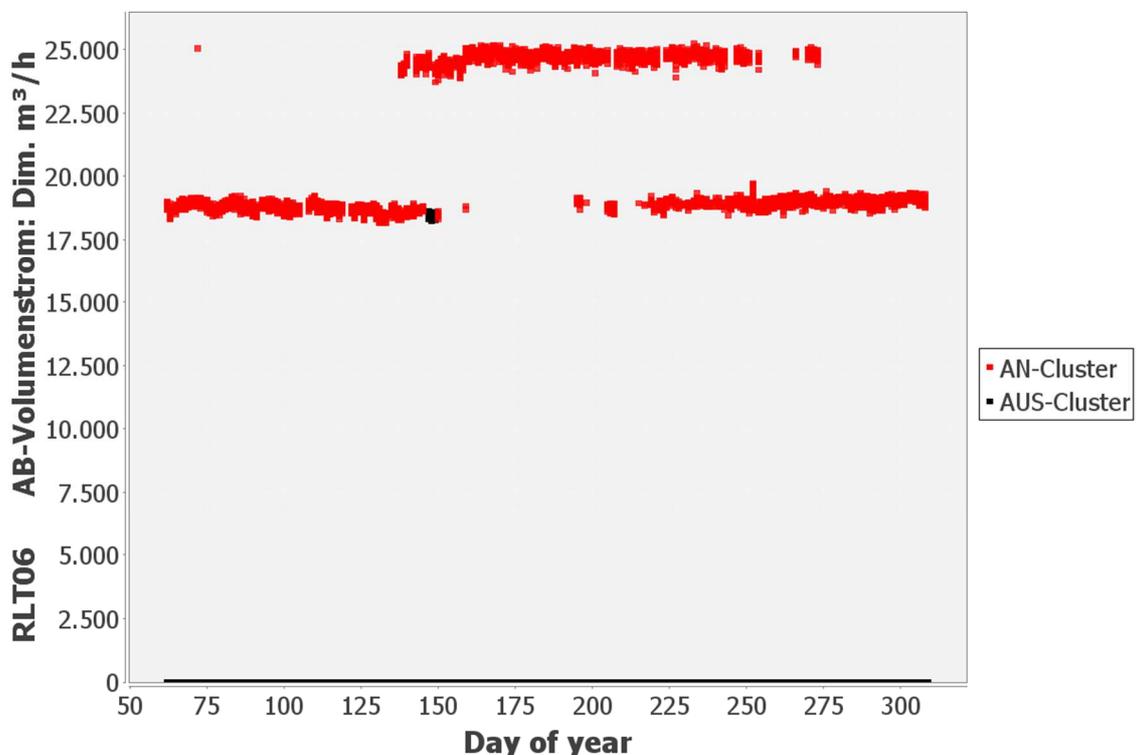


Abbildung 11: RLT06 Darstellung des Abluftvolumenstroms über den Jahresverlauf mit dem Ergebnis der Clusteranalyse in verschiedenen Farben

In Tabelle 6 ist die Qualität der Unterscheidung von AN- und AUS-Zuständen anhand verschiedener Normierungsverfahren dargestellt. Dabei ist der Prozentsatz der jeweiligen Fehlerzustände ein ausschlaggebender Faktor für die Beurteilung der Normierungsqualität. Dieser orientiert sich an der Anzahl falsch zugeordneter Anlagenbetriebszustände bezogen auf den gesamten Betrachtungszeitraum. Aufgrund einer sehr geringen Fehlerquote wird ersichtlich, dass die MINMAX-Normierungsmethode das beste Ergebnis hervorgebracht hat.

Normierungsverfahren	Fehlerquote	AN	AUS
Z-Score	0,44 %	42,21 %	57,37 %
MINMAX [-1, 1]	0,029 %	42,56 %	57,41 %
Decimal Scaling	13,92 %	28,68 %	57,40 %

**Tabelle 6: RLT06 k-Means Szenario 1 Vergleich der Ergebnisse mit verschiedenen Normierungsmethoden**

### Szenario 2:

In Szenario 2 wird nur ein Teil der vorhandenen Datenpunkte verwendet. Da es bei anderen Anlagen nicht zwangsläufig der Fall sein muss, dass ON/OFF-Sensoren vorhanden bzw. intakt, oder korrekt sind, wurden diese entfernt um zu verifizieren, dass auch so ein korrektes Clustering zustande kommt. Auf die Wärmemengenzähler wurde verzichtet, weil diese in Ihrer Beschaffenheit als Tagesmittelwerte mit dem Ergebnis der Clusteranalyse nicht in Korrelation stehen. Dies kann als Wissensgrundlage im Bereich der Datenanalyse, allerdings nicht als Expertenwissen bezüglich der Gebäudeenergie-technik verstanden werden, weswegen es sich mit der von Habib et al. gestellten Maxime deckt. Für das Clustering wurden folgende Datenpunkte ausgewählt:

- RLT06 Ablufttemperatur
- RLT06 Abluftvolumenstrom
- RLT06 Außentemperatur
- RLT06 LE-Rücklauftemperatur
- RLT06 LE-Ventilstellung
- RLT06 WRG-Temperatur
- RLT06 WRG-Ventilstellung
- RLT06 Zulufttemperatur
- RLT06 Zuluftvolumenstrom
- RLT06 Betriebsstufe Abluftventilator
- RLT06 Betriebsstufe LE-Pumpe
- RLT06 Betriebsstufe WRG-Pumpe
- RLT06 Betriebsstufe Zuluftventilator
- RLT06 Betriebsstufe Kältemaschine Stufe 1
- RLT06 Betriebsstufe Kältemaschine Stufe 2
- RLT06 Betriebsstufe Kältemaschine Stufe 3
- RLT06 Betriebsstufe Luftkühler gesamt

Die anteilige Fehlerquote beim zweiten Untersuchungsszenario kann anhand Tabelle 7 nachvollzogen werden. Erneut hat sich bestätigt, dass die Normierung mit der MINMAX-Methode zur geringsten Fehlerquote geführt hat. Durch das Entfernen verschiedener Datenpunkte hat sich das Ergebnis hingegen nur geringfügig verbessert.

Normierungsverfahren	Fehlerquote	AN	AUS
Z-Score	0,26 %	42,38 %	57,36 %
MINMAX [-1, 1]	0,0042 %	42,60 %	57,39 %
Decimal Scaling	17,23 %	25,35 %	57,42 %

**Tabelle 7: RLT06 k-Means Szenario 1 Vergleich der Ergebnisse mit verschiedenen Normierungsmethoden**

Bei beiden Szenarien der Unterscheidung der AN- und AUS-Zustände anhand k-Means lässt sich erkennen, dass das MINMAX-Scaling zu den besten Ergebnissen geführt hat. Dies kann unterschiedliche Ursachen haben, deren Untersuchung den Rahmen dieser Diplomarbeit übersteigt. Bei der MINMAX-Normierung besitzen alle Daten den gleichen Wertebereich. Somit wird angenommen, dass allen Datenpunkten die gleiche Gewichtung im Rahmen der Clusteranalyse zukommt.

Die Auswahl der Sensoren scheint eine untergeordnete Rolle zu spielen, insofern man bei der Datenauswahl alle ausschlaggebenden Komponenten der Lüftungsanlage berücksichtigt. Dennoch kann eine minimale Verbesserung der Ergebnisse durch eine bessere Sensorauswahl erzielt werden. Zusammenfassend kann über das k-Means-Clustering ausgesagt werden, dass es im Rahmen der Diplomarbeit erfolgreich für die Unterscheidung von Betriebszuständen und Nichtbetriebszuständen für RLT06 verwendet wurde.

#### 4.1.2 Klassifizierung der Betriebszustände

Im zweiten Teil der Analyse soll die Klassifizierung von verschiedenen Betriebszuständen vorgenommen werden. Da im Folgenden nur noch AN-Zyklen betrachtet werden, werden sämtliche Daten, die außerhalb der täglichen Betriebszeiten, oder an Sonn- und Feiertagen aufgezeichnet wurden, vernachlässigt.

Bei der Untersuchung wurden verschiedene Parameter erkannt, welche einen Einfluss auf das Ergebnis der Datenanalyse haben. Diese sollen im Folgenden hinsichtlich der Auswirkung auf das Clustering untersucht werden. Es handelt sich dabei um folgende Variablen:

- Anzahl der Cluster aus Cluster-Evaluation anhand eines Schwellenwerts
- Chunk- Anzahl bei der SAX - Transformation & BoWR - Representation
- Auswahl der Sensoren für das hierarchische Clustering

Bevor man die Clusteranalyse durchführt, muss zunächst eine Auswahl der in Betracht kommenden Datenpunkte getroffen werden. Die Auswahl repräsentiert die Komponenten der Anlage, welche für die Beurteilung des Anlagenbetriebs primäre Bedeutung besitzen. Zudem wurde die Außentemperatur hinzugefügt, weil diese die Ursache des jeweiligen Betriebsverhaltens darstellt. Die gewählten Sensoren können in Tabelle 8 nachvollzogen werden. Diese Datenpunktauswahl wird im Folgenden auch als hierarchische Clusteranalyse (HC I) bezeichnet.

Datenpunkte
RLT06 Außentemperatur
RLT06 LE-Ventilstellung
RLT06 WRG-Ventilstellung
RLT06 Betriebsstufe Luftkühler gesamt

Tabelle 8: Verwendete Datenpunkte für die hierarchische Clusteranalyse I bei RLT06

#### 4.1.3 Outlierproblematik

Unabhängig von der Einstellung der bereits erwähnten Parameter haben sich bei jedem Clustering „Ausreißer-Cluster“ ergeben. Dabei handelt es sich um Cluster, die nur wenige Zyklen beinhalten und oftmals in der Clustering-Hierarchie weit oben stehen. In Abbildung 12 ist ein Teil eines Dendrogramms dargestellt, welches mehrere Ausreißer enthält. Diese wurden farbig auf dem Abbildungsausschnitt markiert. Die drei markierten Ausreißer auf der linken Seite resultieren in Clustern, die nur einen Zyklus enthalten. Die Zyklen, die auf der rechten Seite markiert wurden, würden bei genügend hoher Clusterzahl in zwei Clustern mit nur einem Zyklus und in einem Cluster mit drei Zyklen resultieren. Diese Ausreißer-Cluster können auf einen fehlerhaften Anlagenbetrieb hinweisen oder Rückschluss auf fehlerhafte Sensoraufzeichnungen zulassen. Im Rahmen dieser Arbeit war die Untersuchung der genauen Ursache der entstehenden Ausreißer zeitlich nicht möglich.

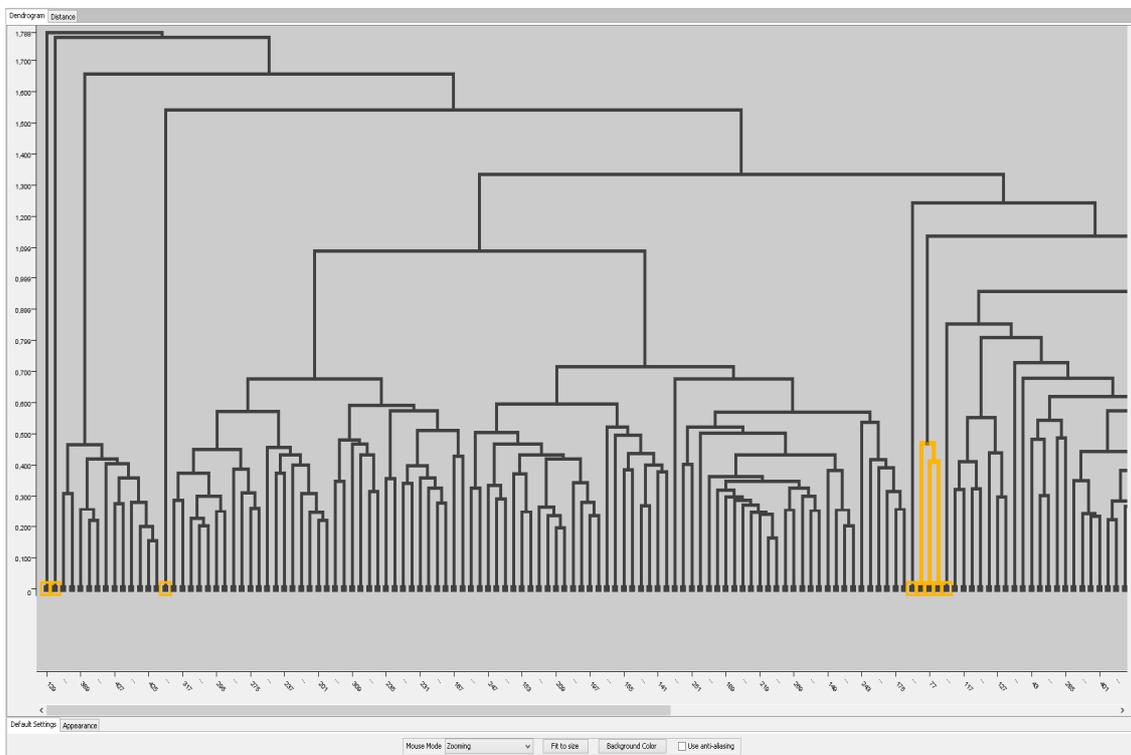


Abbildung 12: Darstellung von Outlier-Problematik im hierarchischen Clustering anhand RLT 06

#### 4.1.4 GAP-Statistic zu Cluster Evaluation

Bei der Evaluation der Cluster-Anzahl unter Verwendung der GAP-Statistic-Methode wurde festgestellt, dass die GAP-Statistic weder ein lokales, noch ein globales Maximum annimmt, sondern monoton ansteigt. Dadurch konnte die Anzahl der Cluster mit der Methode nicht ausgewählt werden. Dies ist vermutlich mit den im Datensatz befindlichen Ausreißer-Clustern in Zusammenhang zu bringen. Die Ausreißer-Cluster haben oftmals nur eine geringe Intra-Cluster-Distanz, was womöglich das Ergebnis verfälscht. Würde man die Ausreißer-Cluster vor der Anwendung der GAP-Statistic entfernen, würde man möglicherweise ein besseres Ergebnis der GAP-Statistic erhalten. Dennoch würde diese Manipulation der Eingangswerte der Clusteranalyse ein verfälschtes Ergebnis zur Folge haben. Die exakte Aufklärung des Sachverhaltes bedarf weiterer Untersuchung.

#### 4.1.5 Einführung des distanzbasierten Schwellenwertes

Um die Datenanalysen untereinander vergleichen zu können, wurde die Wahl der verwendeten Clusterzahl anhand prozentualer Anteile der Clustering Distanzen aus der hierarchischen Clusteranalyse festgemacht. Eine Alternative ist, eine bestimmte Anzahl an Clustern festzulegen. Dennoch bestünde dabei die Gefahr, dass die oben erwähnten Ausreißer-Cluster das Ergebnis verfälschen könnten.

Um einen Schwellenwert für die weitere Untersuchung auszuwählen, wurde der gleiche Datensatz mit sonst identischen Parametern mehrmals bei verschiedenen Schwellenwerten geclustert. Diese sind in Abbildung 13 als horizontale Linien dargestellt. Für die Bestimmung ist das Dendrogramm für das Clustering der RLT06 mit den in Tabelle 8 angegebenen Datenpunkten verwendet worden. Zudem wurden 120 Chunks für die Untersuchung ausgewählt.

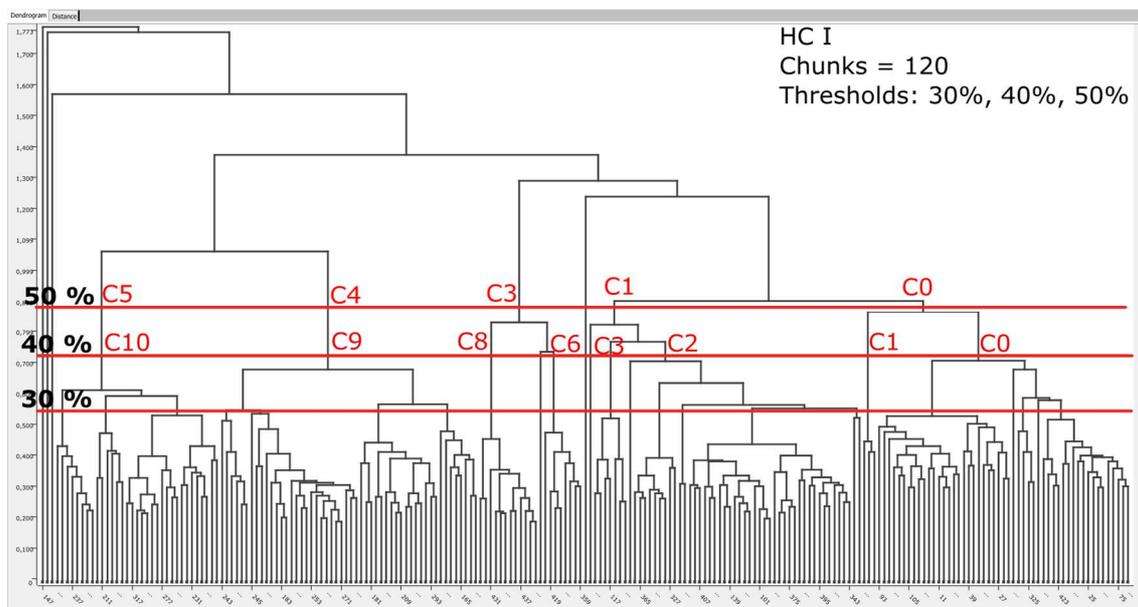


Abbildung 13: Darstellung verschiedener Schwellenwerte für das Clustering

Es wurden Schwellenwerte von 30 %, 40 % und 50 % getestet. Je niedriger man den Grenzwert auswählt, desto genauer wird das Clustering. Es werden mehr Cluster gefunden, da das Dendrogramm auf einer niedrigeren Hierarchiestufe gekappt wird. Die gefundenen Cluster enthalten in der Regel weniger Zyklen und eine geringere Streuung, sind also spezifischer. Durch einen zu geringen Grenzwert entstehen eine hohe Anzahl von Clustern, was wiederum einen erhöhten Analyseaufwand bedeutet, was anhand der Ergebnisse, erzeugt mit einem Grenzwert von 30 % zu erkennen ist. Wird der Grenzwert zu hoch angelegt, werden Cluster, die eigentlich getrennt aufgeführt werden sollen, in zu große und unpräzise Cluster zusammengefasst. Diese großen Cluster enthalten eine zu große Streuung, sind also zu ungenau. Je ungenauer die Clusterbildung, desto höher das Risiko Zyklen mit abweichenden Betriebsmodi als identisch anzunehmen.

Der Schwellenwert von 40 % wurde ausgewählt, weil er im Gegensatz zu der mit 50 % Schwellenwert durchgeführten Clusteranalyse Cluster 2 und Cluster 3 getrennt aufführt. Diese sind in Tabelle 9 dargestellt. Es handelt sich um einen Ausschnitt von Anlage 04. Dabei lässt sich erkennen, dass Cluster 2 eine deutlich niedrigere mittlere Außentemperatur aufweist. Der anteilige WRG-Betrieb weist eine Differenz von 30 % auf. Die mittlere Heizleistung von -47,49 kW weist Cluster 3 zudem als Kühlcluster aus. Wohingegen Cluster 2 mit einer mittleren Heizleistung von -3,14 kW und einer deutlich höheren Standardabweichung ein deutlich diffuseres Bild aufweist. Daher wird angenommen, dass die Trennung der Cluster für das Ergebnis der Analyse von Vorteil ist.

<b>Datenpunkte:</b>	<b>Cluster2</b>	<b>Cluster3</b>
<b>Anz Zyklen je Cluster</b>	47	7
<b>Häufigster Monat (nominal)</b>	September	Mai
<b>Monatszähl (Durchschnitt)</b>	7,57	4,86
<b>Zyklusdauer[h] (Mean)</b>	11,75	12,36
<b>Zyklusdauer[h] (Standard deviation)</b>	2,2	0,28
<b>Durchschnittlicher WRG-Betriebsanteil im Cluster</b>	<b>0,39</b>	<b>0,73</b>
<b>Betriebsstunden WRG (Mean)</b>	4,53	9,06
<b>Betriebsstunden WRG (Standard deviation)</b>	1,95	1,65
<b>Mittlere Heizleistung (Mean)</b>	-3,14	-47,49
<b>Mittlere Heizleistung (Standard deviation)</b>	22,96	20,85
<b>Energie je Zyklus kWh (Mean)</b>	-21,08	-589,55
<b>Energie je Zyklus kWh (Standard deviation)</b>	204,92	267,15
<b>Mittlere Außentemp (Mean)</b>	<b>18,74</b>	<b>26,51</b>
<b>Mittlere Außentemp (Standard deviation)</b>	2,73	3,35
<b>Durchschnittliche Startstunde der Zyklen je Cluster</b>	7,23	6,86
<b>Durchschnittliche Startstunde der Zyklen je Cluster (Standard deviation)</b>	1,6	0,38
<b>Intra-Cluster Distance/Number pairwise distances</b>	0,5	0,45

**Tabelle 9: RLT06 Ausschnitt aus der Evaluationstabelle zur Darstellung von Cluster 2 und Cluster 3 für HC I bei 120 Chunks und einem Schwellenwert von 40 %**

Bei Verwendung des Schwellenwertes von 30 % bei der Clusteranalyse ergab sich ein Analyseergebnis mit einem höheren Detaillierungsgrad. Dabei haben sich insgesamt 29

verschiedene Cluster ergeben, wovon viele Zyklen mit einem ähnlichen Verhalten voneinander getrennt wurden. Da im Rahmen der Analyse jedoch nur vier Betriebszyklen, aufgeführt in Tabelle 5, zu finden waren, wird das Ergebnis als zu detailliert angesehen. Bei der benannten Tabelle handelt es sich um einen Ausschnitt aus Anlage 03. Es sind viele Cluster, mit nur geringen Unterschieden entstanden, deren Trennen als nicht lohnenswert betrachtet wird. Beispielfhaft werden in Tabelle 10 die Cluster 22 bis 25 aufgeführt, die unter Verwendung des Schwellenwertes von 30 % der maximalen Distanz zustande kommen.

Die Zyklen aller dieser Cluster sind bei einem Schwellenwert von 40 % unter Cluster 10 zusammengefasst. Alle in der Tabelle dargestellten Cluster zeichnen sich durch eine negative mittlere Heizleistung, einen geringen WRG-Betriebsanteil im Cluster und eine mittlere Außentemperatur von mindestens 22 °C aus. Man kann also annehmen, dass die Anlage während dieser Zyklen zur Kühlung des Gebäudes genutzt wird und die Wärmerückgewinnungsanlage dabei nur wenig betrieben wird. Ähnliche Werte für die oben genannten Größen hat sich auch bei Cluster 10 der Clusteranalyse mit einem Schwellenwert von 40 % anhand der mittleren Außentemperatur von 24,7 °C, einem durchschnittlichen WRG-Betriebsanteil von 18 % und einer mittleren Heizleistung von -97,83 kW im Cluster erkennen lassen. In Anlage 04 können die Werte für Cluster 10 nachvollzogen werden.

Datenpunkte:	Cluster22	Cluster23	Cluster24	Cluster25
Anz_Zyklen_je_Cluster	8	11	6	8
HäufigsterMonat(nominal)	Juni	August	August	Juli
Monatszähl (Durchschnitt)	7,25	8	7,83	6,87
Zyklusdauer[h] (Mean)	11,41	12,09	12,25	10,69
Zyklusdauer[h] (Standard deviation)	2,39	0,53	0	2,91
Durchschnittlicher WRG-Betriebsanteil im_Cluster	7 %	9 %	29 %	33 %
BetriebsstundenWRG (Mean)	0,85	1,15	3,58	3,51
BetriebsstundenWRG (Standard deviation)	0,69	0,8	0,68	1,33
MittlereHeizleistung (Mean)	-68,08	-115,87	-82,81	-114,03
MittlereHeizleistung (Standard deviation)	13,81	26,57	23,36	23,93
MittlereAussentemp (Mean)	22,23	26,82	23,7	25
MittlereAussentemp (Standard deviation)	1,34	2,08	1,06	2,52
Durchschnittliche_Startstunde_der_Zyklen_je_Cluster	7,88	7,18	7	7,75
In-tra_Cluster_Dist/Number_pairwise_distances	0,39	0,35	0,44	0,37

Tabelle 10: RLT06 Evaluationstabelle für HC I mit 120 Chunks & Schwellenwert 30 % für die Cluster 22 bis 25

Aufgrund der im vorherigen beschriebenen Erkenntnisse wird angenommen, dass ein Schwellenwert von 40 % sich für die Datenanalyse eignet. Er wird auch in der weiteren Untersuchung verwendet werden. Aufgrund der zeitlichen Limitierung der hier angefertigten wissenschaftlichen Arbeit musste eine exakte Verifizierung dessen unterlassen werden. Es handelt sich um eine Annahme.

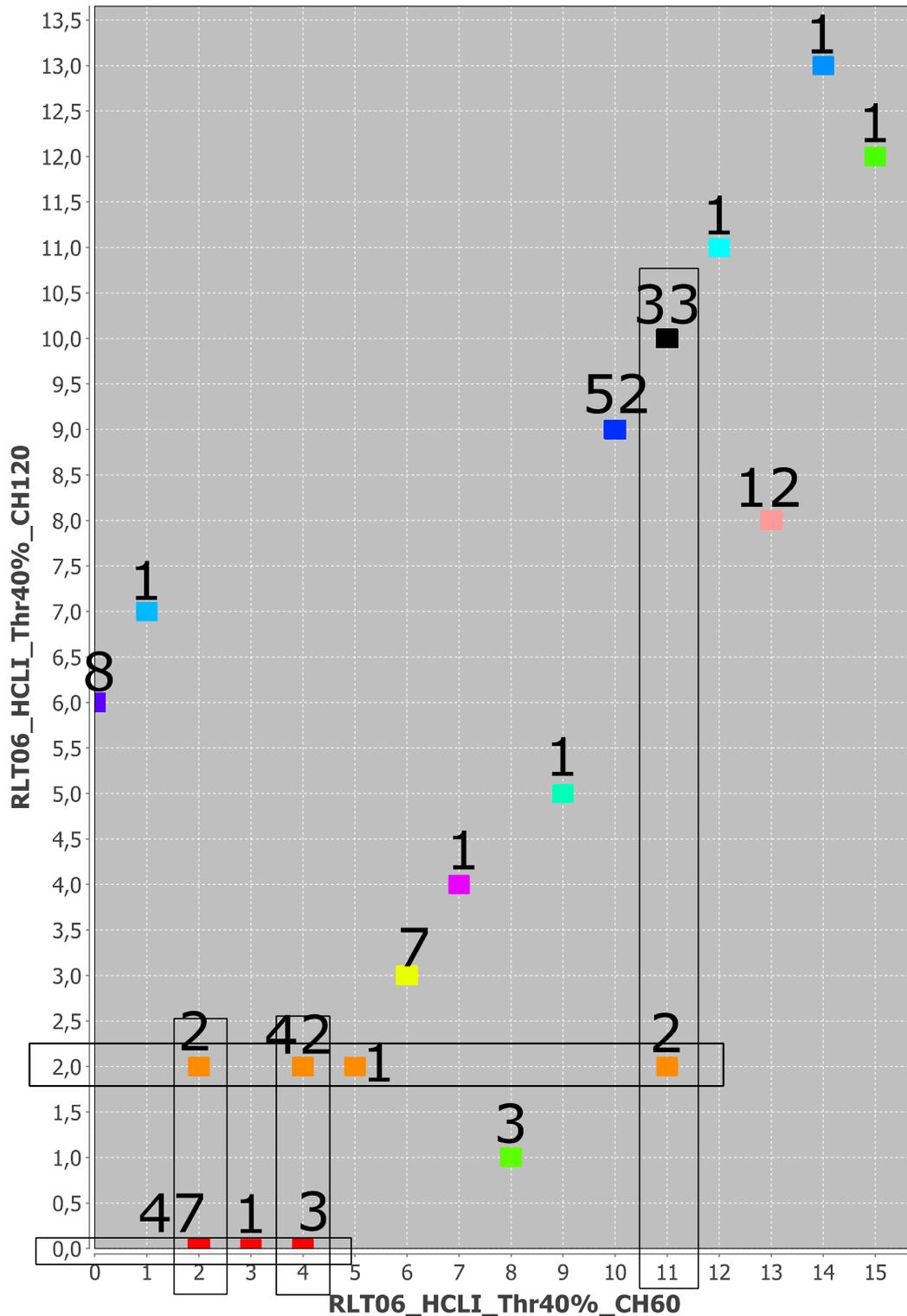
#### 4.1.6 Untersuchung der Auswirkung unterschiedlicher Chunk-Anzahlen

Im Folgenden soll die Auswirkung der Auswahl unterschiedlicher Chunks, also nicht überlappender, gleichgroßer Wertebereiche untersucht werden. Ausgehend von RLT06 wurden hierbei unterschiedliche Anzahlen von Chunks untersucht, um festzustellen welche Auswirkungen diese auf die Analyseergebnisse des hierarchischen Clustering haben. Die Ergebnisse werden im Folgenden dargestellt. Die bisher erwähnten Ergebnisse aus dem optischen Vergleich der Dendrogramme können anhand von Tabelle 11 nachvollzogen werden.

	60 Chunks	120 Chunks
<b>Maximale Distanz</b>	1,8	1,8
<b>Gesamtanzahl Cluster</b>	16	14

**Tabelle 11: maximale euklidische Distanz und Gesamtanzahl gefundener Cluster für verschiedene Chunk-Anzahlen bei Datenpunktauswahl HC I und einem Schwellenwert von 40 %**

Um die Ergebnisse genauer deuten zu können, wurden anhand von Abbildung 14 die Ergebnisse der Clusteranalyse auch auf ihre Relation untersucht. Anhand der Darstellung wird ersichtlich, welche Zyklen in welcher Clusteranalyse zu welchem Cluster zugeordnet wurden. Dabei konnte festgestellt werden, dass bei unterschiedlichen Chunk-Anzahlen bestimmte Zyklen zu unterschiedlichen Clustern zugeordnet werden. Dies ist in Abbildung 14 für das Clustering mit 60 Chunks in der Abszisse und mit 120 Chunks in der Ordinate dargestellt. Für eine leichtere Erkennbarkeit wurden die Cluster der y-Achse mit unterschiedlichen Farben gekennzeichnet. Zudem wurde die Anzahl der Zyklen im jeweiligen Cluster bzw. Cluster-Teil ergänzt und Relationen durch das Einfügen von Rahmen verdeutlicht. Generell lässt sich anhand der Darstellung erkennen, dass die Menge der Zyklen, welche anders referenziert werden nur einen kleinen Teil der Cluster ausmachen. So ist Cluster 2 bei 60 Chunks größtenteils identisch zu Cluster 0 bei 120 Chunks. Gleiches gilt für Cluster 4 bei 60 Chunks und Cluster 2 bei 120 Chunks. Der Ausgleich zwischen Cluster 0 und Cluster 2 bei 120 Chunks mit den anderen beiden Clustern je nach Wahl der Chunkzahl kommt vermutlich dadurch zustande, dass diese sich in ihrem Betriebsverhalten am ähnlichsten sind.



**Abbildung 14: Cluster - Relationen bei HC I mit 120 über 60 Chunks bei einem Schwellenwert von 40 %**

In Anlage 02 können die Relationen der Clusterzugehörigkeiten der Clusterings mit 60 Chunks und mit 120 Chunks nachvollzogen werden. Die Clusterings stimmen in den in Tabelle 12 aufgeführten Ergebnissen überein. Das ist darauf zurückzuführen, dass die gefundenen Cluster sehr eindeutig sind. Bei 120 Chunks zeichnet sich der Cluster 8 durch eine niedrigere mittlere Außentemperatur von 10,85 °C, einen durchschnittlichen

WRG Betriebsanteil von 99 % und eine mittlere Heizleistung von 42,59 kW als Cluster mit dem Betriebsmodus „Heizen mit WRG“ aus. Cluster 9 des gleichen Clustering kann durch eine mittlere Außentemperatur von 28,87 °C, einen hohen WRG-Betriebsanteil von 80 % und die mittlere Heizleistung von -149,28 kW als Cluster im Betriebszustand „Kühlen mit WRG“ klassifiziert werden.

Bei dem mit 120 Chunks gebildeten Cluster 10 wird davon ausgegangen, dass es sich aufgrund der mittleren Außentemperatur von 24,7 °C, einem mittleren WRG-Betriebsanteil von 18 % und einer mittleren Heizleistung von -97,83 kW um den Betriebszustand „Kühlen ohne WRG“ handelt.

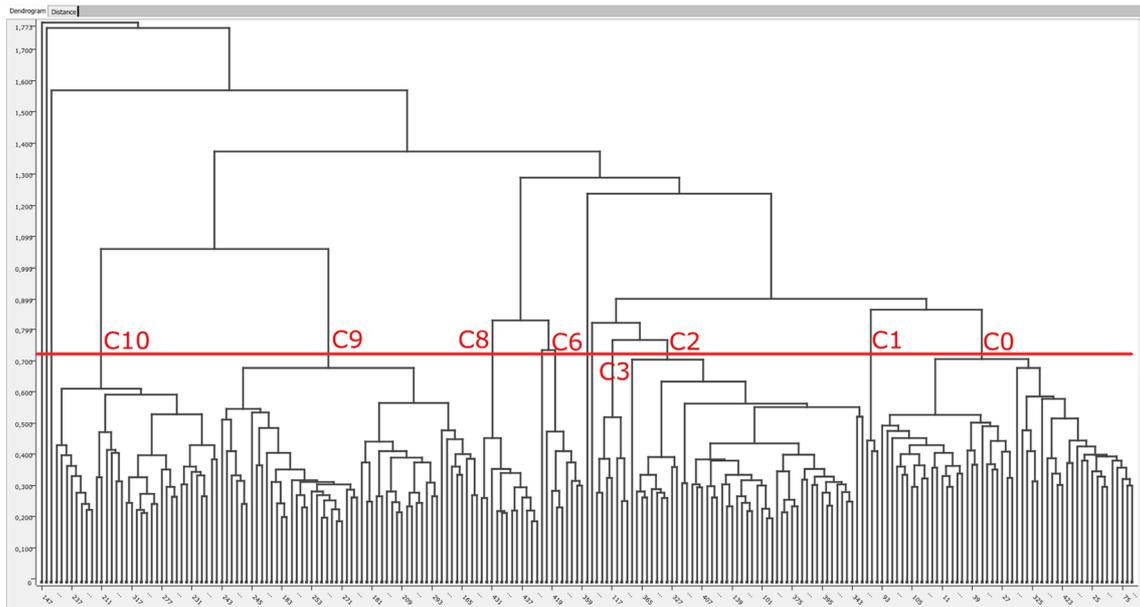
Clustering mit Ch120	Clustering mit Ch60	Clustering mit Ch20	Betriebszustand
Cluster 8	Cluster 13	Cluster 19	Heizen mit WRG
Cluster 10	Cluster 11	Cluster 15 & Cluster 16	Kühlen ohne WRG
Cluster 9	Cluster 10	Cluster 13 & Cluster 14	Kühlen mit WRG

**Tabelle 12: Übereinstimmende Relationen aus Clusteranalysen mit unterschiedlichen Chunk-Anzahlen**

Aufgrund der anhand von Abbildung 14 gewonnenen Erkenntnis, dass bei der Wahl unterschiedlicher Chunk-Anzahlen nur ein geringer Teil der Zyklen anderen Clustern zugeordnet wurde, wird vermutet, dass die Wahl der Anzahl von Chunks bei der Durchführung der Analyse eine untergeordnete Rolle spielt. Dies muss im Zuge weiterer Untersuchungen verifiziert werden, da es den Rahmen dieser wissenschaftlichen Arbeit übersteigt. In den im Folgenden vorgestellten Datenanalysen wird eine Anzahl von 120 Chunks verwendet. Aus diesem Grund wurde keine komplette Betriebszustandszuordnung anhand der Clusteranalyse mit 60 Chunks durchgeführt. Die Evaluationswerte für die Analyse können im Anhang der Arbeit unter Anlage 06 nachvollzogen werden.

#### 4.1.7 Zuordnung von Betriebszuständen zum hierarchischen Clustering I

Hier soll dargestellt werden, welches zu erwartende Anlagenverhalten den einzelnen Clustern bei HC I unter Verwendung von 120 Chunks und einem distanzbasierten Schwellenwert von 40 % zugeordnet werden kann. Dabei wird zunächst in Abbildung 15 das bei der Clusteranalyse entstandene Dendrogramm gezeigt. Es gibt Auskunft über die Nummerierung der entstandenen Cluster.



**Abbildung 15: Dendrogramm von HC I bei 120 Chunks mit Kennzeichnung des 40 % Schwellenwerts und der Nummerierung der Cluster**

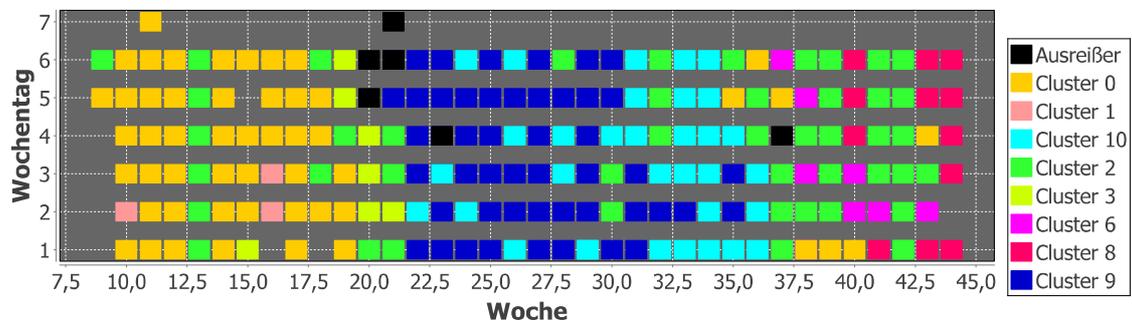
In Tabelle 13 sind die ausschlaggebenden Cluster mit ihrem zugeordneten Betriebszustand, der mittleren Außentemperatur, der mittleren Heizleistung und des durchschnittlichen WRG-Betriebsanteil aufgelistet. Dabei handelt es sich um einen Ausschnitt der in Anlage 04 dargestellten Evaluationstabelle welcher um den zugeordneten Betriebszustand ergänzt wurde.

Cluster Nummer	Betriebszustand	Mittlere Außentemperatur [°C]	Mittlere Heizleistung [kW]	Durchschnittlicher WRG-Betriebsanteil [%]
<b>Cluster0</b>	Heizen mit WRG	12,39	39,35	96
<b>Cluster1</b>	Heizen mit WRG	4,42	118,42	99
<b>Cluster2</b>	Cluster im Übergangsbereich	18,74	-3,14	39
<b>Cluster3</b>	Kühlen mit WRG	26,51	-47,49	73
<b>Cluster6</b>	Heizen mit WRG	14,47	21,17	96
<b>Cluster8</b>	Heizen mit WRG	10,85	42,59	99
<b>Cluster9</b>	Kühlen mit WRG	28,87	-149,28	80
<b>Cluster10</b>	Kühlen ohne WRG	24,7	-97,83	18

**Tabelle 13: HC I - Darstellung gefundener Cluster mit ausschlaggebenden Evaluationswerten und Zuordnung des zu erwartenden Anlagenverhaltens.**

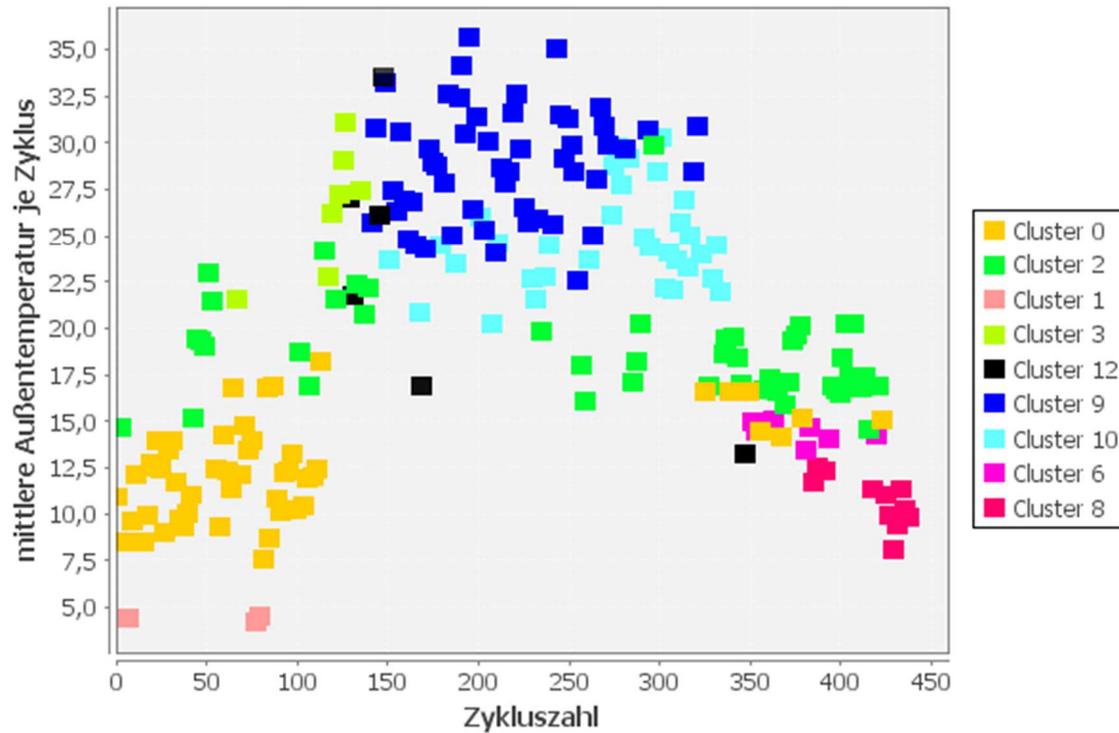
#### 4.1.8 HC I - Darstellung der Ergebnisse anhand ausschlaggebender Daten

In Abbildung 16 ist die Verteilung der gefundenen Cluster über den Jahresverlauf dargestellt. Auf der Abszisse sind die Kalenderwochen des jeweiligen Jahres aufgetragen. Die Ordinate gibt den jeweiligen Wochentag wieder, wobei Montag als Tag eins und Sonntag als Tag sieben definiert ist. Die unterschiedlichen Cluster sind gemäß der Legende farbig dargestellt und Ausreißer, d.h. Zyklen, welche im Clustering in Einzelclustern gelandet sind, wurden schwarz markiert. Lücken im Plot sind auf Feiertage und somit auf Nichtbetrieb der Anlage zurückzuführen. Zudem fällt auf, dass die Anlage in Woche elf und Woche 21 am Sonntag betrieben wurde. Dabei handelte es sich um verkaufsoffene Sonntage. Man kann anhand der Abbildung und unter Verwendung der in Tabelle 13 den Clustern zugeordneten Betriebszuständen erkennen, dass sich die einzelnen Cluster gut in den Jahresverlauf einordnen, was ein Indiz für die den unterschieden im Anlagenbetriebsverhalten entsprechende Clusterbildung ist.



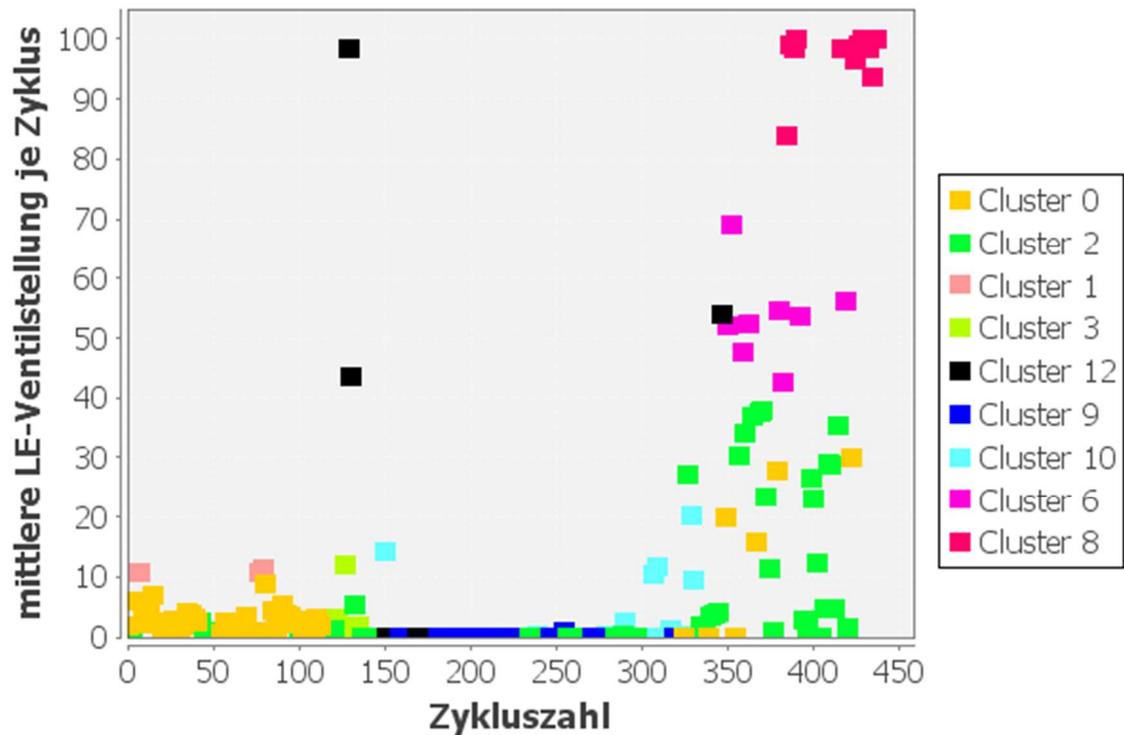
**Abbildung 16: Cluster Heatmap für HC I mit 120 Chunks und Schwellenwert von 40 %**

In den folgenden Abbildungen sollen die Ergebnisse der Clusteranalyse nochmals anhand der gewählten Datenpunkte aufgezeigt werden. Dafür wurde für den jeweiligen Datenpunkt der Mittelwert je Zyklus gebildet und dargestellt. Zudem wurden die gefundenen Cluster farblich hervorgehoben. In Abbildung 17 ist die mittlere Außentemperatur jedes Zyklus über den gesamten Untersuchungszeitraum dargestellt. Man kann erkennen, dass die Cluster 0, 1, 6 & 8 bei niedrigeren Temperaturen erkannt wurden. Auch die Ergebnisse der Evaluation der Cluster in Tabelle 13 bestätigen deren Eigenschaft als Heizcluster. Cluster 9 wurde erfolgreich als Kühlcluster bei den höchsten Temperaturen erkannt. Cluster 12 weist eine geringere mittlere Temperatur und somit einen geringeren Kühlbedarf auf. Bei den als Cluster 12 zusammengefassten Zyklen handelt es sich um die Menge der beim Analyseprozess entstandenen Ausreißer, welche dementsprechend schwarz dargestellt wurden.



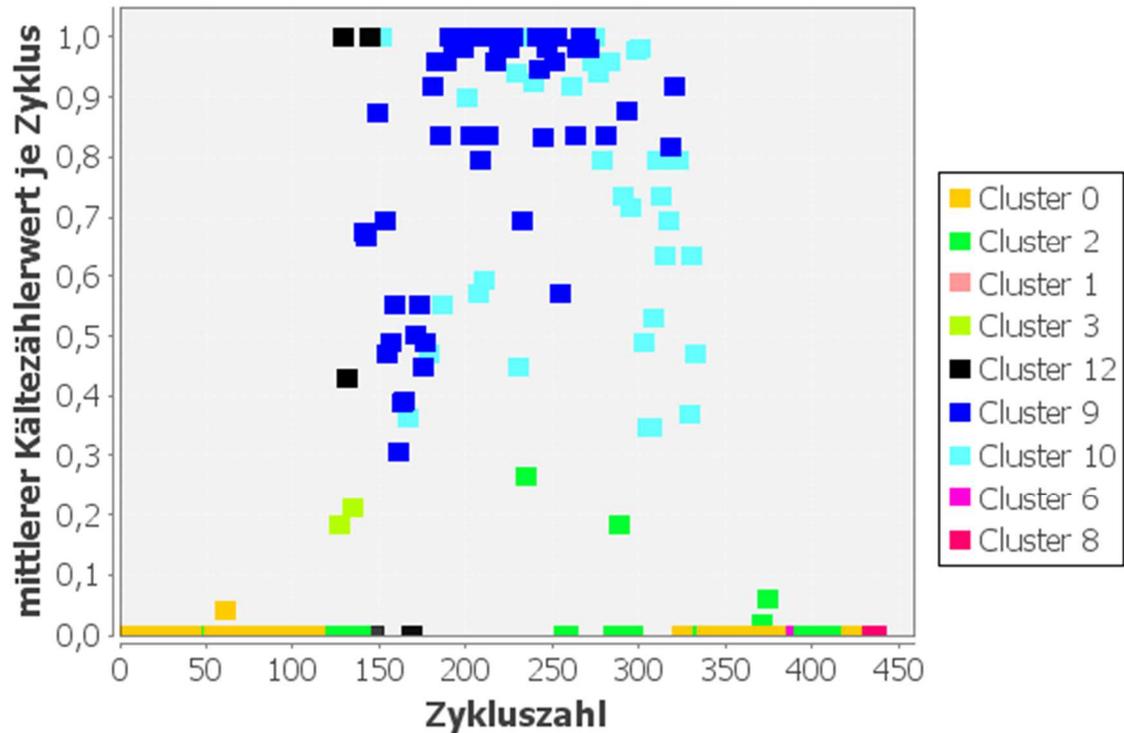
**Abbildung 17: HC I Clustering Ergebnis anhand der mittleren Außentemperatur je Zyklus**

In Abbildung 18 kann man erkennen, dass die Cluster des Heizbetriebes, also Cluster 6 und Cluster 8 sich vor allem durch die mittlere Ventilstellung des Luftherhitzers von den anderen Clustern unterscheiden. Cluster 0 und Cluster 1 weisen eine deutlich geringere Öffnung des LE-Ventils auf, können aber dennoch als Heizcluster verstanden werden. Zudem wird erkennbar, dass die mittlere LE-Ventilstellung zwischen Zyklus 340 und 450 vergleichsweise höhere Werte annimmt als in dem Bereich von Zyklus 0 bis Zyklus 150. Dies konnte durch die Clusterbildung in den meisten Fällen unterschieden werden.



**Abbildung 18: HC I Clustering Ergebnis anhand der mittleren LE-Ventilstellung je Zyklus**

Anhand von Abbildung 19 kann man nachvollziehen, dass es sich bei den Clustern 9 und 10 um die hauptsächlichen Kühlcluster handelt. Bei diesen steigt der Kältezählerwert deutlich höher, als bei den anderen Clustern. Bei Cluster 2 und Cluster 3 lässt sich anhand von Abbildung 18 und Abbildung 19 erkennen, dass beide Cluster sowohl Zyklen mit geöffnetem LE-Ventil, als auch Zyklen mit Betrieb des Luftkühlers enthalten. Daraus wird geschlossen, dass es sich bei den enthaltenen Betriebszyklen um Übergangcluster handelt.



**Abbildung 19: HC I Clustering Ergebnis anhand der mittleren Betriebsstufe des Luftkühlers je Zyklus**

Anhand von Abbildung 20 kann man erkennen, dass die Kühlcluster 9 und 12 anhand ihrer WRG-Ventilstellung eindeutig unterschieden werden können. Da Cluster 9 höhere Temperaturwerte aufweist, kann davon ausgegangen werden, dass es sich um ein authentisches Ergebnis handelt. Zudem fällt auf, dass sich Cluster 2 durch geringe WRG-Nutzung auszeichnet. Dieser Cluster beinhaltet vor allem Zyklen, die mittlere Temperaturen zwischen 16 °C und 21 °C aufweisen.

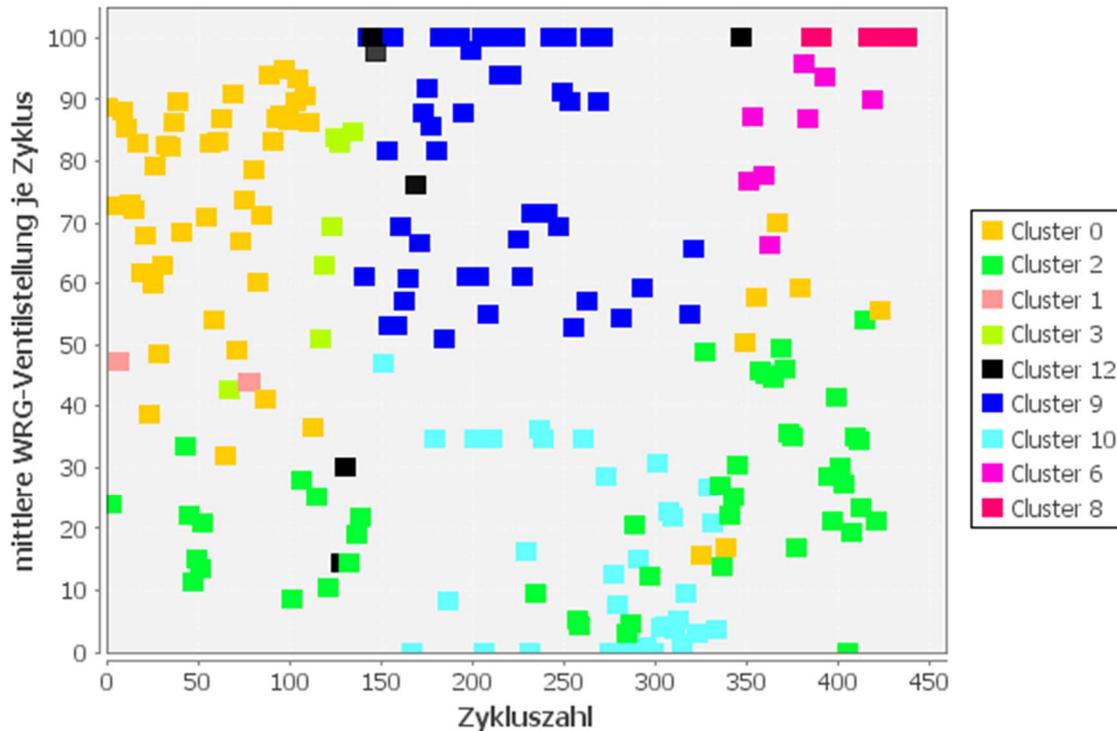


Abbildung 20: HC I Clustering Ergebnis anhand der mittleren WRG-Ventilstellung je Zyklus

#### 4.1.9 Einführung der hierarchischen Clusteranalyse II

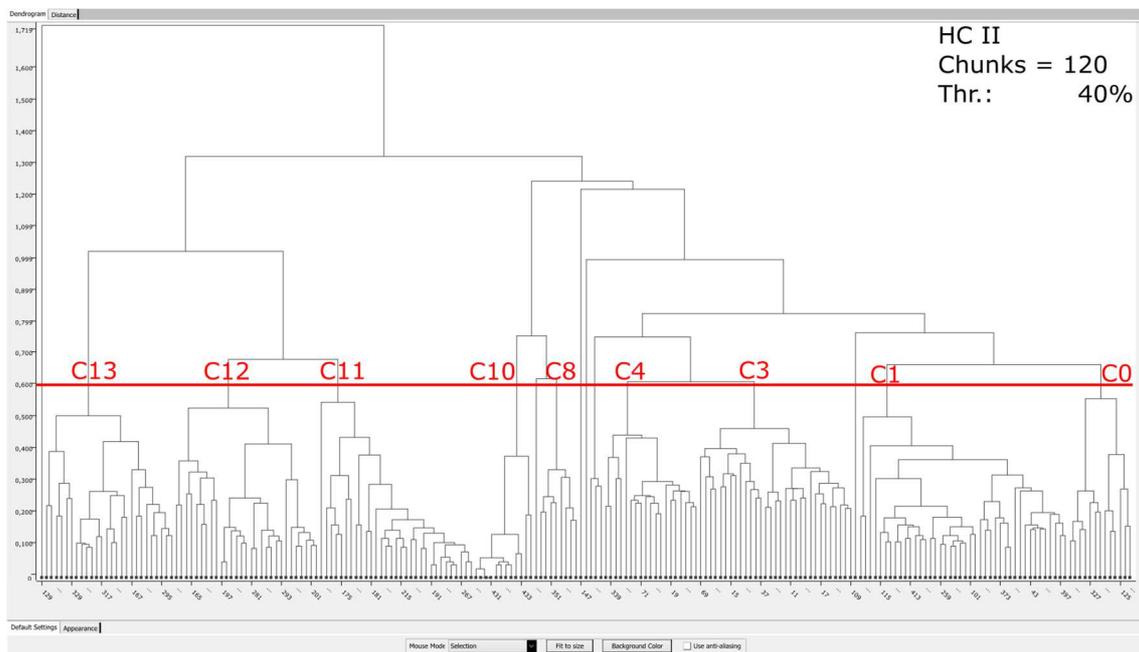
Im folgenden Abschnitt soll aufgezeigt werden, inwiefern sich das Ergebnis des Clusterings durch die Verwendung unterschiedlicher Datenpunkte ändert. Schon bei der ersten Datenpunktauswahl wurde darauf geachtet, dass die Sensoren die entscheidenden Komponenten der Lüftungsanlage bei der Datenpunktauswahl berücksichtigt werden. Im Rahmen dieser Untersuchung wurden die in Tabelle 14 dargestellten Datenpunkte verwendet. Nach wie vor wird das Anlagenverhalten adäquat wiedergespiegelt, jedoch unter Vernachlässigung der Außentemperatur. Dadurch wird versucht die Aussagekraft der Außentemperatur im Clustering aufzuzeigen.

Um entsprechend referenzieren zu können, wurden die anderen Parameter entsprechend beibehalten. Es wurde erneut eine Chunkzahl von 120 und ein Schwellenwert von 40 % gewählt, um somit die Ergebnisse der Clusteranalyse mit den Ergebnissen von HC I vergleichen zu können. Die in diesem Abschnitt eingeführte Datenpunktauswahl für die erneute Clusteranalyse wird als hierarchisches Clustering II (HC II) bezeichnet.

Datenpunkte
RLT06 LE-Ventilstellung
RLT06 WRG-Ventilstellung
RLT06 Betriebsstufe Luftkühler gesamt

Tabelle 14: RLT06 – HC II – Datenpunktauswahl

Das Ergebnis des Clustering ist in Form eines Dendrogramm in Abbildung 21 dargestellt. Der Schwellenwert wurde durch eine horizontale rote Linie ergänzt. Die Nummerierung der Cluster kann ebenfalls nachvollzogen werden.



**Abbildung 21 : Dendrogramm Hierarchical Clustering II mit 120 Chunks und 40 % Schwellenwert**

In Tabelle 15 wurden die größten bei HC II entstandenen Cluster mit dem zugeordneten Betriebszustand, der mittleren Außentemperatur, der mittleren Heizleistung und dem durchschnittlichen WRG-Betriebsanteil im Cluster aufgeführt. Dabei handelt es sich um einen Ausschnitt der in Anlage 07 dargestellten Evaluationstabelle, der um den zugeordneten Betriebszustand erweitert wurde. Dabei konnte Cluster 3, Cluster 4, Cluster 8 und Cluster 10 der Betriebszustand „Heizen mit WRG“ zugesprochen werden. Dies kann mit der positiven mittleren Heizleistung und dem hohen prozentualen WRG-Nutzungsanteil begründet werden. Cluster 0 und Cluster 12 wurde der Betriebszustand Kühlen mit WRG zugeordnet, was man anhand der negativen mittleren Heizleistung und einem hohen WRG-Betriebsanteil festmachen kann. Cluster 1 weist mit  $-5,8$  kW nur eine geringe Abweichung von Null auf. Da es sich bei dem Wert um den Mittelwert für die im Cluster enthaltenen Betriebszyklen handelt, wird angenommen, dass es in diesen Zyklen sowohl zu einem Heizbetrieb als auch zu einem Kühlbetrieb kommt. Cluster 13 wurde der Betriebszustand Kühlen ohne WRG zugeordnet, da der WRG-Betriebsanteil bei 14 % liegt und die mittlere Heizleistung bei  $-92,61$  kW deutlich negativ ist.

Cluster Nummer	Betriebszustand	mittlere Außentemperatur [°C]	mittlere Heizleistung [kW]	Durchschnittlicher WRG-Betriebsanteil
Cluster0	Kühlen mit WRG	21,47	-19,38	69%
Cluster1	Cluster im Übergangsbereich	19,18	-5,8	35%
Cluster3	Heizen mit WRG	11,09	52,14	99%
Cluster4	Heizen mit WRG	14,19	21,37	94%
Cluster8	Heizen mit WRG	14,47	21,17	96%
Cluster10	Heizen mit WRG	10,85	42,59	99%
Cluster11	Kühlen mit WRG	30,53	-167,35	94%
Cluster12	Kühlen mit WRG	26,15	-120,49	57%
Cluster13	Kühlen ohne WRG	24,53	-92,61	14%

Tabelle 15: HC II - Darstellung gefundener Cluster mit ausschlaggebenden Evaluationswerten und Zuordnung des zu erwartenden Anlagenverhaltens

#### 4.1.10 HC II - Darstellung der Ergebnisse anhand ausschlaggebender Daten

In **Abbildung 22** kann die Zuordnung der Cluster für jeden Betriebszyklus des Jahres abgelesen werden. Auf der y-Achse wird der jeweilige Wochentag vermerkt, wobei Montag als erster Tag der Woche festgelegt wurde. Man kann anhand der Abbildung und unter Verwendung der in Tabelle 15 den Clustern zugeordneten Betriebszuständen erkennen, dass sich die einzelnen Cluster gut in den Jahresverlauf einordnen, was ein Indiz für die korrekte Zuordnung der einzelnen Betriebszyklen zu den jeweiligen Clustern ist.

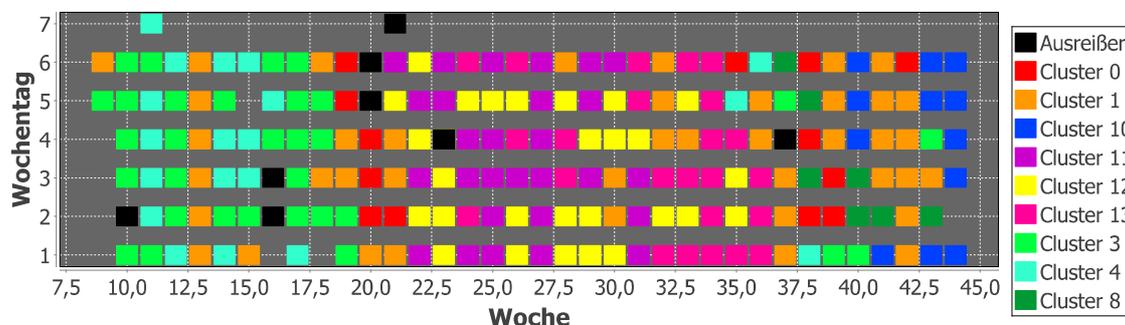
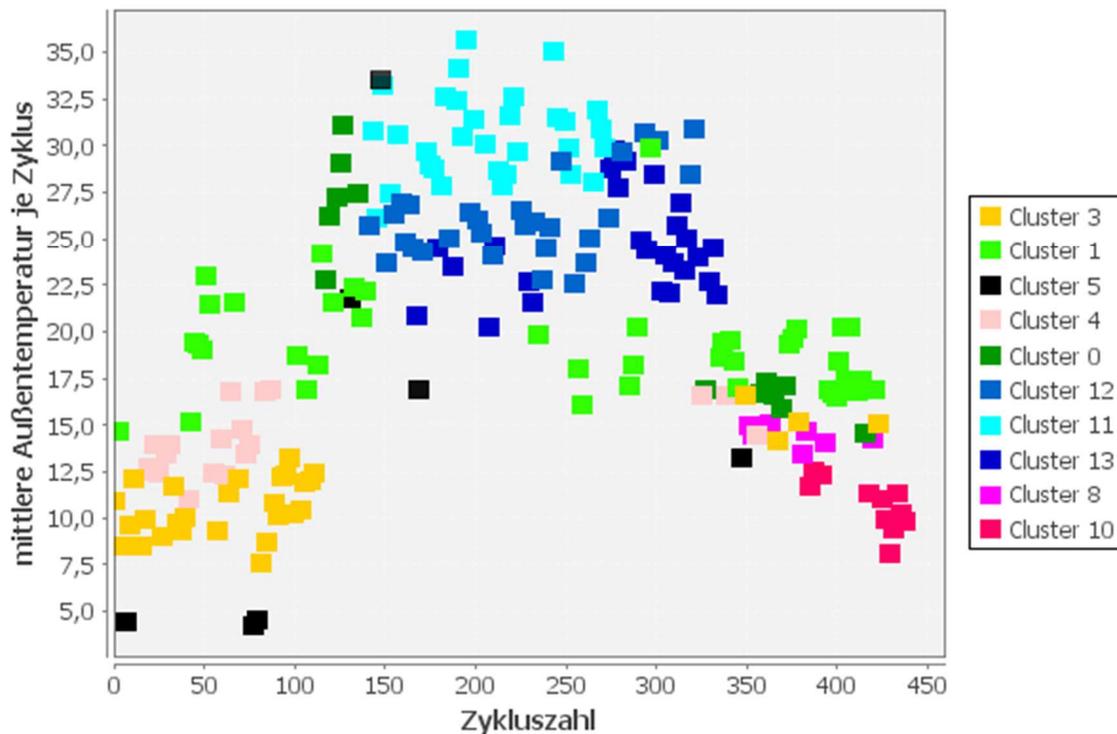


Abbildung 22: Cluster Heatmap HC II - CH 120 - Thr.: 40%

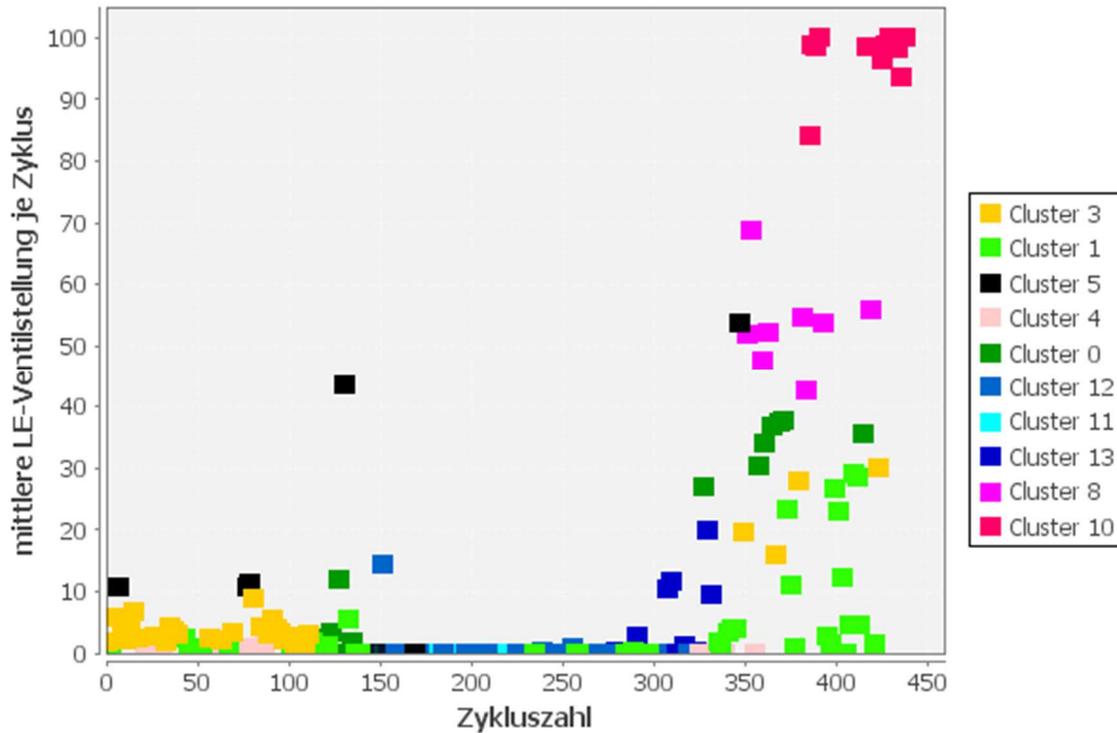
In **Abbildung 23** ist die mittlere Außentemperatur je Zyklus für die Clusteranalyse HC II dargestellt. Man erkennt, dass Cluster 0, Cluster 11, Cluster 12 und Cluster 13 bei den

höchsten mittleren Temperaturen gebildet wurden. Cluster 1 ist vor allem in dem Bereich von 17,5 °C bis 22,5 °C aufzufinden. Cluster 8 und Cluster 10 wurden bei niedrigen Temperaturwerten im Bereich zwischen 350 und 450 Zyklen angesiedelt. Dagegen finden Cluster 3 und Cluster 4 hauptsächlich innerhalb der ersten 125 Zyklen statt. Bezogen auf die mittlere Außentemperatur liegen sie dabei hauptsächlich in einem Bereich zwischen 7,5 °C und 12,5 °C bzw. 12,5 °C und 15 °C.



**Abbildung 23: RLT06 - HC II - mittlere Außentemperatur je Zyklus**

In Abbildung 24 ist die mittlere prozentuale LE-Ventilstellung gegen die Zahl der Zyklen dargestellt. Zunächst stellt man fest, dass die LE-Ventilstellung von Zyklus 150 bis Zyklus 300 nahezu ständig gleich Null ist. Da sich in diesem Zeitraum vor allem Cluster 11, Cluster 12 und Cluster 13 befinden, welchen der Betriebszustand "Kühlen" zugeordnet wurde, kann dies die durch die Evaluationstabelle getroffene Annahme bestätigen. Außerdem ist zu erkennen, dass sich zwischen Zyklus 0 und Zyklus 125 ein anderer Anlagenbetrieb abzeichnet, als zwischen Zyklus 300 und 450. Man kann erkennen, dass das Luftherzerventil im Mittel weiter geöffnet wurde. Dadurch kann die Unterscheidung von Cluster 3 und Cluster 4 zu Cluster 8 und Cluster 9 erklärt werden.



**Abbildung 24: RLT06 - HC II - mittlere LE-Ventilstellung je Zyklus**

In Abbildung 25 ist die mittlere Betriebsstufe des Luftkühlers je Zyklus dargestellt. Zudem wurden die unterschiedlichen Cluster gemäß der Legende farblich hervorgehoben. Man kann erkennen, dass Cluster 11, Cluster 12 und Cluster 13 entsprechend ihres erhöhten Wertes für die Betriebsstufe des Luftkühlers gebildet werden konnten. Allerdings wird ersichtlich, dass die genannten Cluster nicht nur anhand dieses Datenpunktes gebildet wurden. Anhand von Abbildung 26 kann gezeigt werden, dass Cluster 11, Cluster 12 und Cluster 13 hinsichtlich des Anteils ihrer WRG-Nutzung unterschieden werden konnten.

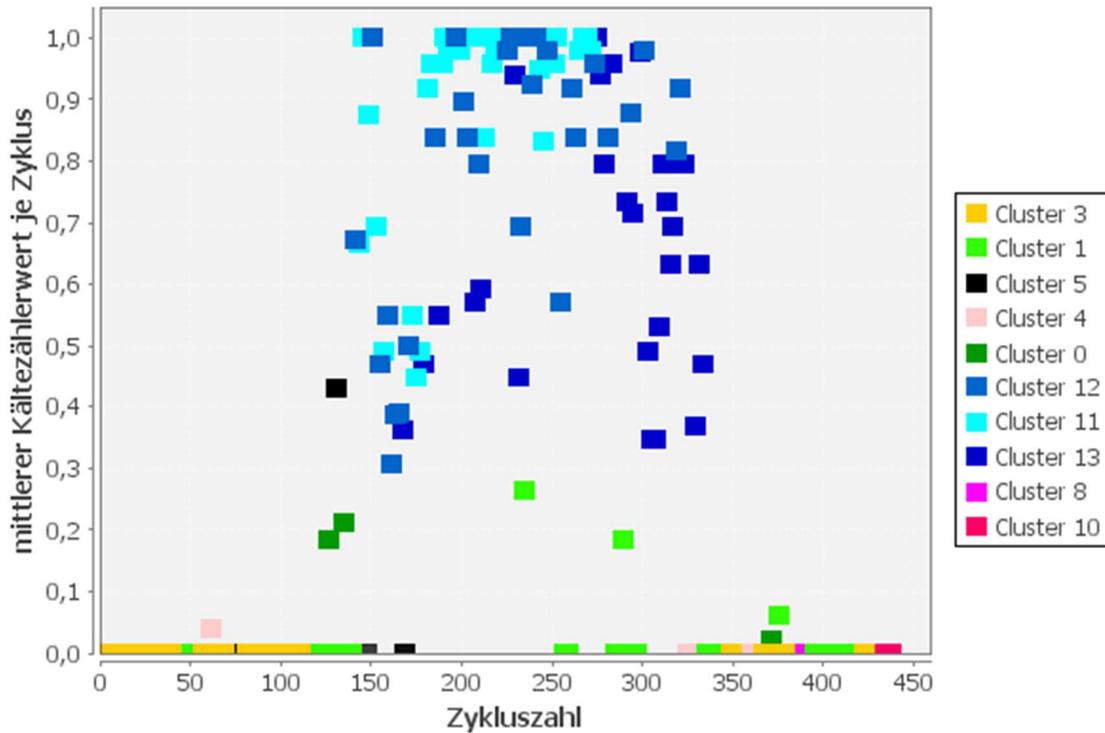


Abbildung 25: RLT 06 - HC II - mittlere Betriebsstufe Luftkühler je Zyklus

In Abbildung 26 lässt sich außerdem erkennen, dass Cluster 3 und Cluster 4 anhand ihrer anteiligen WRG-Nutzung separiert werden konnten. Gleiches gilt für Cluster 8 und Cluster 10.

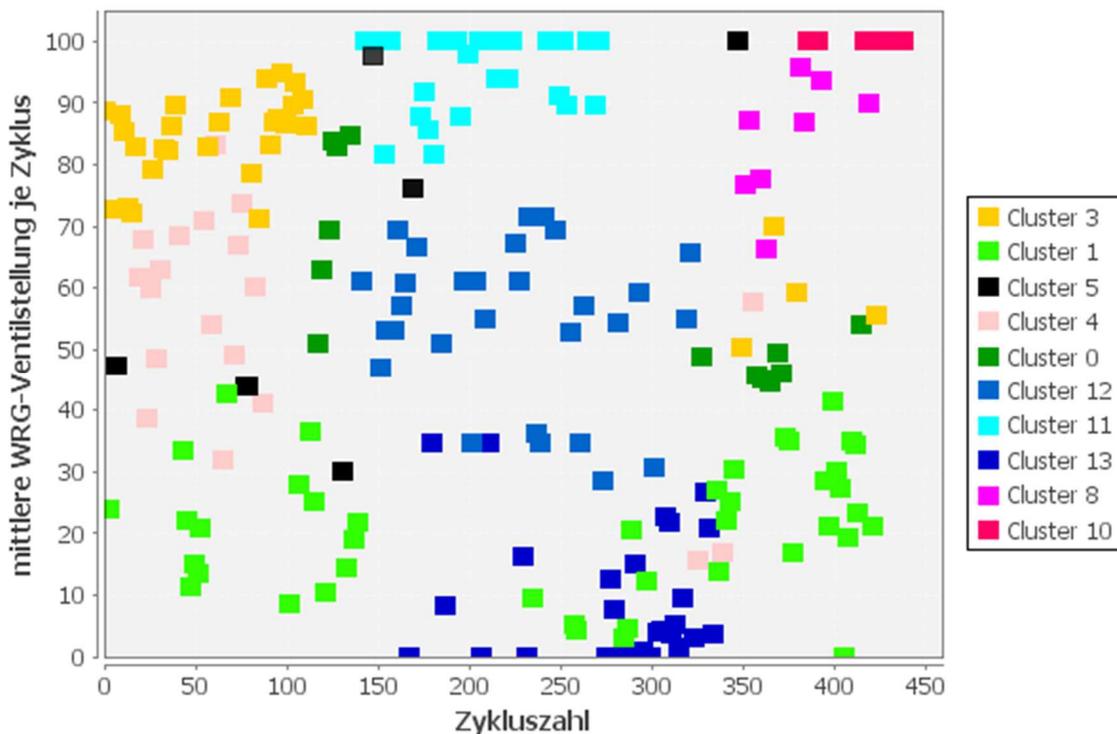


Abbildung 26: RLT06 - HC II - mittlere WRG-Ventilstellung je Zyklus

### 4.1.11 Darstellung der Auswirkung der Außentemperatur auf das Ergebnis der Clusteranalyse

Im Folgenden soll die Auswirkung der Außentemperatur als in der Clusteranalyse verwendeter Datenpunkt auf das Ergebnis der Clusteranalyse dargestellt werden. Dafür werden die in Kapitel 4.1.8 und Kapitel 4.1.10 erläuterten Analyseergebnisse miteinander verglichen. In Abbildung 27 sind die Cluster-Relationen der beiden Analysen dargestellt. Anhand der Darstellung kann nachvollzogen werden, inwiefern sich die Cluster von HC I und die Cluster von HC II entsprechen.

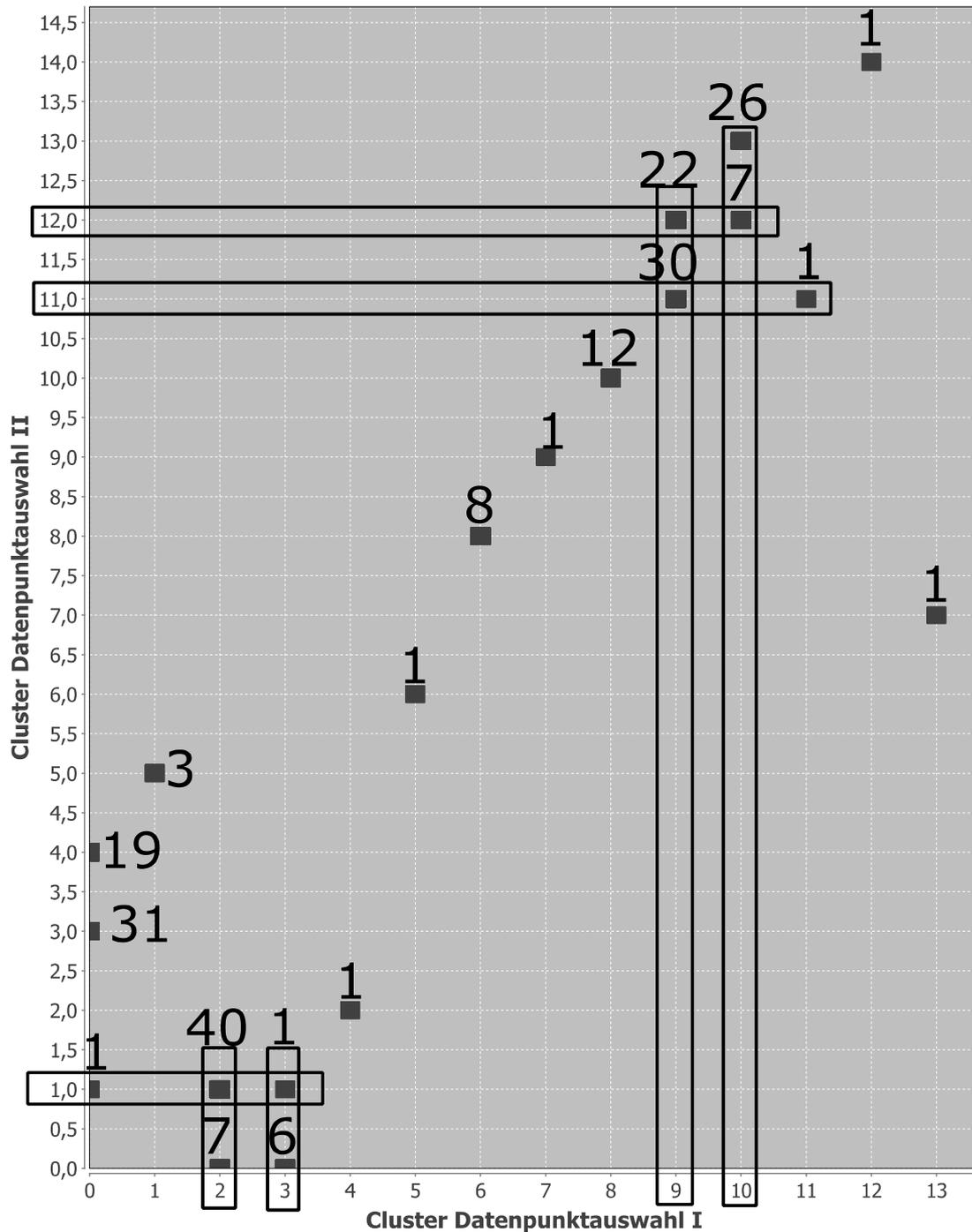


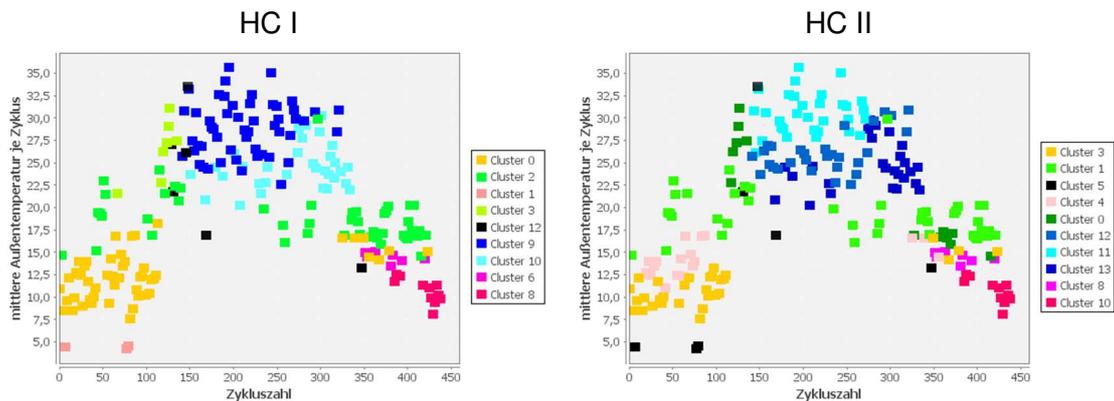
Abbildung 27: Darstellung der Cluster-Relation von HC I zu HC II

Eine Liste identischer Cluster ist in Tabelle 16 ersichtlich. Dabei wurden keine Ausreißer-Cluster, also Cluster mit drei oder weniger Zyklen aufgeführt. Zudem wird ersichtlich, dass Cluster 11 und Cluster 12 das Kühlverhalten der Anlage widerspiegeln und somit bis auf einen Zyklus, welcher allein den Cluster 13 bildet, identisch zur Menge der Cluster 9 und 10 von HC I sind. Zudem wurden mit Cluster 3 und Cluster 4 zwei weitere Cluster entdeckt, welche als Heizcluster gedeutet werden können.

HC I	HC II	Betriebszustand
Cluster 8	Cluster 10	Heizen mit WRG
Cluster 6	Cluster 8	Kühlen ohne WRG

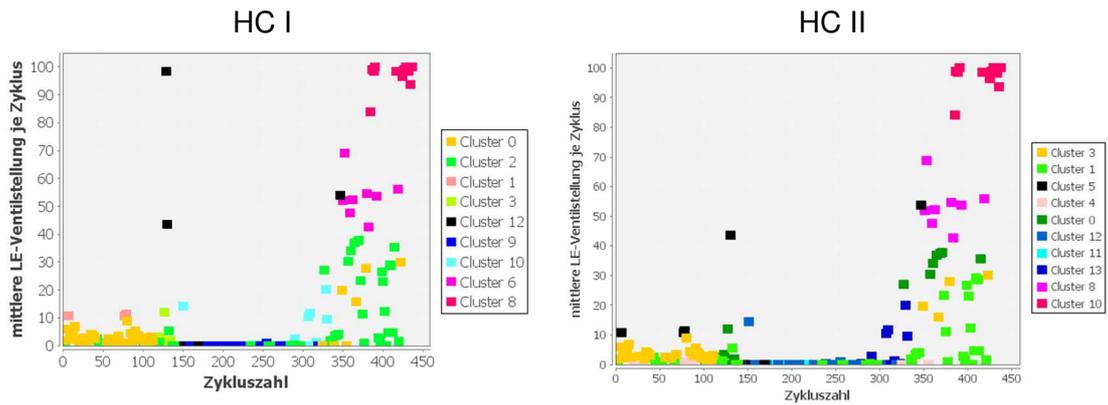
**Tabelle 16: Identische Cluster aus Vergleich HC I und HC II**

In den folgenden Abbildungen wurden für jeden Zyklus ein Mittelwert des jeweilig in der y-Achse dargestellten Datenpunktes bestimmt. In Abbildung 28 wird ersichtlich, dass die beiden Clusteranalysen sich anhand der mittleren Außentemperatur je Zyklus kaum unterscheiden lassen. Farbliche Unterschiede zwischen den beiden Darstellungen lassen sich dadurch erklären, dass beispielsweise Cluster 0 von HC I bei HC II in Cluster 3 und Cluster 4 aufgeteilt wurde. Unter Berücksichtigung der in Tabelle 13 und Tabelle 15 ausgewiesenen Anlagenbetriebszustände der Cluster der jeweiligen Clusteranalyse kann dies genauer nachvollzogen werden.



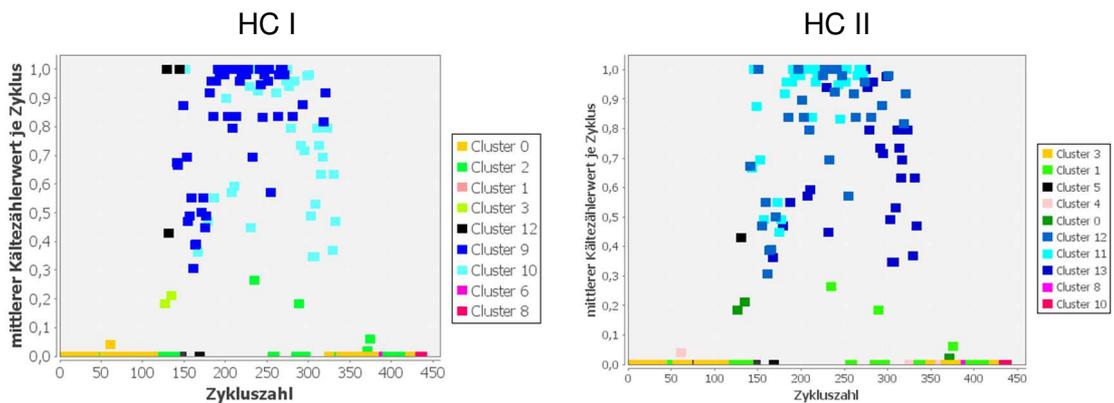
**Abbildung 28: Vergleich der Ergebnisse der Clusteranalysen HC I und HC II anhand der mittleren Außentemperatur je Zyklus**

Anhand von Abbildung 29 ist der Vergleich der mittleren LE-Ventilstellungen zwischen den Clusteranalysen HC I und HC II dargestellt. Man kann erkennen, dass Cluster 6 von HC I und Cluster 8 von HC II übereinstimmen. Gleiches gilt für Cluster 8 von HC I und Cluster 10 von HC II. Kleinere Unterschiede entstehen bei den bei HC II Cluster 0 zugeordneten Zyklen zwischen Zykluszahl 310 und 375. Hier wird ersichtlich, dass die erwähnten Zyklen unter HC I zugeordnet worden sind.



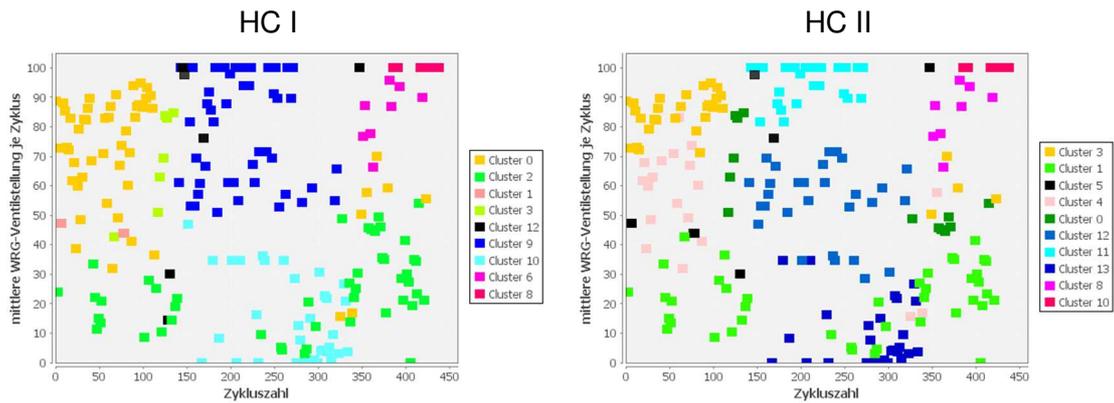
**Abbildung 29: Vergleich der Ergebnisse der Clusteranalysen HC I und HC II anhand der mittleren LE-Ventilstellung je Zyklus**

In Abbildung 30 ist der Vergleich der mittleren Betriebsstufe des Luftkühlers zwischen den Clusteranalysen HC I und HC II dargestellt. Bei beiden Clusteranalysen kann nachvollzogen werden, dass der Algorithmus die in verschiedenen Blautönen dargestellten Cluster als Cluster nach ihrem Kühlbetrieb korrekt eingeordnet hat. Ab einem Wert von 0,3 werden bei beiden Clusteranalysen sämtliche auftretende Zyklen in einen der Cluster zugeteilt, die unter Verwendung der Evaluationstabelle als Cluster im Kühlbetrieb erkannt wurden.



**Abbildung 30: Vergleich der Ergebnisse der Clusteranalysen HC I und HC II anhand der mittleren Betriebsstufe des Luftkühlers je Zyklus**

Anhand von Abbildung 31 kann das Ergebnis der beiden Clusteranalysen bezüglich der mittleren WRG-Ventilstellung jedes Betriebszyklus verglichen werden. Dabei wird ersichtlich, dass in beiden Fällen die mit unterschiedlichen Blautönen markierten Cluster des Kühlbetriebs hinsichtlich ihrer mittleren WRG-Stellung unterschieden wurden.



**Abbildung 31: Vergleich der Ergebnisse der Clusteranalysen HC I und HC II anhand der mittleren WRG-Stellung je Zyklus**

Durch den in diesem Kapitel angestellten Vergleich von HC I und HC II konnte erkannt werden, dass die Verwendung der Außentemperatur im Rahmen der Datenpunktauswahl für das hierarchische Clustering nur eine untergeordnete Rolle spielt.

## 4.2 Analyse und Auswertung von Lüftungsanlage 03

Im folgenden Kapitel soll Lüftungsanlage 03 mit der bei Lüftungsanlage 06 bewährten Methode und identischen Parametern untersucht werden. Grund dafür ist, dass bei RLT03 im Vorfeld ein eindeutiges Fehlverhalten nachgewiesen wurde, da die Anlage oftmals in den frühen Morgenstunden betrieben wird. Die Methodik soll somit auf ihre universelle Einsetzbarkeit untersucht werden.

### 4.2.1 Unterscheidung von AN- und AUS-Zyklen

Zunächst wird eine Liste an Datenpunkten ausgewählt und mittels der k-Means-Clusteranalyse in AN- & AUS- Zyklen unterschieden. Dabei wurden verschiedene Normierungsmethoden getestet, welche zu unterschiedlichen Fehlerquoten im Clustering geführt haben. Die Ergebnisse wurden erneut anhand eines selbstständigen k-Means-Kontrollwertes verifiziert, welcher anhand der Datenpunkte RLT03 Frequenzumrichter Zuluftventilator & RLT03 Frequenzumrichter Abluftventilator erstellt wurde. Bei den beiden Datenpunkten handelt es sich um die binären Betriebssensoren der Frequenzumrichter der Ventilatoren. Unabhängig von der eigentlichen Frequenz kann angenommen werden, dass die Ventilatoren bei einem Wert größer Null in Betrieb sind, weswegen sie sich für die Bildung eines Kontrollwertes eignen. Für das Clustering wurde die folgende Auswahl an Datenpunkten getroffen. Es wurde erneut darauf geachtet, mit einer Datenpunktauswahl zu arbeiten, die kein komplexes Expertenwissen voraussetzt. Es wurden keine binären AN/AUS-Sensoren verwendet, um zu zeigen, dass auch ohne diese ein gutes Ergebnis erzielt werden kann. Zudem wurden bei der Durchführung der Clusteranalyse ebenfalls die bereits als Fehlverhalten identifizierten Anlagenbetriebszyklen miteinbezogen. Liste der verwendeten Datenpunkte:

- RLT03 Ablufttemperatur
- RLT03 Abluftvolumenstrom
- RLT03 Außentemperatur
- RLT03 LE-Rücklauftemperatur
- RLT03 LE-Ventilstellung
- RLT03 WRG-Temperatur
- RLT03 WRG-Ventilstellung
- RLT03 Zulufttemperatur
- RLT03 Zuluftvolumenstrom
- RLT03 Betriebsstufe Abluftventilator
- RLT03 Betriebsstufe LE-Pumpe
- RLT03 Betriebsstufe WRG-Pumpe
- RLT03 Betriebsstufe Zuluftventilator
- RLT03 Betriebsstufe Kältemaschine Stufe 1
- RLT03 Betriebsstufe Kältemaschine Stufe 2
- RLT03 Betriebsstufe Kältemaschine Stufe 3
- RLT03 Betriebsstufe Kältemaschine Stufe 4
- RLT03 Betriebsstufe Luftkühler gesamt
- RLT03 Tagesmittelwert Wärmemengenzähler statische Heizung
- RLT03 Tagesmittelwert Wärmemengenzähler Lüftungsanlage

Die Resultate des k-Means-Clustering zur Unterscheidung von AN- und AUS-Zyklen können in Tabelle 17 nachvollzogen werden. Die angegebene Fehlerquote ist als prozentualer Anteil der Sensorwerte über den gesamten Betrachtungszeitraum angegeben. Wie schon bei der im Vorherigen untersuchten Anlage hat sich gezeigt, dass die MIN/MAX-Normierungsmethode zu den besten Resultaten führt.

Im Gegensatz zu Lüftungsanlage 06 hat sich gezeigt, dass die anderen Normierungsmethoden zu deutlich höheren Fehlern führen können. Bei der Anwendung der Normierungsmethode „Z-Score“ wurden 33,2 % des Datensatzes irrtümlich als AUS-Zyklen geclustert und 5,09 % irrtümlich als AN-Zyklen. Bei der Verwendung des „Decimal Scaling“ wurden 22,74 % des Datensatzes irrtümlich als AN-Zyklen geclustert und 3,51 % irrtümlich als AUS-Zyklen.

Normierungsverfahren	Fehlerquote	AN-Zustand	AUS-Zustand
Z-Score	38,28 %	43,36 %	18,36 %
MINMAX [-1, 1]	0,006 %	48,44 %	51,55 %
Decimal Scaling	26,25 %	25,71 %	48,04 %

**Tabelle 17: RLT03 k-Means-Clustering Ergebnisse mit verschiedenen Normierungsmethoden**

Es wird also angenommen, dass die Wahl der entsprechenden Clustering-Methode auch bei einer Lüftungsanlage mit fehlerhaftem Betrieb zu einem erfolgreichen Ergebnis führt, wenn die Daten im Vorfeld korrekt normiert wurden. Die exakte Auswirkung gewählter Datenpunkte für die Unterscheidung von AN- und Aus-Zyklen konnte für RLT03 im Rahmen dieser Arbeit nicht untersucht werden. Es wird vermutet, dass die Auswahl der Datenpunkte eine untergeordnete Rolle für das Clustering-Ergebnis spielt. Dies muss in weiteren Untersuchungen für die benannte Anlage getestet werden.

#### 4.2.2 Klassifizierung der Betriebszustände

Wie bereits aufgeführt, hat die RLT 03 ein Fehlverhalten mit kurzen Betriebsclustern, die in den frühen Morgenstunden laufen. Es soll mit diesem Abschnitt überprüft werden, inwiefern der Datenanalyseansatz auf eine nicht korrekt laufende Lüftungsanlage Verwendung finden kann. Dabei wurden die bei der Untersuchung von Lüftungsanlage 06 gewählten Parameter verwendet.

Die verwendeten Datenpunkte können in Tabelle 18 nachvollzogen werden. Des Weiteren wurden 120 Chunks für die Quantisierung und ein Schwellenwert von 40 % gewählt.

Datenpunkt	Erklärung
<b>RLT03 Außentemperatur</b>	Außentemperatur
<b>RLT03 LE-Ventilstellung</b>	Öffnungsgrad des LE-Ventils im Vorlauf [%]
<b>RLT03 WRG-Ventilstellung</b>	Öffnungsgrad des WRG-Ventils [%]
<b>RLT03 Betriebsstufe Luftkühler gesamt</b>	Abgeleiteter Stundenzähler für die Kältemaschine [h]

**Tabelle 18: RLT03 – HC I – Datenpunktauswahl**

Unter Verwendung des gesamten Datensatzes hat der Algorithmus mit 151 gefundenen Clustern ein nicht verwertbares Ergebnis geliefert. Die in Anlage 08 dargestellte Evaluationstabelle gibt die Möglichkeit dies nachzuvollziehen. Bei Lüftungsanlage 03 wurde bereits bei der Aufbereitung der Daten festgestellt, dass die Anlage in den Morgenstunden zwischen 2:00 Uhr und 4:00 in Betrieb ist, was als fehlerhafter Betrieb verstanden werden kann. Aus diesem Grund wurden alle Betriebszyklen mit einer Dauer von weniger als 2,5 h aussortiert und das hierarchische Clustering erneut durchgeführt. In Abbildung 32 ist das Dendrogramm als Resultat des Clustering zu erkennen. Der bei 40 % der maximalen Distanz eingeführte Schwellenwert ist in Form der waagerechten roten Linie dargestellt. Zudem wurden die Nummerierungen der größten Cluster beigefügt.

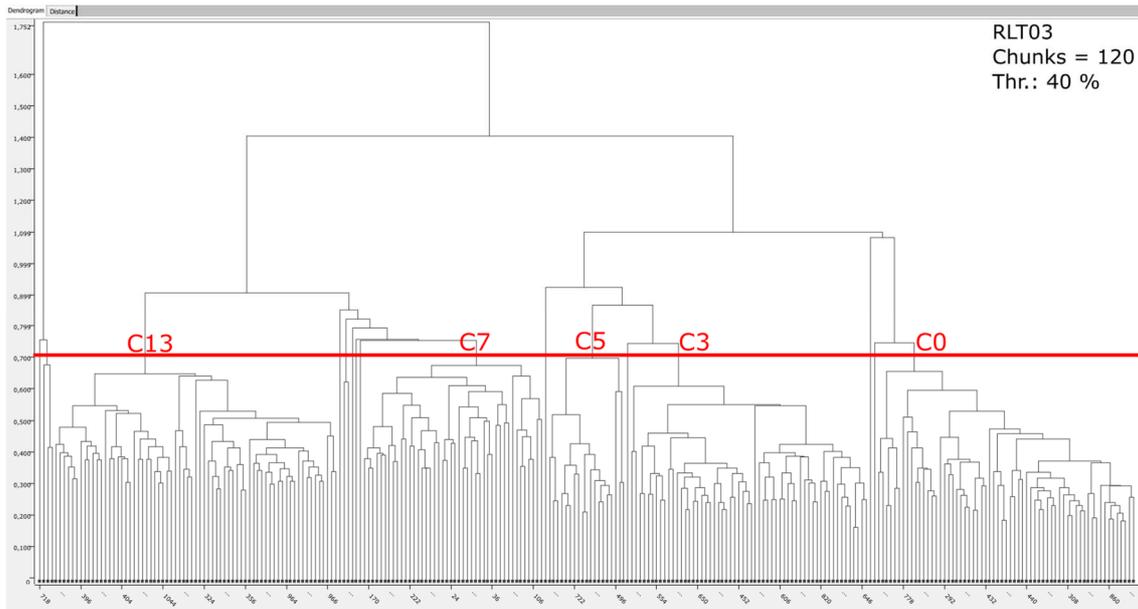


Abbildung 32: RLT03 - Dendrogramm

#### 4.2.3 Zuordnung von Betriebszuständen zum hierarchischen Clustering der Lüftungsanlage 03

Die entscheidenden Werte für die Beurteilung der Cluster können in Tabelle 19 nachvollzogen werden. Dabei handelt es sich um einen Ausschnitt aus Anlage 09, welcher um die zugeordneten Betriebszustände ergänzt wurde. Die erwähnten kurzen Betriebszyklen werden im Folgenden nicht mehr berücksichtigt werden. Aus der Tabelle wird ersichtlich, dass Cluster 3 und Cluster 5 durch ihre mittlere Außentemperatur und ihre negative mittlere Heizleistung als Cluster verstanden werden können, deren Zyklen sich im Kühlbetrieb befinden. Cluster 7 und Cluster 13 hingegen weisen durch ihre niedrigen mittleren Außentemperaturen und die hohe mittlere Heizleistung, sowie die hohen Anteile des WRG-Betriebs auf den Betriebsmodus „Heizen mit WRG“ hin.

Cluster	Betriebszustand	mittlere Außentemperatur	WRG - Betriebsanteil	mittlere Heizleistung [kW]
Cluster0	Übergangskluster	16,05	30%	20,49
Cluster3	Kühlen ohne WRG	21,03	3%	-60,82
Cluster5	Kühlen mit WRG	25,41	48%	-117,65
Cluster7	Heizen mit WRG	1,34	100%	144,02
Cluster13	Heizen mit WRG	10,21	99%	59,45

Tabelle 19: Ergebnis der Betriebszustandsklassifikation von RLTO3

#### 4.2.4 Darstellung der Ergebnisse anhand ausschlaggebender Daten

In Abbildung 33 ist die Verteilung der Cluster über das Kalenderjahr dargestellt. Dabei wird erkennbar, dass Cluster 7 vor allem in den ersten 10 Wochen des Kalenderjahres vertreten ist. Cluster 13 wurde vor allem im März, im April und im Oktober erkannt. Die beiden als Kühlbetrieb erkannten Cluster befinden sich vor allem im Zeitabschnitt zwischen Kalenderwoche 20 und 32, also hauptsächlich in den Sommermonaten. Bei Cluster null ist eine exakte Abschätzung des vorherrschenden Anlagenzustands komplex. Es wird vermutet, dass es sich um einen Cluster handelt bei dem sowohl gekühlt, als auch geheizt wird. Man kann anhand der Abbildung und unter Verwendung der in Tabelle 19 den Clustern zugeordneten Betriebszuständen erkennen, dass sich die einzelnen Cluster gut in den Jahresverlauf einordnen, was ein Indiz für die korrekte Zuordnung der einzelnen Betriebszyklen zu den jeweiligen Clustern ist.

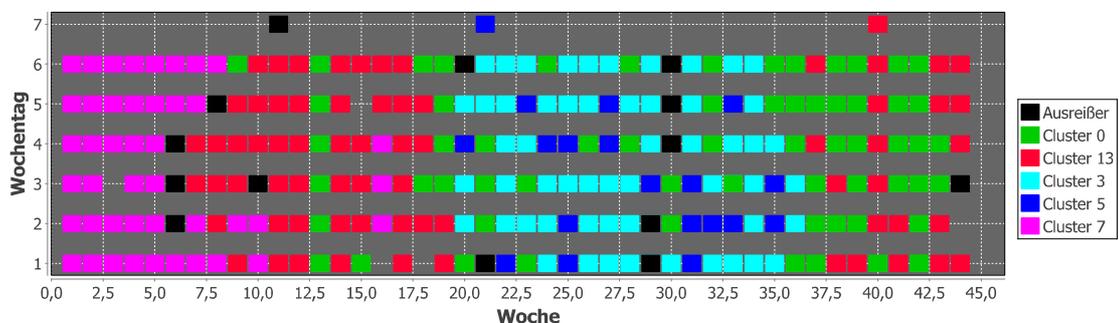
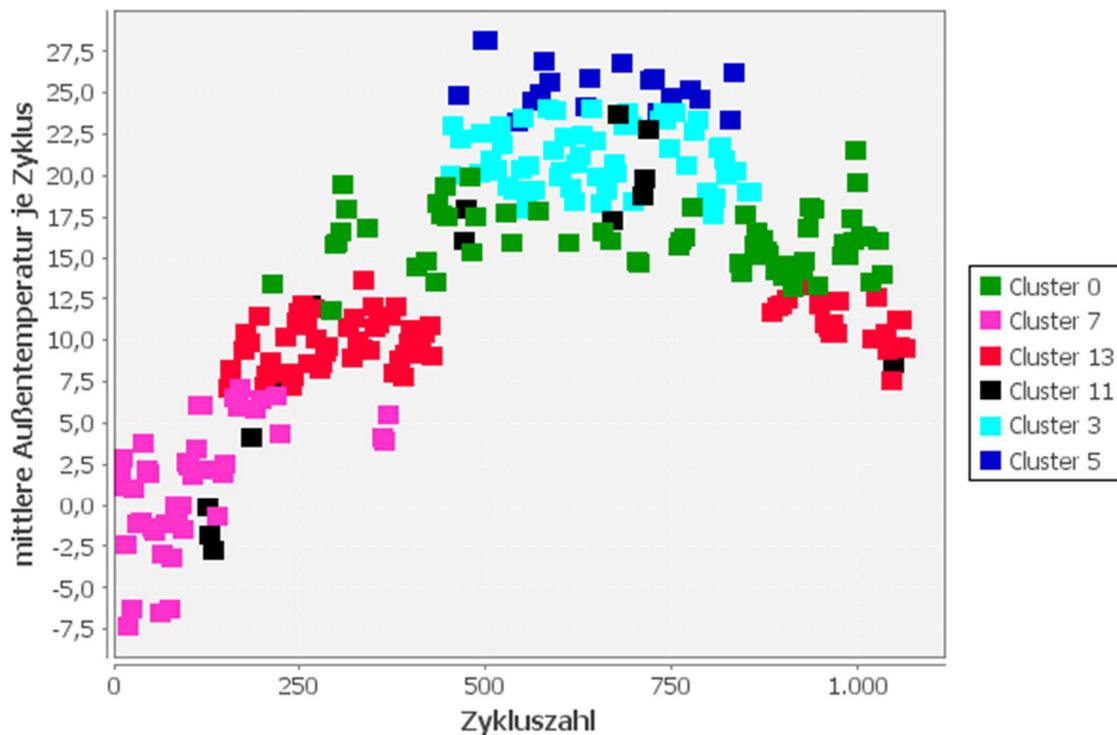


Abbildung 33: RLTO3 Cluster Heatmap über den gesamten Betrachtungszeitraum

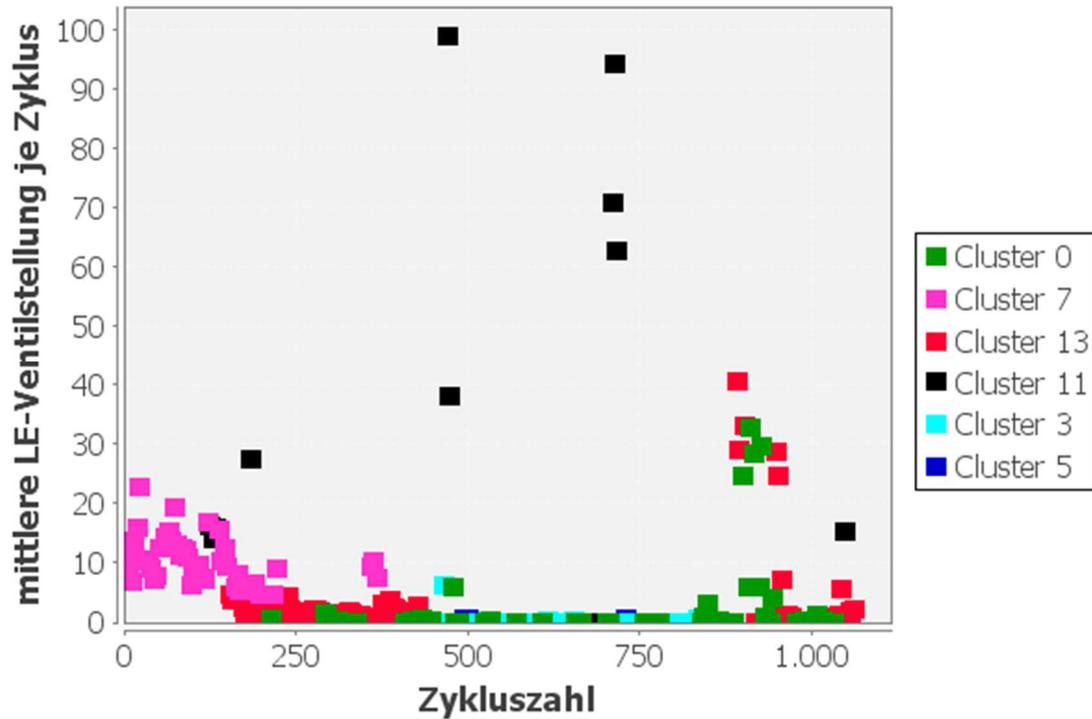
In den folgenden Abbildungen sind die gewählten Kriterien gegen die Zyklusanzahl dargestellt worden. Die gefundenen Cluster wurden farblich hervorgehoben, um das Ergebnis kenntlich zu machen. Dabei wurden alle als Ausreißer-Cluster erkannten Zyklen unter Cluster 11 zusammengefasst und geschwärzt.

In Abbildung 34 ist die mittlere Außentemperatur jedes Gebäudebetriebszyklus dargestellt. Die unterschiedlichen Cluster wurden farblich unterschieden, um das Analyseergebnis zu verdeutlichen. Zu erkennen ist, dass die Cluster abhängig von der Außentemperatur gebildet worden sind. Sie ist die Voraussetzung für den jeweiligen Anlagenbetrieb. Da die einzelnen Cluster sich augenscheinlich nach der Außentemperatur unterscheiden lassen, kann man davon ausgehen, dass die verschiedenen Betriebsmodi der Lüftungsanlage erfolgreich voneinander unterschieden werden konnten.



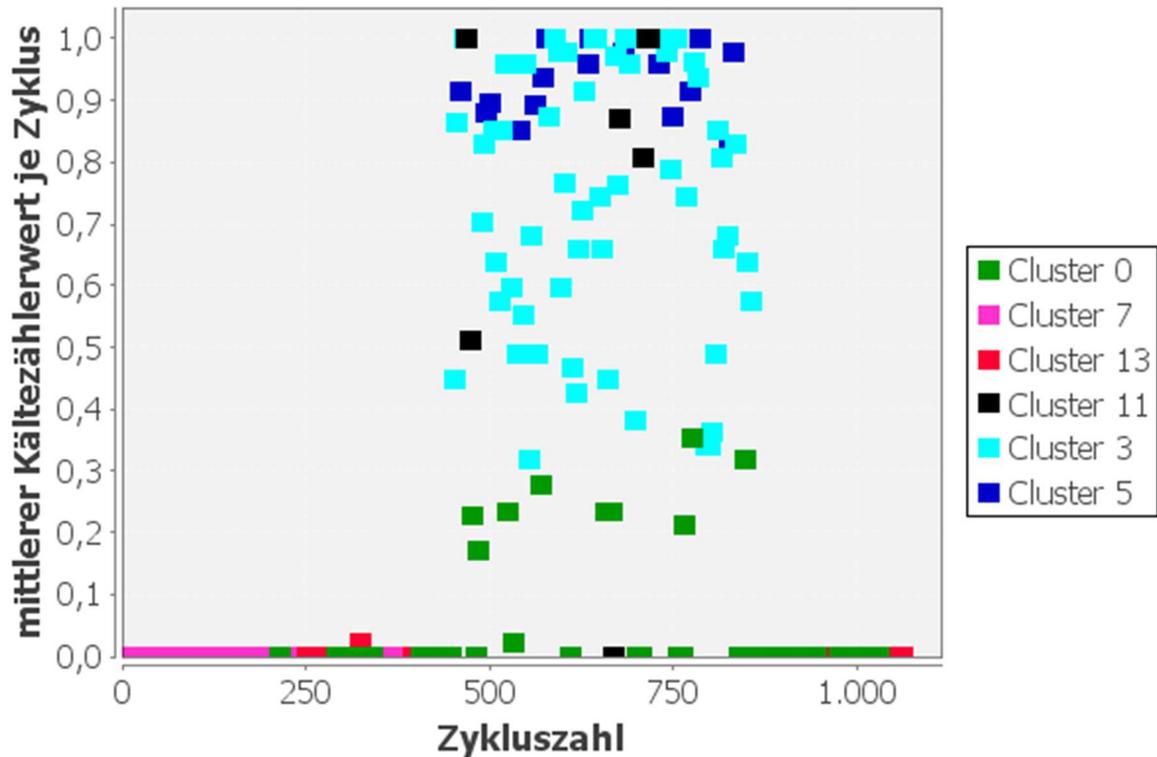
**Abbildung 34: RLT03 - Ergebnis des Clustering anhand der mittleren Außentemperatur**

Anhand von Abbildung 35 können die mittleren LE-Ventilstellungen je Zyklus nachvollzogen werden. Es wird ersichtlich, dass die Zyklen von Cluster 13 und Cluster 7 einen höheren Heizbedarf aufweisen was sich anhand der des dargestellten Wertes erkennen lässt. Dies bestätigt die korrekte Zuordnung des Anlagenverhaltens „Heizen“ zu Cluster 7 und Cluster 13. Außerdem ist erkennbar, dass die schwarz abgebildeten Ausreißer-Cluster über einen unverhältnismäßig hohen Ventilstellungswert aufweisen. Daher wird vermutet, dass es sich bei den in schwarzer Farbe dargestellten Betriebszyklen um Zyklen handelt, bei denen entweder ein Messungsfehler entstanden ist oder die Anlage nicht korrekt betrieben wurde. Dies sollte in weiteren Untersuchungen genauer betrachtet werden



**Abbildung 35: RLT03 - Ergebnis des Clustering anhand der mittleren LE-Ventilstellung**

Durch Abbildung 36 lässt sich erkennen, dass Cluster 3 und Cluster 5 korrekt als Zyklen mit Kühlbetrieb erkannt wurden. Auch weisen einige Zyklen von Cluster 0 einen Kühlbetrieb auf. Da in Abbildung 35 zu erkennen ist, dass auch einige Zyklen einen Heizbedarf aufweisen, kann Cluster 0 als Übergangsbereich zwischen dem Heizbetrieb und dem Kühlbetrieb verstanden werden. Gerade diese Cluster sollten im Anschluss an die hier angefertigte Diplomarbeit genauer untersucht werden.



**Abbildung 36: RLT03 - Ergebnis des Clustering anhand des mittleren Kältezählerwertes**

In Abbildung 37 wird die mittlere WRG-Ventilstellung der Zyklen über den Jahresverlauf dargestellt. Es wird ersichtlich, dass die gefundenen Heizcluster, Cluster 7 und Cluster 13, auch durch den Anteil der WRG-Nutzung unterschieden werden. Da bei den Zyklen von Cluster 7 niedrigere Temperaturen herrschen als bei Cluster 13 wird vermutet, dass es sich hierbei um einen fehlerhaften Anlagenbetrieb handelt. Bei steigendem Heizbedarf sollte die WRG stärker genutzt werden um Energie zu sparen.

Auch Cluster 3 und Cluster 5 werden vor allem durch den unterschiedlichen Grad der WRG-Nutzung unterschieden. Die WRG-Nutzung zur Unterstützung der Kühlung ist erst dann sinnvoll, wenn die Rücklauftemperatur niedriger ist, als die Außentemperatur. Der stärkere WRG-Betrieb bei den wärmsten Zyklen gemäß Abbildung 34 lässt also den Schluss zu, dass das Anlagenverhalten korrekt klassifiziert werden konnte.

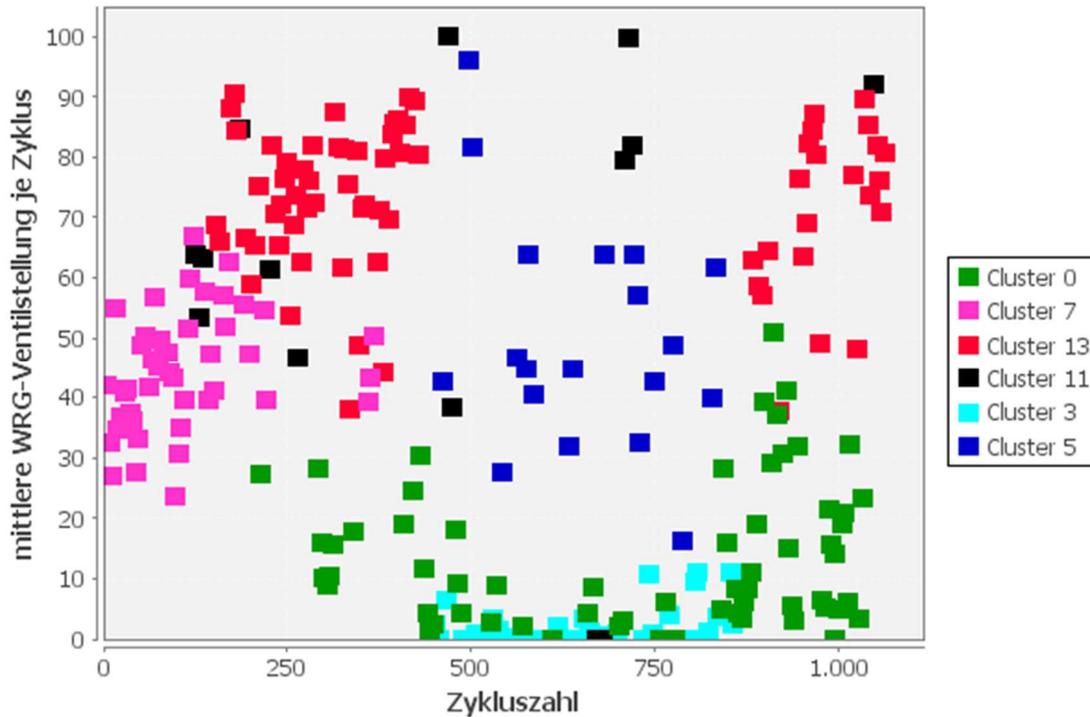


Abbildung 37: RLT03 - Ergebnis des Clustering anhand der mittleren WRG-Ventilstellung

### 4.3 Vergleich der Ergebnisse von Lüftungsanlage 06 und Lüftungsanlage 03

Anhand von Abbildung 38 kann der Vergleich der automatisierten Betriebszustandsklassifikation von Lüftungsanlage 06 unter Verwendung der Datenpunktauswahl HC I und der Lüftungsanlage 03 nach Entfernen der fehlerhaften Betriebszyklen nachvollzogen werden. Der Vergleich wird anhand der Darstellung der mittleren Außentemperatur je Zyklus vorgenommen. Dabei ist zu erkennen, dass bei beiden raumluftechnischen Anlagen eine ähnliche Aufteilung der Cluster erfolgt ist.

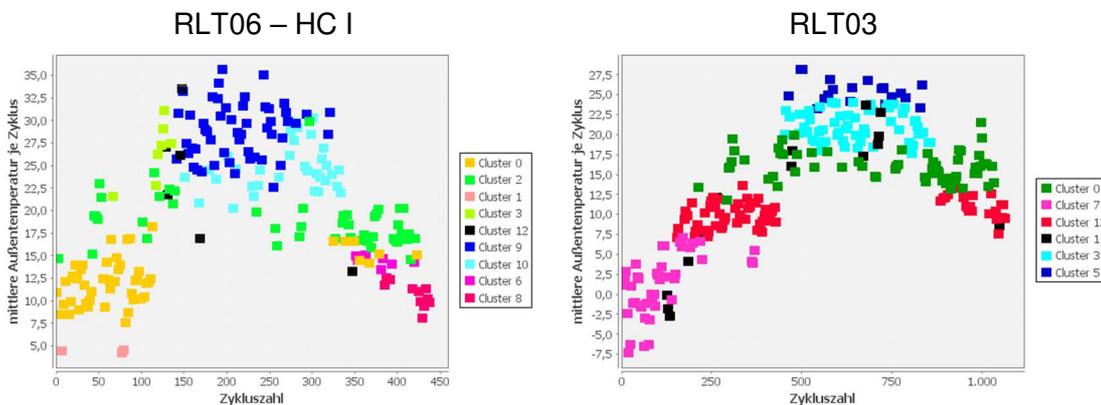
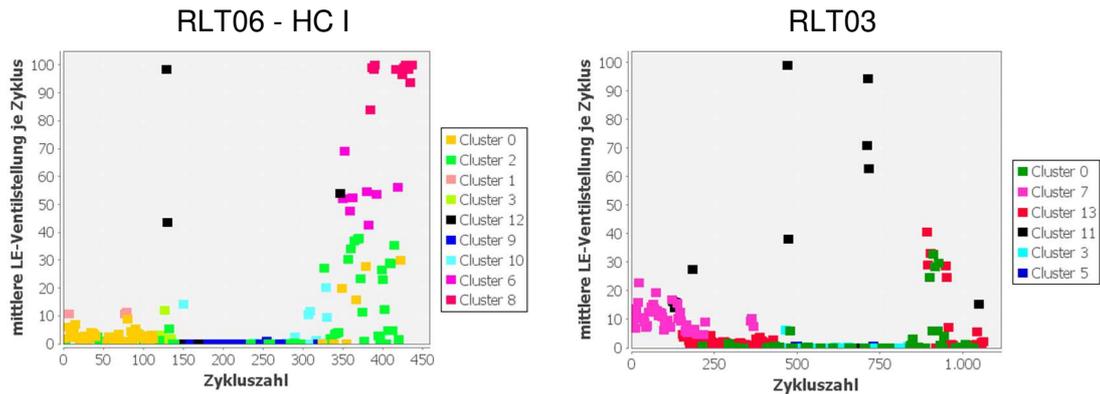


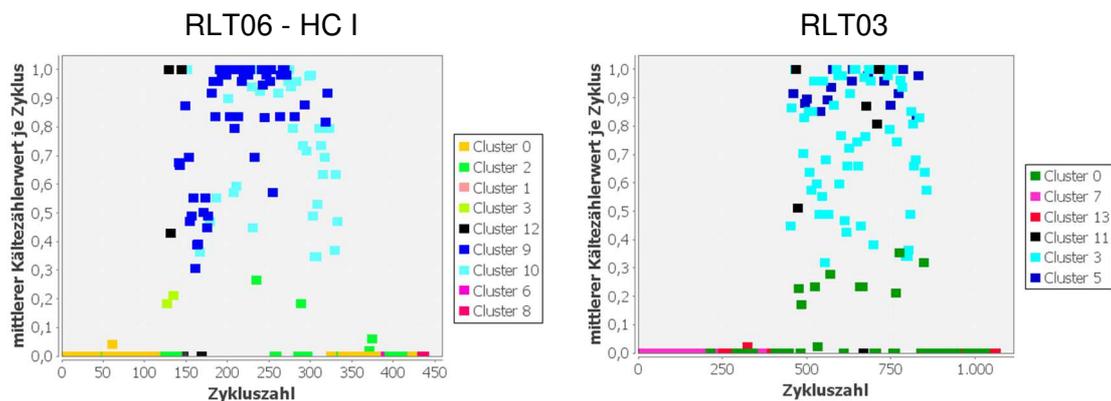
Abbildung 38: Vergleich der Ergebnisse der automatisierten Betriebszustandsklassifikationen von Lüftungsanlage 06 HC I und Lüftungsanlage 03 nach der Entfernung fehlerhafter Betriebszyklen anhand der mittleren Außentemperatur je Betriebszyklus

Da sich bei der Klassifizierung beider Anlagen ein qualitativ ähnliches Ergebnis abzeichnet, wird geschlussfolgert, dass die automatisierte Klassifikation von Betriebszuständen zu positiven Ergebnissen führt. Anhand von Abbildung 39 kann der Vergleich der Ergebnisse anhand der mittleren LE-Ventilstellungen der Analysen beider Lüftungsanlagen nachvollzogen werden. Dabei stellen die schwarzen Punkte Menge der Ausreißer-Cluster der jeweiligen Anlage dar. Da diese im Vergleich zu Zyklen in einem ähnlichen Zeitraum deutlich höhere Werte für die mittlere LE-Ventilstellungen aufweisen, kann vermutet werden, dass es sich bei diesen Zyklen um ein Fehlverhalten der Anlage handelt. Dies bedarf weiterer Untersuchung.



**Abbildung 39: Vergleich der Ergebnisse der automatisierten Betriebszustandsklassifikationen von Lüftungsanlage 06 HC I und Lüftungsanlage 03 nach der Entfernung fehlerhafter Betriebszyklen anhand der mittleren LE-Ventilstellung je Betriebszyklus**

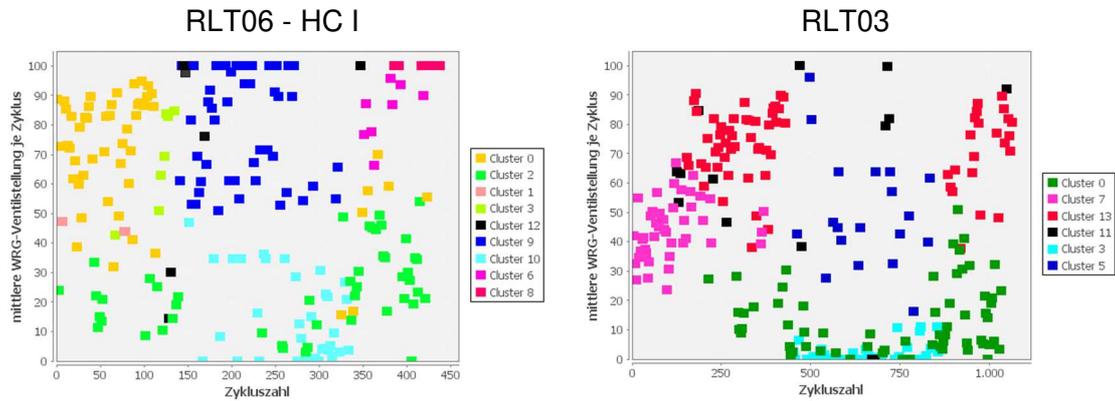
Anhand von Abbildung 40 können die mittleren Betriebsstufenwerte des Luftkühlers je Betriebszyklus nachvollzogen werden. Der Vergleich bestätigt ein ähnliches Ergebnis für beide Anlagen, wobei sich Cluster 9 und Cluster 10 links und Cluster 3 und Cluster 5 rechts durch eine deutlich höhere Betriebsstufe des Luftkühlers im Vergleich zu den anderen Clustern absetzen. Den erwähnten Clustern wurde der Betriebszustand „Kühlen“ zugeordnet.



**Abbildung 40: Vergleich der Ergebnisse der automatisierten Betriebszustandsklassifikationen von Lüftungsanlage 06 HC I und Lüftungsanlage 03 nach der Entfernung fehlerhafter Betriebszyklen anhand der mittleren Betriebsstufe Luftkühler je Betriebszyklus**

Anhand von Abbildung 41 kann der Vergleich der Ergebnisse der Betriebszustandsklassifikation anhand der mittleren WRG-Stellung je Zyklus für beide Anlagen nachvollzogen

werden. Man erkennt bei beiden Lüftungsanlagen die Unterteilung der mit „Kühlen“ klassifizierten Betriebszustände an der mittleren WRG-Ventilstellung je Zyklus.



**Abbildung 41: Vergleich der Ergebnisse der automatisierten Betriebszustandsklassifikationen von Lüftungsanlage 06 HC I und Lüftungsanlage 03 nach der Entfernung fehlerhafter Betriebszyklen anhand der mittleren WRG- je Betriebszyklus**

## 5 Schlussbetrachtung

### 5.1 Zusammenfassung

In der hier vorliegenden Diplomarbeit wurde untersucht, inwiefern sich die Analyse von Betriebsdaten raumluftechnischer Anlagen nach dem Ansatz von [Habib et al., 2016] automatisieren lässt. Dabei wurde das Verfahren auf raumluftechnische Anlagen übertragen und unter Verwendung des Konstanz Information Miners implementiert. Damit sollte ermöglicht werden, die zu erwartenden Betriebszustände raumluftechnischer Anlagen zu klassifizieren. Es wurden zunächst die notwendigen theoretischen Grundlagen bezüglich der Gebäudeleittechnik, sowie die verwendeten statistischen Verfahren erläutert. Im Umsetzungsteil der Arbeit wurden die zu untersuchenden raumluftechnischen Anlagensysteme vorgestellt. Es wurde auf deren täglichen Betriebszeitraum eingegangen und deren Aufbau vorgestellt. Zudem wurde die Interaktion der enthaltenen Anlagen-Systemkomponenten erläutert und eine Liste der in der statistischen Analyse verwendeten Datenpunkte geliefert. Zudem wurde dargestellt, wie das Datenanalyseverfahren auf die raumluftechnischen Anlagen angepasst wurde. Daraufhin wurden die zu erwartenden Anlagenbetriebszustände anhand ausschlaggebender Systemkomponenten veranschaulicht. Der nachfolgende Teil der Arbeit wurde darauf verwendet, die Ergebnisse der Analyse zweier unterschiedlicher Anlagen aufzuzeigen. Zunächst wurde anhand der Eingangsdaten von Lüftungsanlage 06 untersucht, inwiefern die Unterscheidung von AN- und AUS-Zuständen unter Verwendung verschiedener Normierungsverfahren und Datenpunkt-Auswahlen umsetzbar ist. Beurteilt wurde dies anhand von Datenpunkten, die den Betrieb der Ventilatoren bestätigen.

Danach wurde untersucht inwiefern sich die Analysealgorithmen zur automatisierten Klassifizierung von Betriebszustände eignen. Dafür wurde eine geeignete Datenpunktauswahl für das hierarchische Clustering anhand entscheidender Systemkomponenten festgelegt. Bei der Klassifizierung der Betriebszustände wurde aufgezeigt, dass bei der hierarchischen Clusteranalyse Ausreißer-Cluster entstehen. Zudem wurde aufgezeigt, dass die Evaluation der Anzahl der Cluster unter Verwendung der GAP-Statistic Methode im Rahmen der Untersuchung zu keinem verwertbaren Ergebnis geführt hat. Deswegen wurde ein Schwellenwert zur Festlegung der idealen Clusterzahl eingeführt, diskutiert und auf einen Wert von 40 % der maximalen Euklidischen Distanz festgelegt. Im darauffolgenden Abschnitt wurde aufgezeigt, inwiefern sich das Ergebnis der Betriebszustandsklassifizierung ändert, wenn unterschiedliche Anzahlen an Chunks gewählt werden. Zudem wurde eine Anzahl von 120 Chunks für die Analyse bestimmt. Daraufhin wurde den gefundenen Clustern der Analyse von Lüftungsanlage 06 anhand der dafür entworfenen Evaluationstabelle ein Betriebszustand zugeordnet. Die Ergebnisse wurden anhand der wichtigsten Datenpunkte aufgezeigt.

Anschließend sollte anhand einer zweiten Datenpunktauswahl untersucht werden, welchen Einfluss die Außentemperatur auf das Ergebnis der Clusteranalyse hat. Wiederum wurden den somit gefundenen Clustern anhand der Werte der Evaluationstabelle Anlagenbetriebszustände zugeordnet. Daraufhin wurden die Ergebnisse der Clusteranalysen der verschiedenen Datenpunktauswahlen untereinander verglichen, um aufzuzeigen, dass die Außentemperatur für das Ergebnis der Betriebszustandsklassifikation im gezeigten Fall eine untergeordnete Rolle spielt.

Im zweiten Teil der Ergebnisse der Datenanalyse wurde Lüftungsanlage 03 untersucht. Da bei dieser bereits im Vorfeld ein Fehlverhalten erkannt werden konnte, sollte untersucht werden inwiefern sich die Anlage für die automatisierte Erkennung von Gebäudebetriebszyklen und die Zuweisung von Betriebszuständen eignet. Dabei wurde die Unterscheidung von AN- und AUS-Zuständen unter Miteinbezug der im Vorfeld als fehlerhaft erkannten Betriebszustände durchgeführt. Anschließend wurde untersucht, inwiefern die Klassifikation von Betriebszuständen unter Miteinbezug des erwähnten Fehlverhaltens der Lüftungsanlage möglich ist. Da dies zu keinem verwertbaren Ergebnis geführt hat wurde das Fehlverhalten entfernt. Die Klassifikation wurde daraufhin erneut durchgeführt. Den bei der Analyse entstandenen Clustern wurden erneut Betriebszustände zugeordnet und im nächsten Schritt die Ergebnisse der Datenanalyse dargestellt.

## 5.2 Ergebnisse der Arbeit

Bei der Untersuchung von Lüftungsanlage 06 wurde gezeigt, dass die Unterscheidung von Betriebs- und Ruhezuständen erfolgreich durchgeführt werden konnte. Dabei wurde festgestellt, dass die Wahl des Normierungsverfahrens einen entscheidenden Einfluss auf die Qualität des Ergebnisses hat. Die Verwendung unterschiedlicher Datenpunktauswahlen in Szenario 1 und Szenario 2 hat erkennen lassen, dass die Wahl geeigneter Datenpunkte bei der untersuchten Anlage nur eine untergeordnete Rolle spielt. Es wird daher vermutet, dass die Datenpunktauswahl auch bei anderen raumluftechnischen Anlagen nur eine untergeordnete Rolle spielt. Im Rahmen der Klassifikation der Betriebszustände konnte festgestellt werden, dass sich die GAP-Statistic nicht zur Festlegung der idealen Clusterzahl bei den untersuchten Anlagensystemen eignet. Dies steht vermutlich mit den Ausreißer-Clustern in Zusammenhang. Zudem wird vermutet, dass diese Ausreißer-Cluster auf ein Fehlverhalten der Lüftungsanlage hinweisen. Eine genaue Untersuchung dessen übersteigt den Rahmen der hier angefertigten Diplomarbeit und sollte im Folgenden untersucht werden. Als Alternative zur GAP-Statistic wurde ein Schwellenwert eingeführt, der sich prozentual an der maximalen euklidischen Distanz der hierarchischen Clusteranalyse orientiert.

Es wurde anhand des Vergleichs von Analysen mit unterschiedlichen Schwellenwerten aufgezeigt, warum die Auswahl auf einen Schwellenwert von 40 % gefallen ist. Es wird jedoch angenommen, dass die Wahl eines exakten Schwellenwertes im hohen Maß von der Beschaffenheit der jeweiligen Datenpunkte abhängt, da sich durch diese die Distanzen ergeben, welche im Rahmen der hierarchischen Clusteranalyse verwendet werden. Anschließend wurde untersucht, welche Auswirkung die Wahl unterschiedlicher Chunk Anzahlen auf das Analyseergebnis hat. Dabei hat sich gezeigt, dass nur ein geringer Anteil der Betriebszyklen anders zugeordnet wurde.

Daraufhin konnte angenommen werden, dass sich die Verwendung von 120 Chunks zur Datenanalyse für die untersuchten Anlagensysteme eignet. Des Weiteren konnte aus dem Vergleich der Clusteranalyse mit 60 und mit 120 Chunks erschlossen werden, dass die Anzahl der Chunks nur eine geringe Auswirkung auf das Ergebnis der automatisierten Klassifizierung von Betriebszuständen hat. Aus dem Vergleich der Ergebnisse der Klassifikation der Betriebszustände unter Verwendung unterschiedlicher Datenpunktauswahlen konnte geschlussfolgert werden, dass die Außentemperatur das Ergebnis der Klassifikation wenig beeinflusst.

Bei der Untersuchung von Lüftungsanlage 03 hat sich gezeigt, dass die Unterscheidung von Ruhe- und Betriebszuständen möglich ist, obwohl die Anlage ein Fehlverhalten aufweist. Wiederum hat sich gezeigt, dass die Wahl der Normierungsmethode einen Einfluss auf das Ergebnis hat. Es wird vermutet, dass die Unterscheidung von AN- oder AUS-Zuständen auch bei anderen raumluftechnischen Anlagen erfolgreich durchgeführt werden kann. Dies bedarf weiterer Analyse. Das hierarchische Clustering wurde anhand der im Vorherigen dargestellten Datenpunktauswahl erfolgreich durchgeführt.

Dabei konnte gezeigt werden, dass die Klassifizierung der Betriebszustände erfolgreich war. Dies war nur möglich, wenn die im Rahmen der Voranalyse als Fehlverhalten erkannten Betriebszustände vor der Betriebszustandsklassifizierung entfernt wurden. Es wird daher angenommen, dass die aufgezeigte Analysemethodik nur eingeschränkt für die Klassifizierung von Gebäudebetriebszuständen geeignet ist. Lediglich könnte die Tatsache, dass eine Clusteranalyse keine klare Klassifizierung liefert als Indikator verstanden werden, dass bei der Anlage ein Fehlverhalten vorliegt.

Bei dem Vergleich der Ergebnisse von Lüftungsanlage 06 und Lüftungsanlage 03 konnte eine qualitativ ähnliche Clusterbildung festgestellt werden, woraus geschlussfolgert werden kann, dass die automatisierte Betriebszustandsklassifikation zu einem positiven Ergebnis führt. Wie sich gezeigt hat, können die verwendeten Algorithmen ohne die professionelle Expertise hinsichtlich der Gebäudeenergie-technik umgesetzt werden. Dennoch ist ein profundes Verständnis raumluftechnischer Anlagen bei der Beurteilung und somit weiteren Verwendung der Ergebnisse erforderlich. Sensoren müssen im Vorfeld der Analyse auf ihre Verwendbarkeit beurteilt werden und ggf. aufbereitet werden. Fehlende Expertise bei der Vorbereitung der Daten könnte dazu führen, dass beispielsweise Sensorausfälle nicht berücksichtigt werden, oder Stundenzähler als absolute Werte gedeutet werden. Auch bei der Wahl der Sensoren für die Clustering-Algorithmen sollte darauf geachtet werden, dass alle essentiellen Bestandteile der jeweiligen Anlage berücksichtigt wurden. Aufgrund möglicher Unterschiede im Aufbau von raumluftechnischen Anlagen unter Verwendung unterschiedlicher Komponenten könnte eine zu generische Auswahl an Datenpunkten zu fehlerhaften Ergebnissen führen. Bei dem k-Means-Clustering hat sich gezeigt, dass bei der Verwendung einer großen Menge unterschiedlicher Datenpunkte ein positives Ergebnis erzielt werden kann, sofern eine geeignete Normierungsmethode gewählt wird und alle essentiellen Anlagenteile berücksichtigt werden. Bei Anlage 03 war erkennbar, dass die Verwendung der Eingangsdaten beim hierarchischen Clustering zu keinem befriedigenden Ergebnis geführt hätten. Nachdem offensichtliches Fehlverhalten entfernt wurde, konnte die Anlage bearbeitet werden, jedoch setzt dies eindeutig professionelle Expertise und Verständnis von Anlagentechnik voraus.

Die größten Vorzüge von KNIME sind die schnelle und intuitive Bedienbarkeit sowie das Vorhandensein einer Bibliothek an verfügbaren Modulen, die dem Nutzer einen großen Teil der Implementierungsarbeit abnehmen. Auf der anderen Seite wirkt sich die Struktur aus vorgefertigten Lösungen nachteilig auf den Grad der Individualisierbarkeit der Implementierung aus. So können bei einem k-Means Clustering bspw. Anzahl der Cluster und auch Anzahl der Iterationen eingestellt und ggf. auch dynamisch verändert werden. Beschränkt man sich bei der Implementierung auf die KNIME Nodes, so hat man die Möglichkeit zu einer agilen Datenanalyse. Wenn der benötigte Analyse-Algorithmus nicht als KNIME-Modul vorliegt, kann entweder ein externer Programmcode integriert

oder der Algorithmus in seine Teilschritte zerlegt und diese mit KNIME-Nodes nachgebildet werden. Bezogen auf den aufgebrauchten Arbeitsaufwand ist die direkte Anbindung von externem Programmcode die schnellere Möglichkeit. Bezogen auf die Berechnungsdauer haben sich die KNIME-internen Lösungen als deutlich performanter erwiesen. Die im Rahmen dieser Diplomarbeit entstandene Implementierung kann in Anlage 1 nachvollzogen werden.

### 5.3 Ausblick

Die hier angefertigte Arbeit soll einen Beitrag zur automatischen Fehlererkennung beim Betrieb von raumluftechnischen Anlagen bilden. Dabei soll dem bearbeitenden Datenanalysten die Möglichkeit gegeben werden, über die zu analysierenden Daten einen Überblick zu bekommen bzw. sich eine erste Einordnung der zu analysierenden Datenmenge generieren zu lassen.

Ein möglicher Ansatz zur Verbesserung der Genauigkeit der Ergebnisse der Klassifikation von Betriebszuständen kann es sein, die gefundenen Betriebszyklen in mehrere Teile zu zerlegen. Da die Außentemperatur über den Tagesverlauf großen Veränderungen unterworfen ist, kann nicht ausgeschlossen werden, dass über den Tagesverlauf unterschiedliche Anlagenbetriebszustände auftreten. Daher könnte eine Unterteilung der Betriebszyklen, beispielsweise anhand fester zeitlicher Rahmen, dafür Sorge tragen, ein genaueres Ergebnis zu erzielen. Die im Rahmen dieser Arbeit nicht genauer beleuchteten Ausreißer-Cluster können in zukünftigen Arbeiten auf ihr Entstehen untersucht werden. Sie könnten ihre Ursache in dem Fehlverhalten der Anlage haben. Dadurch bestünde die Möglichkeit, das entwickelte Untersuchungsverfahren zu nutzen um Ausreißer-Cluster festzustellen und nur noch diese zu untersuchen. Dies könnte zu einer Zeitersparnis bei der Fehleranalyse der Betriebsbeurteilung von raumluftechnischen Anlagen führen. Zudem könnte die Untersuchung der im ersten Schritt entfernten AUS-Zyklen durch einen ähnlichen Klassifikationsansatz untersucht werden. Dabei könnte Fehlverhalten, wie beispielsweise der Betrieb einzelner Systemkomponenten trotz ausgeschalteter Ventilatoren nachgewiesen werden.

In Zukunft kann der in dieser Diplomarbeit dargestellte Ansatz dafür genutzt werden, erkannte Betriebszustände nicht nur zu klassifizieren, sondern auch nach ihrer Betriebsqualität zu beurteilen. Dafür sollten einzelne Betriebszyklen hinsichtlich energetisch sinnvoller bzw. ungünstiger Zustände untersucht und beleuchtet werden, inwiefern Unterschiede der Betriebsqualität unter Verwendung der hierarchischen Clusteranalyse erkannt werden können.

Bei der Justierung der verwendeten Parameter hat sich gezeigt, dass die MINMAX-Normierungsmethode die besten Ergebnisse im Rahmen der Datenanalyse liefert. Im Anschluss an diese Erkenntnis können bezüglich der Ursachen weitere Untersuchungen angestellt werden.

Da bei der Umsetzung darauf verzichtet wurde, die Piecewise Aggregate Approximation Methode auf den Datensatz anzuwenden, kann in zukünftigem untersucht werden, inwiefern die Größe der gewählten Zeitintervalle bei der Approximation der Daten Einfluss auf das Ergebnis der Clusteranalyse nimmt.

## Literaturverzeichnis

### Monographien und Aufsätze in Fachzeitschriften

- [Habib et al, 2016] Habib, U./Hayat, K./Zucker, G.: Complex building's energy system operation patterns analysis using bag of words representation with hierarchical clustering, in: Complex Adaptive Systems Modeling, 4:8, (2016), S. 1–20
- [Habib et al, 2015] Habib, U./Zucker, G.: Finding the Different Patterns in Buildings Data Using Bag of Words Representation with Clustering, in: 2015 13th International Conference on Frontiers of Information Technology, (2015), S. 303–308
- [Lin et al., 2007] Lin, J./Keogh, E./Wei, L./Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series, in: Data Mining Knowledge Discovery, 15(2), (2007), S. 107–144
- [Tibshirani et al., 2001] Tibshirani, R./Walther, G./Hastie, T.: Estimating the number of clusters in a data set via the gap statistic, in: Journal of the Royal Statistical Society 63(2), (2001), S. 411–423
- [Zucker et al., 2015] Zucker, G./Habib, U./Blöchle, M./Judex, F./Leber, T.: Sanitation and Analysis of Operation Data in Energy Systems, in: Energies, 11, (2015), S. 12776–12794
- [Lloyd, 1982] Lloyd, S.: Least squares quantization in PCM, in: IEEE Transactions on Information Theory, 2, (1982), S. 129 – 137 [CrossRef]
- [Jain, 2010] Jain, A. K.: Data clustering: 50 years beyond k-Means, in: Pattern Recognit. Lett., 31, (2010), S. 651 - 666
- [Vesanto, J. et al, 2000] Vesanto, J./Alhoniemi, E.: Clustering of the self-organizing map, in IEEE Transactions on Neural Networks, Vol. 11, No. 3, (2000), S. 586-600
- [Bollin et al., 2009] Bollin, E. (Hrsg.)/ Becker, M./Boggasch, E./Fraaß, M./Karbach, A. /Ritzenhoff, P./Striebel, D.: Automation regenerativer Wärme- und Kälteversorgung von Gebäuden; Verlag: Vieweg+Teuber; 1. Auflage; (2009)
- [Buderus, 1994] Buderus Heiztechnik GmbH: Handbuch für Heizungstechnik; Verlag: Beuth; 33. Auflage; (1994)
- [Eichmann, 1998] Eichmann, R. A.: Grundlagen der Klimatechnik; Verlag: C. F. Müller; 1. Auflage; (1998)
- [Schmid et al., 2005] Schmid, C. (Hrsg.)/Zürcher, C./Nipkow J./Baumgartner, T.: Heizung, Lüftung, Elektrizität - Energietechnik im Gebäude; Band 5; Verlag: vdf; 3. Auflage; (2005)

- [Schmidt et al., 2011] Schmidt, M. (Hrsg.)/Hörner, B.: Handbuch der Klimatechnik; Band 2; Verlag: VDE; 5. Auflage (2011)
- [Schramek et al., 2009/2010] Schramek, E.-R./Recknagel, H./Sprenger, E.: Taschenbuch für Heizung und Klimatechnik; Verlag: Oldenbourg Industrieverlag; 74. Auflage; (2009/2010)

## Internetquellen

<https://www.fibav.de/wp-content/uploads/2018/02/w%C3%A4rmer%C3%BCckgewinnung.jpg>, Stand 25.04.2018

<https://www.baunetzwissen.de/imgs/1/3/2/1/4/7/5/5e6b3ec8c50bcc5d.jpg>, Stand 30.05.2018

<https://wirtschaftslexikon.gabler.de/definition/clusteranalyse-28221>, Stand 27.05.2018

[https://cdn-images-1.medium.com/max/800/0\\*rrzG3LyOnAvOepbJ.png](https://cdn-images-1.medium.com/max/800/0*rrzG3LyOnAvOepbJ.png), Stand 26.05.2018

<https://wirtschaftslexikon.gabler.de/definition/dendrogramm-31008#references>, Stand 13.05.2018

<https://wirtschaftslexikon.gabler.de/sites/default/files/styles/keyword/public/images/wilex/73528.png?itok=KI07mdMF>, Stand 26.05.2018

[https://jmotif.github.io/sax-vsm\\_site/morea/assets/sax\\_paa1.png](https://jmotif.github.io/sax-vsm_site/morea/assets/sax_paa1.png), Stand 24.04.2018

<https://www.knime.com/knime-open-source-story>, Stand 27.02.2018

<https://www.knime.com/software>, Stand 27.02.2018

## Anlagenverzeichnis

Anlage 01	Daten-CD
Anlage 02	Dendrogramm Relation RLT06 mit Schwellenwert 40%, 60 Chunks und 120 Chunks
Anlage 03	Cluster-Evaluationstabelle RLT06 HC I mit Schwellenwert 30%, 120 Chunks
Anlage 04	Cluster-Evaluationstabelle RLT06 HC I mit Schwellenwert 40%, 120 Chunks
Anlage 05	Cluster-Evaluationstabelle RLT06 HC I mit Schwellenwert 50%, 120 Chunks
Anlage 06	Cluster-Evaluationstabelle RLT06 HC I mit Schwellenwert 40%, 60 Chunks
Anlage 07	Cluster-Evaluationstabelle RLT06 HC II mit Schwellenwert 40%, 120 Chunks
Anlage 08	Cluster-Evaluationstabelle RLT03 mit Schwellenwert 40%, 120 Chunks, mit Fehlverhalten
Anlage 09	Cluster-Evaluationstabelle RLT03 mit Schwellenwert 40%, 120 Chunks, ohne Fehlverhalten

## **Erklärung zur selbstständigen Abfassung der Diplomarbeit**

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und ist auch noch nicht veröffentlicht worden.

Dresden, den 31. Mai 2018

Markus Bär

## Anlage 02: Dendrogramm Relation RLT06 mit Schwellenwert 40%, 60 Chunks und 120 Chunks

