

HTW Berlin
Fachbereich 2 - Ingenieurwissenschaften
Umweltinformatik

smartB Energy Management GmbH
Hardenbergstr. 9a
10623 Berlin

Bachelorarbeit

Entwicklung und Integration eines Systems zur Witterungsanalyse und -bereinigung in ein bestehendes Energie Metering und Monitoring System

Knut Hühne

Matrikel Nr.: 0539442

2015-09-03

Erstgutachter: Prof. Dr. Volker Wohlgemuth

Zweitgutachter: Dr. Marco Blumendorf

Abstract

Äußere Faktoren wie etwa die Außentemperatur haben oft einen Einfluss auf den Stromverbrauch von Gebäuden. Diese Einflüsse können analysiert werden und eine Bereinigung des Stromverbrauchs um die äußeren Einflüsse kann durchgeführt werden, was zu einer besseren Vergleichbarkeit führen kann. Hierfür kann eine Witterungsbereinigung - ähnlich der verbreiteten Witterungsbereinigung von Heizdaten - genutzt werden. Dabei müssen die Unterschiede beim Stromverbrauch berücksichtigt werden. In dieser Arbeit wurden öffentliche Stromverbrauchsdatensätze auf ihre Korrelation mit der Außentemperatur untersucht und die Besonderheiten des Zusammenhangs zwischen Stromverbrauch und Außentemperatur dargestellt. Aufbauend auf diesen Ergebnissen wurde ein Modul zur Witterungsanalyse und -bereinigung konzipiert. Das Modul wurde anschließend mithilfe von Python, angular.js und anderen Webtechnologien entwickelt und in ein bestehendes Energiemanagementsystem integriert. Das neu entwickelte Modul ermöglicht es Nutzern, den Stromverbrauch ihrer Gebäude auf Korrelation mit der Außentemperatur zu untersuchen und eine Witterungsbereinigung um Heiz- und Kühlgradtage durchzuführen. Die Bereinigung von Stromdaten um äußere Einflüsse ist bisher im Vergleich zur Bereinigung von Heizdaten wenig verbreitet. Die Arbeit zeigt jedoch, dass auch hier eine genauere Untersuchung und Bereinigung einen Mehrwert für Nutzer bieten kann.

External factors like the outside temperature influence the power consumption of buildings. Those influences can be analyzed and the power consumption can be adjusted for the external factors. This can lead to a better comparability. In order to achieve this, a weather adjustment - similar to the popular weather adjustment for heating power - can be done. In this paper, existing public power consumption datasets were analyzed and the characteristics of the link to the outside temperature were shown. Based on those findings, a new module for the analysis and adjustment of outside temperature was planned. This module was implemented using Python, angular.js and other web technologies and integrated into an existing energy management system. The new module allows users to check the power consumption of their buildings for correlations with the outside temperature. It furthermore enables them to do a weather adjustment for heating or cooling degree days. Adjusting power consumption of buildings for the outside temperature is less popular than adjusting heating consumption but this paper shows that the analysis and adjustment of those factors can also be of value to the users.

Inhaltsverzeichnis

Inhaltsverzeichnis	ii
Tabellenverzeichnis	iv
Abbildungsverzeichnis	v
Quellcodeverzeichnis	vi
Abkürzungsverzeichnis	vii
1 Einleitung	1
1.1 Motivation	2
1.2 Zielstellung	5
1.3 Aufbau der Arbeit	6
2 Bezug zur Umweltinformatik	8
3 Vorstellung der smartB Energy Management GmbH	10
4 Wetterdaten zur Datenanalyse	13
4.1 Anforderungen an die Daten	13
4.2 Daten des Deutschen Wetterdienstes	14
4.2.1 Historische Wetterdaten	14
4.2.2 Klimafaktoren	14
4.2.3 TRY - Testreferenzjahre	15
4.3 Daten der NOAA	15
5 Korrelationsanalyse von Beispieldatensätzen	18
5.1 Vorgehen	18
5.1.1 Datenquellen	18
5.1.2 Methode zur Bewertung der Korrelation	20
5.2 Die Gebäude	21
5.2.1 Haushalte in Austin, USA	21
5.2.2 Haushalt in Sceaux	23
5.2.3 Bürogebäude in Berlin	24
5.2.4 Simulierter Datensatz	24
5.3 Diskussion der Ergebnisse	25

6	Konzeptionierung des Systems	27
6.1	Visualisierung	27
6.2	Korrelationsanalyse	29
6.3	Witterungsbereinigung	30
6.3.1	Definition Gradtagzahl	31
6.3.2	Bereinigung über Gradtagzahl	32
6.3.3	Umsetzung der Witterungsbereinigung	33
6.4	Nonfunktionale Anforderungen	35
7	Entwicklung des Systems	37
7.1	Genutzte Technologien	38
7.1.1	Datenanalyse	38
7.1.2	Frontend	41
7.2	Automatisierte Datenbeschaffung	42
7.3	Einbindung der Gebäudekoordinaten	44
7.4	Visualisierung von Korrelationen	47
7.5	Automatisierte Korrelationsanalyse	48
7.6	Witterungsbereinigung über Gradtagzahlen	51
8	Test des Systems	55
8.1	Expertengespräch	56
8.2	Nutzertest mit unerfahrenem Nutzer	56
8.3	Erfüllung der funktionalen Anforderungen	58
8.4	Erfüllung der nonfunktionalen Anforderungen	58
8.5	Diskussion der Ergebnisse	59
9	Fazit und Ausblick	61
A	Übersicht über die veröffentlichten NILM-Datensätze	65
	Literatur	72

Tabellenverzeichnis

5.1	Öffentliche NILM-Datensätze	20
5.2	Korrelation der Beispieldatensätze	26
6.1	Beispiel Heizgradtage (Basis 15° Celsius)	32
6.2	Beispiel Witterungsbereinigung (Bereinigt um das langjährige Mittel 1200)	33

Abbildungsverzeichnis

3.1	Das smartB Web Interface	12
4.1	Die TRY-Regionen	17
5.1	Diagramme für Pecan Street Id 585	22
5.2	Diagramme für Frankreich	23
5.3	Diagramme für Berlin	24
5.4	Diagramme für simulierten Datensatz	25
7.1	Architekturdiagramm	37
7.2	Die jupyter Programmierumgebung	39
7.3	Ordnerstruktur der Wetterdaten	44
7.4	Die neue Kartenkomponente in den Gebäudeeinstellungen	46
7.5	Die Datenvisualisierung im Frontend	48
7.6	Vorgehen zur Korrelationsanalyse	50
7.7	Das Modul zur Witterungsbereinigung	53

Quellcodeverzeichnis

1	SQL Anfrage an Pecan Street Metadatenbank	22
2	Beispiel zur Datenanalyse mit Pandas	40
3	Zwei-Achsen Liniendiagramm mit matplotlib	40
4	Aufruf der Funktion zum Download der Wetterdaten	43
5	JSON Beispiel für vom System gefundene Korrelation	51
6	Beispiel für Gradtagzahl JSON	54

Abkürzungsverzeichnis

API Application Programming Interface.

BBSR Bundesinstituts für Bau-, Stadt- und Raumforschung.

CSV Comma Separated Values.

DWD Deutscher Wetterdienst.

EnEV Energieeinsparverordnung.

FTP File Transfer Protocoll.

ISD Integrated Surface Data.

ISH Integrated Surface Hourly.

JSON Javascript Object Notation.

METAR Message d'observation météorologique régulière pour l'aviation.

NILM Non Intrusive Load Monitoring.

NOAA National Oceanic and Atmospheric Administration.

ORM Objekt Relationaler Mapper.

REST Representational State Transfer.

SPA Single Page Application.

SVG Scalable Vector Graphics.

UBA Umweltbundesamt.

URL Uniform Resource Locator.

VDI Verein Deutscher Ingenieure.

1 Einleitung

Im *Nationalen Aktionsplan Energieeffizienz*¹ hat die Bundesregierung das Ziel festgelegt, bis zum Jahr 2050 den Primärenergiebedarf der Bundesrepublik um 80 Prozent im Vergleich zu 2008 zu reduzieren. Um dieses Ziel zu erreichen, ist es wichtig, dass Unternehmen und Energiemanager Informationen zum Verbrauch ihrer Gebäude erhalten und nutzen und somit Energietransparenz erlangen können. Das Bundesministerium für Wirtschaft und Energie zeigt in seinem Bericht, dass es hier jedoch noch erhebliches Potential gibt.

Viele Haushalte und Unternehmen können ihren Energieverbrauch nur grob oder gar nicht quantifizieren oder die Einsparpotenziale akkurat erkennen und bewerten. So verfügen nur etwa 14 Prozent der Unternehmen laut einer Umfrage der dena über ein Energiemanagementsystem. Oft ist nicht bekannt, wie viel Energie im Jahr verbraucht wird, und folglich ebenso wenig, wie Energieverbrauch und Kosten gesenkt werden können.²

Das webbasierte System der smartB Energy Management GmbH ermöglicht es Unternehmen, mit Hilfe von Smart Metern und Methoden des Machine Learning ihren Stromverbrauch in einer Auflösung von 50 Hz zu analysieren und Rückschlüsse auf einzelne Verbraucher und den Energieverbrauch als Ganzes zu ziehen.

Laut einer Untersuchung des Umweltbundesamt (UBA) variiert der Energieverbrauch von betrieblich genutzten Gebäuden sehr, die meisten Gebäude haben jedoch ein Einsparpotential von etwa 10 bis 30 Prozent. Hier kann das von smartB entwickelte System helfen,

¹Energieeffizienz 2014.

²Energieeffizienz 2014, S. 33.

Klarheit zu schaffen und Einsparpotentiale aufzudecken.³

Ein Thema, das vom bestehenden System noch nicht abgedeckt wird, ist die Analyse des Zusammenhangs zwischen Klimaeinflüssen und dem Stromverbrauch. Diese Einflüsse zu untersuchen und eine Witterungsbereinigung - ähnlich der Witterungsbereinigung von Heizdaten für den Energieausweis von Gebäuden - zu konzipieren ist das Thema dieser Arbeit. Dem Nutzer soll eine einfache Möglichkeit an die Hand gegeben werden, Zusammenhänge zu visualisieren und bereinigte Verbräuche zu betrachten.

1.1 Motivation

Seit der Einführung der *Richtlinie über die Gesamtenergieeffizienz von Gebäuden*⁴ muss europaweit für alle Bestandsgebäude und Neubauten ein Energieausweis vorliegen, der über ihre Energieeffizienz informiert.⁵ Bei der Ausstellung des Energieausweises nach §17 Energieeinsparverordnung (EnEV) werden zwei unterschiedliche Möglichkeiten zur Auswahl gestellt: Ein Energieausweis auf Basis des berechneten Energiebedarfs (Energiebedarfsausweis) oder auf Basis des gemessenen Verbrauchs (Energieverbrauchsausweis).⁶

Wenn der Energieausweis auf Basis des berechneten Energiebedarfs ausgestellt werden soll, werden in technischen Normen festgelegte Faktoren genutzt, um z. B. anhand der Qualität der Fenster, Außenwände und der Heizanlage den Verbrauch des Gebäudes zu bestimmen.⁷

Die Alternative hierzu ist die Berechnung des Energieverbrauchs anhand des tatsächlich gemessenen Verbrauchs der vergangenen Jahre. Um hier aber eine Vergleichbarkeit zwischen mehreren Jahren und unterschiedlichen Regionen zu erreichen, müssen die Werte nach §19(1) EnEV witterungsbereinigt werden. Hierfür werden die Energieverbräuche

³Bundesregierung / Energie"tatort" Büro – Wo kann ich am Schreibtisch Energie sparen? 2015, vgl.

⁴<http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=uriserv:en0021>, abgerufen 2015-08-18

⁵vgl. Weglage 2010, S. 3.

⁶vgl. *Verordnung über energiesparenden Wärmeschutz und energiesparende Anlagentechnik bei Gebäuden (Energieeinsparverordnung - EnEV)* 2007, §17.

⁷vgl. Weglage 2010, S. 25.

durch einen orts- und zeitabhängigen Faktor bereinigt, der vom Deutschen Wetterdienst bereitgestellt wird. Durch die Witterungsbereinigung werden zwei Problemfelder abgedeckt: der Vergleich der Energieeffizienz von Gebäuden in unterschiedlichen Regionen und der Vorher/Nachher Vergleich nach der Einführung einer Maßnahme zur Steigerung der Energieeffizienz.

Der Energieverbrauch eines Gebäudes besteht zu einem großen Teil aus der Energie, die zum Heizen der Räume benötigt wird. Hier fallen etwa 70% des Gesamtverbrauchs an.⁸ Eine Witterungsbereinigung ist im Zusammenhang mit Heizdaten besonders wichtig, weil das Wetter und die Außentemperatur einen großen Einfluss auf den Verbrauch einer Heizanlage haben.

Ein Beispiel für ein Szenario im betrieblichen Umfeld, das zeigt, warum eine Witterungsbereinigung sinnvoll ist, lässt sich wie folgt vorstellen: Eine elektrische Heizanlage in einem Gebäude wurde auf Anraten des Energiemanagers durch ein effizienteres Modell ausgetauscht. Der folgende Winter wird nun kälter als der vorherige. Der Vorgesetzte des Energiemanagers sieht auf seiner Endabrechnung einen höheren Stromverbrauch und somit größere Kosten als im vergangenen Jahr und stellt den Energiemanager zur Rede. Er teilt ihm mit, dass die Anschaffung der neuen Heizanlage keinerlei Effekt hatte und sogar noch zu einer Verschlechterung der Situation geführt hat. Der Energiemanager erklärt, dass ein einfacher Vergleich des Stromverbrauchs beider Jahre alleine nicht sinnvoll ist, sondern auch die äußeren Bedingungen mit in die Berechnung einbezogen werden müssen. Er führt eine Witterungsbereinigung durch, bei der die Stromverbräuche um die Gradtagzahlen der Jahre (und somit die klimatischen Verhältnisse) bereinigt werden. Hiermit ist er in der Lage, zu belegen, dass der Verbrauch im zweiten Jahr bei den gleichen Außentemperaturen wie im ersten Jahr deutlich geringer gewesen wäre und der Austausch der Anlage folglich doch sinnvoll war. Dies beruhigt den Vorgesetzten, und er kann anschaulich erkennen, wie der Austausch der Anlage sich in den kommenden Jahren amortisieren wird.

Dieses Szenario lässt sich noch weiterdenken, indem man etwas weiter vorne beginnt. Durch eine einfache Korrelationsanalyse zwischen Außentemperatur und Stromverbrauch

⁸vgl. Peter Hofer 2013, S. 1 (Standort Schweiz, 2002).

ist es dem Energiemanager möglich, einen besseren Überblick über seine Verbraucher zu erlangen. So könnte es etwa in obigem Beispiel sein, dass der Energiemanager sich den Verbrauch und die Außentemperatur in einem Streudiagramm angesehen und dabei erkannt hat, dass der Verbrauch unter einer gewissen Temperaturgrenze deutlich ansteigt. Dies eröffnete ihm den Rückschluss, dass eine Anlage, die mit der Heizung des Gebäude befasst ist, einen großen Einfluss auf den Gesamtverbrauch hat. Mit diesem Wissen könnte er weitere Untersuchungen anstellen und schließlich den Austausch der Anlage vorschlagen.

Doch es gibt noch weitere Gründe, warum eine Witterungsbereinigung sinnvoll oder sogar nötig sein kann. Wenn Projektionen über den zukünftigen Energieverbrauch angestellt werden sollen, ist es wichtig, die Jahreswerte witterungsbereinigt zu betrachten. Wenn dieser Schritt weggelassen wird, kann es zu dem Fall kommen, dass ein besonders heißes oder kaltes Jahr zu verzerrten Ergebnissen in der Projektion führt. Mit einer Witterungsbereinigung kann sichergestellt werden, dass ein langjähriger Trend und nicht nur kurzfristige Schwankungen in die Projektion einfließen.⁹

Ein weiteres Szenario, für das eine Witterungsbereinigung wichtig ist, ist der Vergleich mehrerer Gebäude. Soll die Energieeffizienz von Gebäuden an unterschiedlichen Standorten miteinander verglichen werden, muss der Verbrauch der Gebäude unter anderem um Grundfläche und Gebäudetyp, aber auch um die äußeren klimatischen Bedingungen, bereinigt werden, um tatsächlich Vergleichbarkeit zu schaffen.¹⁰

Für Stromdaten ist eine Witterungsbereinigung bisher eher unüblich, doch auch hier bietet sie sich an. So haben in Bürogebäuden Beleuchtung, elektrische Heizung und Lüftung einen Anteil von 32 Prozent am Gesamtstromverbrauch¹¹ und eine Bereinigung um Klimafaktoren, die diese Verbraucher beeinflussen können, ist auch hier sinnvoll. Wichtig hierbei ist es, dem Nutzer die Möglichkeit an die Hand zu geben, eine Witterungsbereinigung eigenständig durchzuführen. Wird die Interpretation der Ergebnisse weiterhin allein dem Nutzer überlassen, ist es möglich, dass er ein Gebäude, bei dem die Heizung kaum einen

⁹vgl. Peter Hofer 2013, 1f.

¹⁰vgl. Peter Hofer 2013, Anhang: S.1.

¹¹vgl. Weber 2002, S. 120.

Einfluss auf den Gesamtverbrauch hat, um Kühlgradtage bereinigt. Hierbei werden die gleichen Ergebnisse anfallen wie bei einer sinnvollen Bereinigung, diese werden aber nicht sinnvoll nutzbar sein. In dieser Situation ist es nun denkbar, dass das System den Nutzer zu einer Überprüfung der Ergebnisse auf ihre Sinnhaftigkeit auffordert.

Bereits der erste Schritt, das Aufzeigen der Korrelation zwischen Temperatur und Stromverbrauch, kann Gebäudemanagern einen interessanten Einblick in die Verbrauchsdaten und ihre Zusammenhänge geben. Die Witterungsbereinigung selbst hilft dann beim Schaffen von Vergleichbarkeit zwischen unterschiedlichen Jahren.

1.2 Zielstellung

Das Ziel der vorliegenden Arbeit ist die Entwicklung eines Konzepts zur Korrelationsanalyse von Strom- und Wetterdaten und zu ihrer Witterungsbereinigung. Sie gliedert sich dabei in zwei Teile: einen theoretischen Teil, der Stromdaten grundsätzlich auf Korrelation untersucht, und einen zweiten Teil, der die Entwicklung eines automatisierten Systems beschreibt. Dieses automatisierte System soll in ein bestehendes Energie Metering und Monitoring System eingebunden werden. Das System enthält bereits Informationen über den Stromverbrauch, nicht aber über die äußeren Faktoren, die einen Einfluss darauf haben können. Im Laufe der Arbeit soll geklärt werden, ob - und wenn ja wie - eine Witterungsanalyse und -bereinigung auch für Stromdaten durchgeführt werden kann. Eine weitere Aufgabe wird es sein, Nutzern des Systems ein einfaches Tool an die Hand zu geben, solch eine Analyse durchzuführen.

Dabei steht diese Arbeit unter drei Leitfragen, die im Laufe der Arbeit beantwortet werden. Die erste Leitfrage lautet „Inwiefern gibt es eine Korrelation von Strom- und Temperaturdaten und welche Unterschiede gibt es zur Korrelation bei Heizdaten?“. Ihr widmet sich vor allem das Kapitel 5. Die nächste Leitfrage lautet „Wie kann der Zusammenhang zwischen Temperatur und Stromverbrauch für (unerfahrene) Nutzer sichtbar gemacht werden?“ und die letzte Frage, die beantwortet werden soll, lautet „Wie kann eine Witterungsbereinigung

von Stromdaten umgesetzt werden?“. Die beiden letzten Fragen werden in den Kapiteln 6 und 7 beantwortet.

1.3 Aufbau der Arbeit

Im vorherigen Abschnitt wurde eine Einleitung in das Themengebiet Witterungsanalyse- und bereinigung von Stromdaten gegeben und der Grund dafür aufgezeigt, warum die Entwicklung eines Systems zur Umsetzung hiervon sinnvoll ist.

Im nächsten Teil wird der Bezug zum Fachbereich Umweltinformatik aufgezeigt. Dafür wird die Umweltinformatik in ihrer Definition vorgestellt und diese Arbeit in den Fachbereich eingeordnet.

Bei dem in dieser Arbeit entwickelten System handelt es sich um eine Ergänzung zur von smartB entwickelten Webanwendung. In Kapitel 3 werden das Unternehmen und der Funktionsumfang des Systems zum Zeitpunkt dieser Arbeit beschrieben.

Damit Stromdaten auf Korrelationen mit äußeren Einflüssen untersucht werden können, müssen historische Wetterdaten vorliegen. Kapitel 4 zeigt auf, welche Anforderungen an solche Daten gestellt werden. Weiterhin wird in dem Kapitel das Angebot des Deutschen Wetterdienstes aufgezeigt, da dies häufig auch zur Witterungsbereinigung für den Energieausweis genutzt wird. Um auch globale Datensätze untersuchen zu können, wird weiterhin eine Quelle für historische Klimadaten vorgestellt, die eine globale Abdeckung liefert. Die in diesem Kapitel gefundenen Daten werden für die Korrelationsanalyse im darauffolgenden Kapitel genutzt. Kapitel 7.2 beschreibt, warum diese Datensätze sich zwar für eine manuelle Analyse eignen, im späteren System aber nicht mehr zur Anwendung kommen können. Es folgt ein Abschnitt, in dem von wissenschaftlichen Institutionen veröffentlichte Verbrauchsdatensätze auf ihre Korrelation mit der Außentemperatur untersucht werden. Nachdem die Anforderungen an Datensätze zur Analyse aufgezeigt wurden, wird ein Überblick über alle öffentlichen Datensätze gegeben, die sich mit dem Stromverbrauch von Gebäuden beschäftigen. Die Datensätze werden anhand der gegebenen Kriterien bewertet

und für diese Arbeit passende Datensätze werden ausgewählt. Hierzu ist an die Arbeit ein umfassender Anhang angefügt, der alle veröffentlichten Verbrauchsdatensätze beschreibt und ihre Eignung für das System weiter untersucht. Es folgt eine Untersuchung von vier Datensätzen - zwei von ihnen aus den öffentlichen Datensätzen, einer aus den von smartB gesammelten Daten und ein vierter simulierter Datensatz, der ein Randszenario abdecken soll. Das Kapitel endet mit einer Zusammenfassung und Diskussion der Ergebnisse.

Nachdem die Grundlagen geklärt sind, beschreibt Kapitel 6 einen möglichen Ansatz, um eine Analyse und Bereinigung der Daten durchzuführen. In diesem Teil der Arbeit wird eine vertiefende Beschreibung des üblichen Vorgehens zur Witterungsbereinigung von Heizdaten gegeben und eine analoge Methode zur Witterungsbereinigung von Stromdaten entwickelt. Es werden weiterhin Methoden zur Visualisierung von Korrelationen in diesem Zusammenhang und zur automatisierten Korrelationsanalyse beschrieben.

Das nächste Kapitel beschäftigt sich mit der technischen Umsetzung des Systems. Ein Einblick in die genutzten Technologien wird gegeben und die Entwicklung der einzelnen Module wird beschrieben. Hierbei wird besonders herausgestellt, welche Anpassungen an das bestehende System gemacht werden mussten und wie sich der Prozess der Analyse und Visualisierung - insbesondere im Bezug auf die automatisierte Beschaffung von Wetterdaten - automatisieren lässt.

In Kapitel 8 wird das entwickelte System darauf untersucht, ob es der Anforderung, für Laien und Experten einfach nutzbar zu sein, gerecht wird. Hierfür werden die Ergebnisse eines Gesprächs mit einem erfahrenen Energiemanager beschrieben. Zusätzlich wird ein einfacher Usabilitytest mit einem Laien beschrieben und die Ergebnisse beider Untersuchungen werden diskutiert.

Die Arbeit schließt mit einem Fazit, das sich mit der Beantwortung der Leitfragen beschäftigt, und einem Ausblick auf weitere mögliche Ausbaustufen des Systems.

2 Bezug zur Umweltinformatik

Diese Arbeit entstammt dem Fachbereich der Umweltinformatik. Dieser Fachbereich wird im Folgenden kurz vorgestellt und der Bezug des Ansatzes der Arbeit zur Umweltinformatik geklärt.

Die Umweltinformatik ist eine angewandte Informatik und versteht sich als interdisziplinäre Wissenschaft, die verschiedene wichtige Themenfelder zusammenbringt, um es Anwendern zu ermöglichen, nachhaltigere Entscheidungen zu treffen.¹

Die Themenbereiche, die in den Bereich der Umweltinformatik fallen, sind vielfältig. Sie reichen von der Simulation von Umweltsystemen über die Entwicklung von Datenbanksystemen wie etwa Gefahrstoffkatastern bis hin zur Entwicklung von geographischen Umweltinformationssystemen, die räumliche Daten abbilden können. Ziel der Anwendungen, die im Bereich der Umweltinformatik entwickelt werden, sind etwa die Verbreitung von Informationen, das Schaffen von Verständnis über Umweltsysteme und die Auswirkungen, die Menschen auf diese verursachen und die Erleichterungen von Vorgängen etwa im Umweltcontrolling.

In ihrem Aufsatz *Umweltinformatik*² stellen die Autoren sieben methodische Schwerpunkte der Umweltinformatik vor. Bei diesen handelt es sich um *Datenbanksysteme und Geographische Informationssysteme, Modellbildung und Simulation, Wissensbasierte Systeme, Benutzungsschnittstellen und Software-Ergonomie, Computergraphik und Visualisierung,*

¹vgl. Fuchs-Kittowski und Wohlgemuth 2011, S. 99,100.

²vgl. Page und Hilty 1995, 20ff.

Künstliche Neuronale Netze und Integration. Die vorliegende Arbeit nimmt vor allem Bezug auf die Themengebiete *Computergraphik und Visualisierung* und *Benutzungsschnittstellen und Software-Ergonomie*. Die Autoren beschreiben die Notwendigkeit, Systeme zu entwickeln, die komplizierte Umweltproblematiken und -informationen einfach aufbereiten und die Ergebnisse auch für „gelegentliche Endbenutzer“ entsprechend bereitstellen.³

Das von smartB entwickelte System versucht, das immer wichtiger werdende Themengebiet des Stromverbrauchs (und somit auch des CO₂-Ausstoßes) auf einfache Art sowohl für Unerfahrene als auch für Fachnutzer darzustellen. Im Rahmen dieser Arbeit soll das bestehende System um eine weitere Komponente ergänzt werden, die die Wetterdaten einbindet und mit bestehenden Daten in Verbindung setzt, diese einfach visualisiert und es so Nutzern ermöglicht, bessere Entscheidungen zu treffen.

³vgl. Page und Hilty 1995, S. 20.

3 Vorstellung der smartB Energy Management GmbH

Die smartB Energy Management GmbH wurde im November 2013 gegründet und hat ihren Sitz in Berlin. Das Unternehmen stellt eine Hard- und Softwarelösung bereit, mit der es Kunden möglich ist, ihren Stromverbrauch zu messen und genauere Informationen über die einzelnen Verbraucher zu erhalten. Hierfür wird ein Smart Meter in den Verteilerkasten des Kunden eingebaut. Dieser Meter misst mit einer Auflösung von 50 Hz die anfallende Spannung und Stromstärke. Aus diesen Messwerten werden Wirkleistung und Blindleistung bestimmt.

Bei den Smart Metern, mit denen smartB die Gebäude ausstattet, handelt es sich um Meter die mit Non Intrusive Load Monitoring (NILM) Methoden Messungen durchführen. *Non Intrusive* beschreibt hier die Tatsache, dass die bestehende Stromverbindung nicht unterbrochen werden muss, sondern das Messgerät einfach in das bestehende System integriert werden kann.¹

Die von den Smart Metern gemessenen Werte werden im Folgenden auf unterschiedliche Arten genutzt. Zunächst werden sie mit Methoden des Machine Learnings analysiert und eine Disaggregation des Gesamtverbrauchs auf einzelne Verbraucher wird durchgeführt. Die Disaggregation basiert darauf, dass Stromverbraucher Lastprofile haben, die sich voneinander unterscheiden. So können Ein- und Ausschaltvorgänge im System erkannt und Anhand der Unterschiede im Stromverbrauch zu bekannten Geräten oder Gerätegrup-

¹Für eine genauere Erklärung von NILM vergleiche auch Paris 2006.

pen zugeordnet werden. Mit Hilfe dieser Methode ist es möglich, einzelne Verbraucher zu erkennen, ohne jeden Verbraucher einzeln bemessen zu müssen, was viel Geld kosten würde. Die so gefundenen Verbraucher werden dann als Representational State Transfer (REST)-Application Programming Interface (API) bereitgestellt.

Als Benutzungsoberfläche zu diesen Daten wird eine webbasierte Anwendung bereitgestellt. Mit dieser Anwendung können Nutzer die disaggregierten Werte einsehen und so mehr Energietransparenz erlangen. Das System lässt sich zum aktuellen Zeitpunkt in drei weitere große Teilbereiche einteilen: Den Vergleich mit Benchmarks, die Analyse des Verbrauchs mit unterschiedlichen zeitlichen Auflösungen und die automatische Erstellung von Benachrichtigungen bei vom System entdeckten Anomalien.

Dem Nutzer werden wichtige Benchmarks - zum Beispiel Verbrauch pro Jahr - dargestellt und mit entsprechenden Vorgaben - etwa den Werten aus der EnEV verglichen. Des Weiteren werden die gemessenen Verbrauchsdaten in jährlicher, monatlicher, wöchentlicher und täglicher (bis zum aktuellen Zeitpunkt) Auflösung angezeigt, der Vergleich zweier Messreihen wird ermöglicht. Der letzte Teil, die Benachrichtigungen, ermöglichen es zum Beispiel, Nutzer bei ungewöhnlich hohem Verbrauch am Wochenende oder zu hoher Grundlast zu informieren, so dass sie direkt korrigierende Maßnahmen treffen und etwaige Einsparpotentiale in ihren Gebäuden umsetzen können.

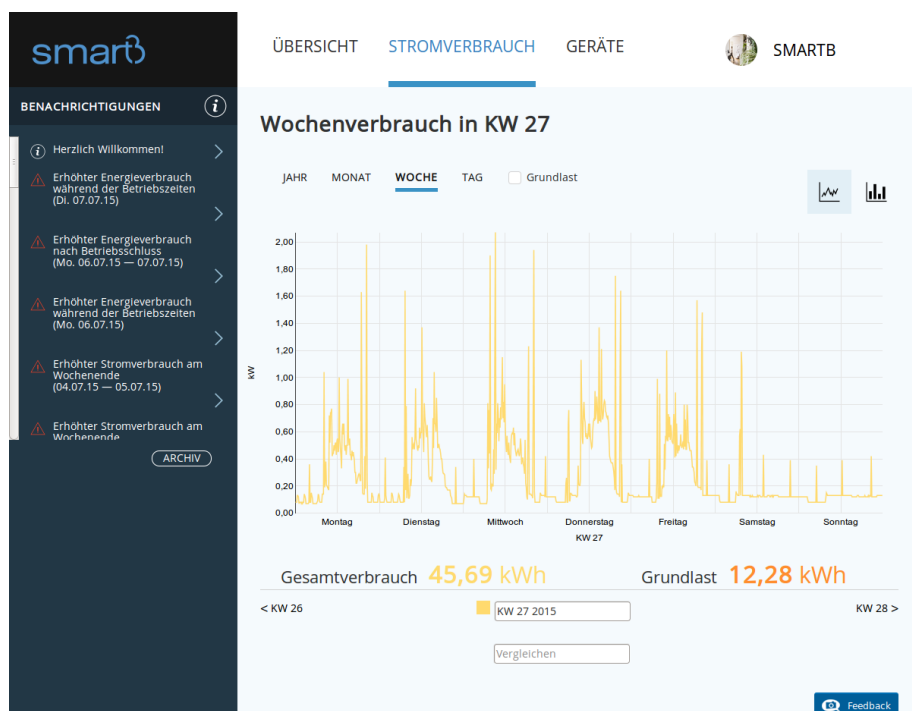


Abbildung 3.1: Das smartB Web Interface

4 Wetterdaten zur Datenanalyse

4.1 Anforderungen an die Daten

Im nächsten Kapitel werden öffentliche Verbrauchsdatensätze beschrieben und auf Korrelation mit Wetterdaten untersucht. Um dies zu ermöglichen, mussten Wetterdaten gefunden werden, die sich zur Analyse eignen. Dabei mussten einige Anforderungen an die Daten gestellt werden. Die Wetterdaten sollten kostenlos verfügbar sein, eine Zeitspanne von mehreren Jahren umfassen und qualitätsgesichert sein. Inhaltlich sollten sie die Außentemperatur und - wenn möglich - auch die Sonnenscheindauer abbilden.¹ Dies ermöglicht es später, Korrelationen mit Temperatur (also Heizung, Kühlung) und Lichtverhältnissen (also Innenbeleuchtung) anzustellen. Im Folgenden werden eine Datenquelle für Deutschland - die Daten des Deutschen Wetterdienstes - und eine internationale Datenquelle - die Integrated Surface Hourly (ISH) Datensätze der National Oceanic and Atmospheric Administration (NOAA) - vorgestellt.

¹Die Sonnenscheindauer wurde später jedoch nicht verwendet, da schon die Temperaturdaten relativ gut mit der Sonnenscheindauer korrelieren und die Betrachtung eines weiteren Faktors den Rahmen der Arbeit zu sehr vergrößert hätte.

4.2 Daten des Deutschen Wetterdienstes

4.2.1 Historische Wetterdaten

Der Deutsche Wetterdienst (DWD) stellt als Bundesoberbehörde eine Vielzahl von Daten kostenlos zur Verfügung. Neben den unten genannten Klimafaktoren werden auch historische Rohdaten zur Verfügung gestellt. Die Temperaturdaten von 613 Stationen² lassen sich per File Transfer Protocol (FTP) abrufen.³ Laut der Datei,⁴ die die Metainformationen der Stationen enthält, gehen die ältesten Aufzeichnungen hier bis 1893 zurück. Die Nutzungsbedingungen verlangen nur die Namensnennung⁵ und die Werte der Stationen sind qualitätsgesichert. Diese Daten erfüllen also alle Voraussetzungen und können daher im folgenden Kapitel zur Analyse von Gebäuden, die sich in Deutschland befinden, genutzt werden.

4.2.2 Klimafaktoren

Zur Erstellung des Energieausweises stellt der Deutsche Wetterdienst (DWD) die sogenannten Klimafaktoren⁶ bereit. Hierbei handelt es sich um dimensionslose Werte, die den Vergleich zwischen Regionen oder Jahren ermöglichen sollen. Zur Berechnung der Klimafaktoren werden die Werte aus der *TRY Messreihe* (siehe nächster Abschnitt) genutzt. Die Grundlage der Berechnung bildet der Vergleich mit der TRY-Datenreihe für den Messstandort Potsdam (früher Würzburg). Die Klimafaktoren beschreiben immer einen Zeitraum von 12 Monaten, wobei dieser gleitend ist. Es gibt also beispielsweise Klimafaktoren für die Zeiträume 01.06.2013-31.05.2014 und 01.07.2013-30.06.2014. Zur Berechnung werden die Gradtagzahlen (siehe Erklärung zu Gradtagzahlen Kapitel 6.3.1) des Standorts Potsdam aus dem Testreferenzjahresdatensatz durch die Gradtagzahl des entsprechenden

²Stand: 2015-08-10

³<ftp://ftp-cdc.dwd.de/pub/CDC/>

⁴ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/hourly/air_temperature/historical/TU_Stundenwerte_Beschreibung_Stationen.txt

⁵ftp://ftp-cdc.dwd.de/pub/CDC/Nutzungsbedingungen_German.pdf

⁶<http://www.dwd.de/klimafaktoren>

Jahres vor Ort geteilt. Dies ermöglicht es, alle Jahre und Standorte miteinander zu vergleichen.

4.2.3 TRY - Testreferenzjahre

Der TRY (Test Reference Year)-Datensatz ⁷ umfasst stündliche und über die Jahre 1988 bis 2007 gemittelte meteorologische Daten für 15 Klimaregionen. Für diese Klimaregionen wird jeweils ein charakteristischer Witterungsverlauf abgebildet, der möglichst nah an den Durchschnittswerten der untersuchten 20 Jahre ist. Zusätzlich zu den Durchschnittswerten werden außerdem jeweils zwei Extremreihen veröffentlicht. Diese Jahreswitterungsverläufe beschreiben besonders heiße und besonders kalte Jahre. Mit Hilfe dieser Daten kann eine Witterungsbereinigung durchgeführt werden, wenn keine genaueren Messwerte für den Standort des Gebäudes verfügbar sind.

4.3 Daten der NOAA

Für Datensätze, die sich außerhalb von Deutschland befinden, können die Daten des Deutschen Wetterdienstes nicht benutzt werden, da diese nur Deutschland umfassen. Die US-Amerikanische Wetterbehörde NOAA veröffentlicht aber seit 1988 in der sogenannten Integrated Surface Data (ISD) historische und aktuelle Klimadaten von über 20.000 Messstationen.⁹ Bei vielen dieser Messstationen handelt es sich um Flughäfen, die Message d'observation météorologique régulière pour l'aviation (METAR) Berichte veröffentli-

⁷<http://www.dwd.de/TRY>

⁸Quelle: http://www.dwd.de/bvbw/generator/DWDWWW/Content/0effentlichkeit/KU/KU1/KU12/Klimadaten/TRY/TRY__Karte,templateId=raw,property=publicationFile.pdf/TRY_Karte.pdf,
abgerufen 2015-08-05

⁹Neal Lott 2008.

chen.¹⁰ Diese Daten lassen sich jahresweise per FTP von der Website¹¹ der NOAA herunterladen.

Die Qualität der Daten ist allerdings sehr schwankend. Gerade in den USA ist die Abdeckung recht gut, für Deutschland umfasst der Datensatz aber nur noch zwei Stationen, die aktuelle Messwerte liefern. Für das Gebäude in Frankreich, das im kommenden Kapitel untersucht wird, lagen glücklicherweise Daten einer nahegelegenen Messstation vor, so dass diese Daten genutzt werden konnten.

Die Daten liegen in einem speziellen Format vor, das teilweise dem METAR Format folgt. Um die Daten mit Python zu analysieren, kann die unter der Open Source MIT-Lizenz veröffentlichte Python-Bibliothek *ish-parser*¹² genutzt werden.

¹⁰Hierbei handelt es sich um ein standardisiertes Format, das die aktuellen Werte und eine Vorhersage für die nächsten zwei Stunden enthält. Vergleiche auch <http://www.wetterklima.de/flug/metar/Metarcode.htm>, abgerufen 2015-08-31

¹¹<ftp://ftp.ncdc.noaa.gov/pub/data/noaa/>

¹²https://github.com/haydenth/ish_parser

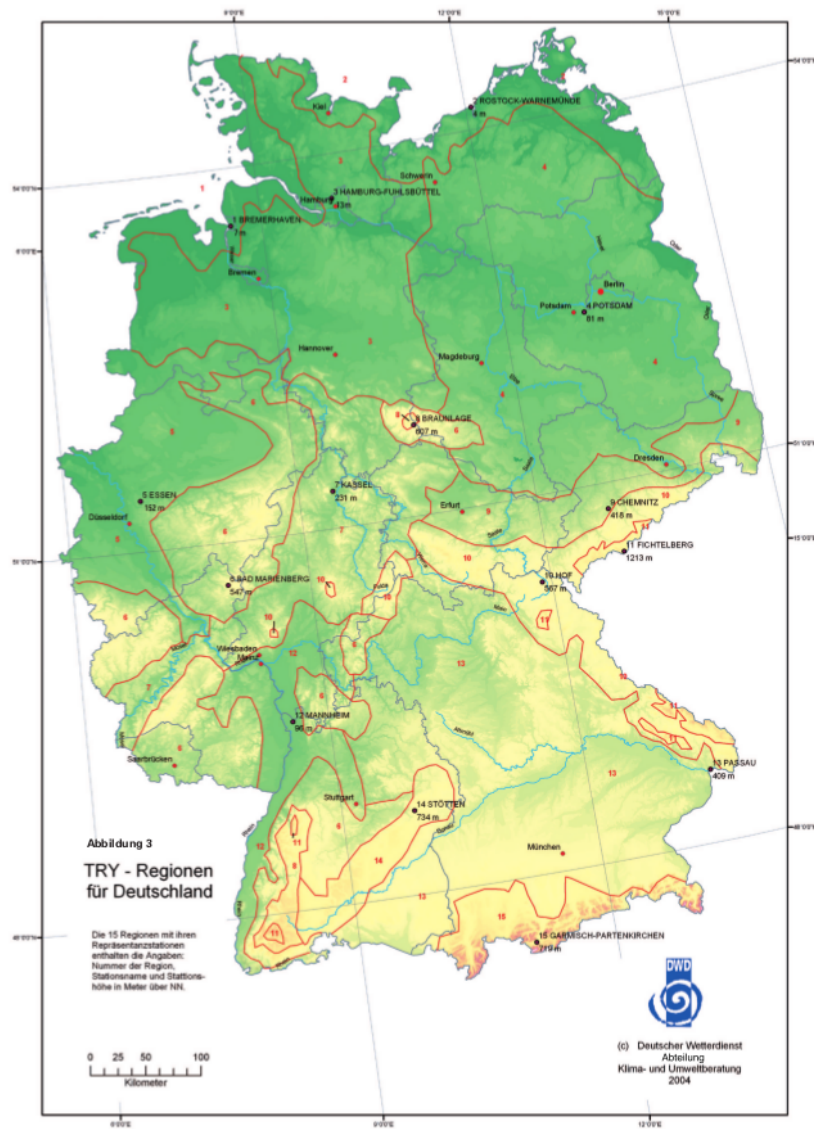


Abbildung 4.1: Die TRY-Regionen⁸

5 Korrelationsanalyse von Beispieldatensätzen

Um Stromdaten um Temperaturschwankungen bereinigen zu können, muss zuerst untersucht werden, ob es für Stromdaten überhaupt einen Zusammenhang zwischen dem Verbrauch und der Außentemperatur gibt. Dafür werden im folgenden Kapitel die Verbrauchsprofile von drei Standorten mit den Klimadaten für die entsprechenden Regionen verglichen und Korrelationen aufgezeigt. Außerdem wird ein potentielles Lastmuster aufgezeigt und ein simulierter Datensatz untersucht.

Die Analysen in diesem Kapitel liefern auch die Grundlage für die Visualisierungen, die später im Laufe der Entwicklung erstellt und den Nutzern zugänglich gemacht werden.

5.1 Vorgehen

5.1.1 Datenquellen

Um zu überprüfen, ob eine Korrelation zwischen Außentemperatur und Stromverbrauch von Gebäuden vorliegt, werden zunächst Daten zum Stromverbrauch eines Gebäudes und die Wetterdaten aus der entsprechenden Region benötigt. Dabei gibt es einige Anforderungen, die an die Stromdaten gestellt werden. Die Daten sollten einen langen Zeitraum umfassen (idealerweise mehr als ein Jahr, um einen ganzen Jahresklimazyklus zu durchlaufen), möglichst wenig Datenverlust beinhalten und mit einem möglichst genauen Standort

versehen sein, um die passenden Wetterdaten finden zu können. Um zum Anwendungsgebiet des Systems von smartB zu passen, sollte es sich weiterhin im Idealfall um Stromdaten von gewerblichen Gebäuden und nicht von Privathaushalten handeln.

Zum Zeitpunkt dieser Arbeit gab es 12 öffentliche Datensätze, die näher betrachtet wurden. Zehn dieser Datensätze beschreiben private Haushalte, die restlichen zwei bilden Universitätsgebäude ab. Alle Datensätze sind in Tabelle 5.1 zusammengefasst. Die Tabelle basiert auf Daten der Website von Oliver Parson,¹ dem GREEND-Paper² und eigener Recherche. Um den Rechercheprozess für kommende Arbeiten zu vereinfachen, wurden die Ergebnisse auf github³ veröffentlicht, wo sie einfach einsehbar und auch einfach für Dritte änderbar und aktualisierbar sind. Eine detaillierte Beschreibung mit Angabe der zugehörigen wissenschaftlichen Veröffentlichungen befindet sich zur besseren Lesbarkeit in Anhang A.

Bei den Datensätzen, die am besten zu den geforderten Kriterien passen, handelt es sich um den IHEPCDS-Datensatz⁴ und die Daten aus dem Pecan Street Projekt. Beide Datensätze beinhalten Daten aus mehreren Jahren und lassen somit einen guten Schluss auf die saisonalen Schwankungen zu. Es handelt sich aber um Daten aus privaten Haushalten - beim Anwendungsbereich des zu entwickelnden Algorithmusses handelt es sich um gewerbliche Gebäude. Hierfür stand ein weiterer Datensatz von smartB zur Verfügung, der ein Bürogebäude in Berlin über den Zeitraum von etwa einem halben Jahr beschrieb.

¹<http://blog.oliverparson.co.uk/2013/06/comparison-of-public-disaggregation.html>, abgerufen 2015-08-03

²Monacchi u. a. 2014.

³<https://github.com/k-nut/nilm-datasets/blob/master/nilm-datasets.csv>

⁴Hébrail u. a. 2010.

Name	Standort	Gebäudeart	Gebäudeanzahl	Max. Messdauer (Tage)
AMPds	Region Vancouver, Kanada	Haushalt	1	730
BERDS	Berkley, USA	Universität	3	7
COMBED	Neu Delhi, Indien	Universität	5 (?)	30
ECO	Thun, Schweiz	Haushalt	6	244
GREEND	Österreich/Italien	Haushalt	9	310
iAWE	Neu Delhi, Indien	Haushalt	1	73
IHEPCDS	Sceaux, Frankreich	Haushalt	1	1456
Pecan Street	{Austin, San Diego, Boulder}, USA	Haushalt	725	1281
REDD	Boston, USA	Haushalt	10	19
Smart*	Massachussets, USA	Haushalt	3	90
SustData	Madeira, Portugal	Haushalt	50	504
UK Dale	Großbritannien	Haushalt	5	499

Tabelle 5.1: Öffentliche NILM-Datensätze

5.1.2 Methode zur Bewertung der Korrelation

Im folgenden Abschnitt werden die Gebäude vorgestellt und die Stromdaten gemeinsam mit den Temperaturdaten der Region dargestellt. Die Daten liegen in den meisten Fällen in mindestens täglicher, teils stündlicher Auflösung vor. Die Schwankungen, die sowohl Stromverbrauch als auch Außentemperatur an einem Tag aufweisen, sind relativ hoch, weshalb es sich anbietet, die Daten in einer größeren Auflösung zu betrachten. Besonders eignet sich hier der wöchentliche Vergleich, da dieser die Schwankungen in den Daten reduziert und eine relativ große Menge an Datenpunkten liefert. Alle Datensätze wurden folglich durch Bilden des Mittelwertes der täglichen bzw. stündlichen Daten auf eine

wöchentliche Auflösung umgerechnet. Diese wöchentlichen Daten werden im Folgenden als Liniendiagramme dargestellt, um einen Eindruck vom saisonalen Verlauf zu erhalten. Weiterhin bietet ein Streudiagramm einen guten Überblick über die Korrelation. Außerdem wird der Wert der Pearson-Korrelation ⁵ angegeben. Die Analysen wurden mit der Python-Bibliothek Pandas durchgeführt (mehr Informationen hierzu in Kapitel 7.1).

5.2 Die Gebäude

5.2.1 Haushalte in Austin, USA

Bei dem von der Pecan Street Inc veröffentlichten Datensatz handelt es sich um einen der größten verfügbaren Datensätze seiner Art. Insgesamt umfasst er 1318 Haushalte. Nicht alle dieser Haushalte waren aber auch mit smart-Metern versehen, die einen Rückschluss auf den täglichen Stromverbrauch erlaubten. Die Daten werden als PostgreSQL-Datenbank⁶ bereitgestellt, auf die Anfragen ausgeführt werden können. Eine Anfrage an die Metadata Datenbank ergab, dass insgesamt 724 Haushalte über Messwerte von smart-Metern verfügten. Diese Haushalte wurden in der gleichen Anfrage nach der Dauer der Messzeit geordnet und die drei Haushalte mit der längsten Messdauer ausgewählt. Bei diesen Haushalten handelt es sich um die Haushalte mit den IDs 585, 744 und 5129 mit einer Messdauer von 1281, 1280 und 1202 Tagen. Alle Haushalte befinden sich in Austin, Texas.

Die Klimadaten wurden ebenfalls vom Pecan Street Projekt bereitgestellt und per Anfrage an die Datenbank heruntergeladen. Die SQL-Anfrage, mit der die Gebäude gefunden wurden, die den längsten Messzeitraum umfassen, findet sich in Beispiel 1.

⁵Diese Art der Korrelationsberechnung bietet sich für normalverteilte Daten mit linearen Zusammenhängen an, wovon bei den hier betrachteten Werte auszugehen ist vgl. Mittag 2014, 118f

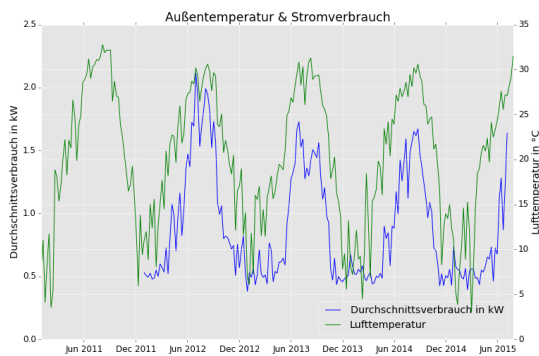
⁶<http://www.postgresql.org/>

```
select *, (egauge_max_time - egauge_min_time) as egauge_duration
  from university.metadataata
 where egauge_min_time is not null
 order by egauge_duration desc
 limit 3;
;
```

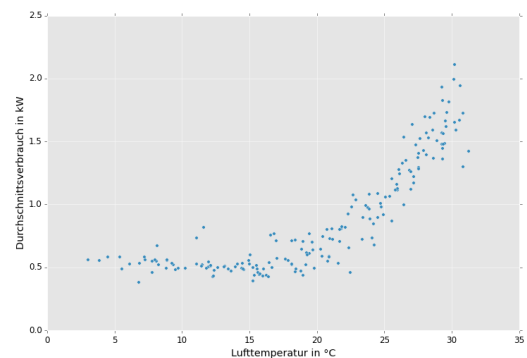
Listing 1: SQL Anfrage an Pecan Street Metadatenbank

Im Streudiagramm für diesen Datensatz wird gut sichtbar, dass es einen Zusammenhang von Stromverbrauch und Außentemperatur gibt. Dieser könnte zum Beispiel von einer Klimaanlage kommen. So bleibt der Stromverbrauch im Bereich von 2 bis 20 Grad Celsius relativ gleichbleibend bei etwa 0,5 kW. Ab 20 Grad steigt der Stromverbrauch linear, bis auf ca. 1,8kW bei 32 Grad an.

Die Berechnung des pearsonschen Korrelationskoeffizienten ergibt für diesen Haushalt einen Wert von 0.85 was einer sehr guten Korrelation entspricht.



(a) Stromverbrauch und Außentemperatur



(b) Streudiagramm

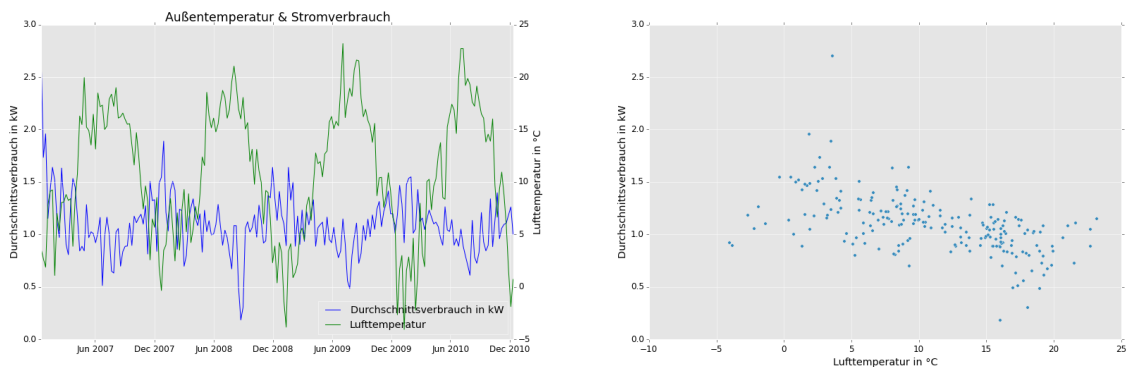
Abbildung 5.1: Diagramme für Pecan Street Id 585

5.2.2 Haushalt in Sceaux

Der Haushalt aus dem IHEPCDS Datensatz⁷ befindet sich in Sceaux, südlich von Paris.⁸ Daten sind für den Zeitraum von Dezember 2006 bis November 2010 verfügbar. Da es sich hierbei um einen Haushalt in Frankreich handelt, für den es keine Daten des DWD gibt, wurden ISH Daten (Siehe hierzu Kapitel 4.3) des nahe gelegenen Flugplatzes Villacoublay genutzt. Zu finden sind diese Daten auf dem FTP-Server⁹ der NOAA unter der Stations-ID 071470 bzw. LFPV.

Ein Blick auf die Daten zeigt, dass bei diesem Haushalt eine negative Korrelation vorliegt - sinkt die Temperatur, steigt der Stromverbrauch. Dies spricht dafür, dass der Haushalt viel Strom zum Heizen benötigt. Dies könnte zum Beispiel eine elektrische Heizung sein.

Der Korrelationskoeffizient beträgt $-0,54$, ist also deutlich schwächer als der vorhergehende Koeffizient.



(a) Stromverbrauch und Außentemperatur

(b) Streudiagramm

Abbildung 5.2: Diagramme für Frankreich

⁷<http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption>

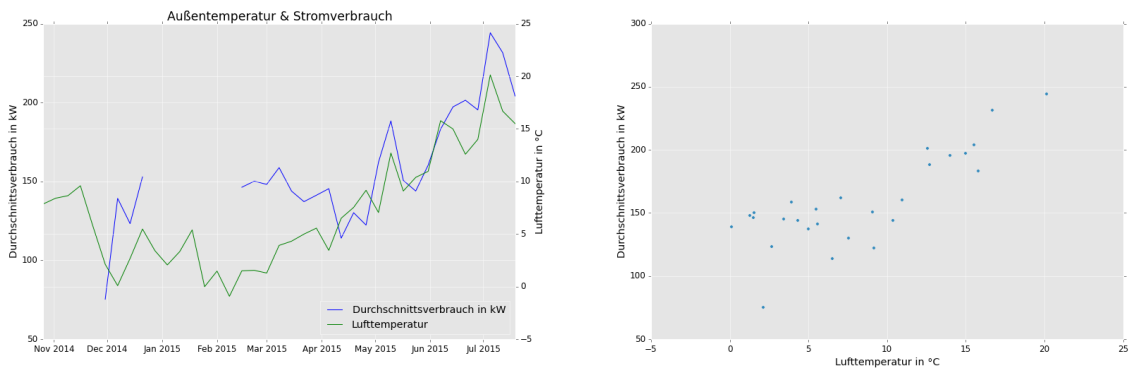
⁸persönliche Korrespondenz mit Autoren vom 05.07.2015

⁹<ftp://ftp.ncdc.noaa.gov/pub/data/noaa/2006/>

5.2.3 Bürogebäude in Berlin

Das Bürogebäude in Berlin kommt den Anforderungen für das zu entwickelnde System am nächsten (Bürogebäude in Deutschland), umfasst aber auch nur den kürzesten Zeitraum. Im Januar gab es einen Ausfall des smart-Meters, so dass für diesen Zeitraum keine Daten zur Verfügung stehen.

Auch hier (Abbildung 5.3) ist zu erkennen, dass ein Anstieg in der Temperatur im Laufe des Jahres mit einem Anstieg des Stromverbrauchs einhergeht. Der Korrelationskoeffizient liegt hier bei 0,80. Er ist aber, da es sich nur um Daten für etwas mehr als ein halbes Jahr handelt, nicht so verlässlich wie der für die anderen Datensätze.



(a) Stromverbrauch und Außentemperatur

(b) Streudiagramm

Abbildung 5.3: Diagramme für Berlin

5.2.4 Simulierter Datensatz

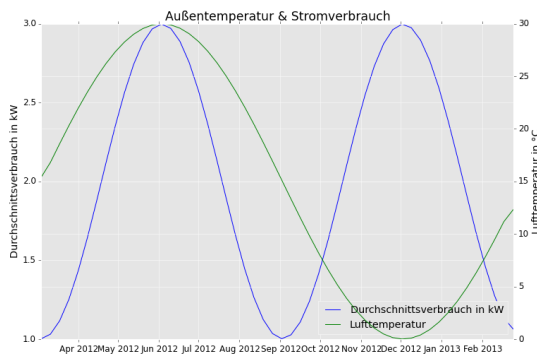
Es wäre vorstellbar, dass ein Bürogebäude in der Realität noch ein anderes Muster aufzeigen würde: Nämlich einen Anstieg des Stromverbrauchs bei sehr tiefen Temperaturen (zum Beispiel aufgrund einer elektrischen Heizung) und bei sehr hohen Temperaturen (zum Beispiel aufgrund einer Klimaanlage). Dies würde das Erkennen einer Korrelation schwieriger machen, ist aber gerade im betrieblichen Umfeld vorstellbar und muss daher im Folgenden

berücksichtigt werden. Hierfür wird ein stark vereinfachter Datensatz angenähert mit:

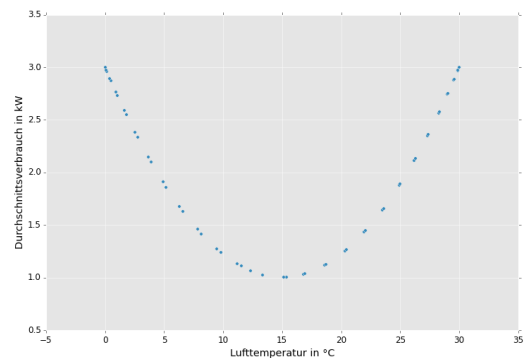
$$Verbrauch = \sin\left(\frac{4\pi}{365} \times x + \frac{3\pi}{2}\right) + 2 \quad (5.2.1)$$

$$Temperatur = 15 \sin\left(\frac{2\pi}{365}x\right) + 15 \quad (5.2.2)$$

Betrachtet man hier die normale lineare Korrelation über den gesamten Zeitraum, erhält man einen Wert, der ungefähr 0 entspricht, da die Werte sich gegenseitig aufheben. Betrachtet man aber nur die Werte über 23 bzw. unter 10 Grad, ergibt sich eine Korrelation von 0,998 bzw. -0,993. Auch ein Fall wie dieser soll folglich vom zu entwickelnden System erkannt werden.



(a) Stromverbrauch und Außentemperatur



(b) Streudiagramm

Abbildung 5.4: Diagramme für simulierten Datensatz

5.3 Diskussion der Ergebnisse

Eine Zusammenfassung der Ergebnisse ist in Tabelle 5.2 zu finden. Für die Daten aus Texas ist eine klare Korrelation zwischen der Temperatur und dem Stromverbrauch zu

erkennen. Dies liegt mit hoher Wahrscheinlichkeit an den genutzten Klimaanlage.¹⁰ Auch das Bürogebäude in Berlin zeigt eine positive Korrelation, hier ist aber zu bemängeln, dass nur ein kurzer Zeitraum erfasst wurde und der Großteil des Winters nicht gemessen wurde. Der Haushalt in Frankreich schließlich unterscheidet sich von den anderen Datensätzen, da es sich hierbei um eine negative Korrelation handelt.

Insgesamt lässt sich sagen, dass es eine recht große Korrelation zwischen Stromverbrauch und Temperatur geben kann und somit eine genauere Untersuchung und Witterungsreinigung des Stromverbrauchs im System von smartB sinnvoll ist.

Datensatz	Wöchentlich	Monatlich
Pecan-585	0.85	0.92
Pecan-744	0.86	0.93
Pecan-5129	0.77	0.86
Frankreich	-0.54	-0.74
Berlin	0.80	0.70

Tabelle 5.2: Korrelation der Beispieldatensätze

¹⁰Der hohe Stromverbrauch der Klimaanlage ist auch in den disaggregierten Daten zu sehen, die zusätzlich zu den Summen veröffentlicht wurden und bestätigt somit diese Hypothese

6 Konzeptionierung des Systems

Auf Basis der im vorherigen Kapitel gefundenen Ergebnisse soll nun ein neues Modul für das bestehende System entwickelt werden. Im vorherigen Kapitel wurde deutlich, dass nicht nur bei Heizdaten, sondern auch beim Stromverbrauch eine Korrelation zwischen Verbrauch und Außentemperatur auftreten kann. Auf Basis dieser Erkenntnis soll nun ein System entwickelt werden, das es Nutzern ermöglicht, selbstständig den Stromverbrauch ihrer Gebäude auf Korrelation zur Außentemperatur zu untersuchen und eine Witterungsbereinigung durchzuführen. Hierfür wird ein dreiteiliges Modul konzipiert, das in das Webinterface von smartB eingebunden werden soll. Bei den drei Teilen handelt es sich um die Visualisierung der Daten, die automatisierte Korrelationsanalyse und die Witterungsbereinigung. Diese drei Teile und die zu Grunde liegenden Konzepte werden im Folgenden vorgestellt.

6.1 Visualisierung

Der erste Teil des neuen Moduls soll die Visualisierung der Daten ermöglichen. Im vorherigen Kapitel wurden bereits einige Diagramme gezeigt, die eine Korrelation aufzeigen. In dem neuen System werden zur Visualisierung zwei unterschiedliche Diagramme dargestellt. Ein Balken-/Liniendiagramm, das die Außentemperatur als Linie und den monatlichen Stromverbrauch als Balken darstellt und ein Streudiagramm, das die Korrelation zwischen den beiden Faktoren sichtbar macht.

Bei dem ersten Diagramm, das dem Benutzer angezeigt wird, handelt es sich um ein gemischtes Linien- und Balkendiagramm. Diese Art der Darstellung wurde gewählt, weil die Nutzer bereits aus dem bestehenden Modul „Stromverbrauch“ die Darstellung ihres monatlichen Stromverbrauchs in Form des Balkendiagramms kennen. Zusätzlich zum Stromverbrauch wird in dem Diagramm die mittlere monatliche Außentemperatur aufgezeigt. Die Temperatur wird als Linie dargestellt, da dies die für Temperaturen übliche Art der Darstellung ist und Nutzer diese Formatierung gewöhnt sind. Ein Beispiel für diese Darstellungsart ist in Abbildung 7.5 zu finden.

Als Auflösung wurde aus mehreren Gründen ein Monat gewählt. Einerseits soll dieses erste Diagramm den Nutzern ermöglichen, einen Überblick über die Art der Schwankung zu gewinnen, der sowohl Stromverbrauch als auch Temperatur über das Jahr hinweg unterliegen. Hier bietet sich die monatliche Auflösung an, da sie mit 12 Datenpunkten pro Jahr die Diagramme nicht zu breit werden lässt und trotzdem noch einen guten Rückschluss auf den Verbrauch liefert. Des Weiteren ist es Nutzern des Systems möglich, zusätzlich zu den von smartB gemessenen Werten historische Werte einzugeben. Diese Werte stammen meist von Stromrechnungen, die im betrieblichen Bereich monatlich gestellt werden. Aus diesem Grund ist auch die Eingabe der historischen Verbräuche in einer monatlichen Auflösung vorgegeben. Dadurch, dass auch in dem vorgestellten Diagramm monatliche Werte angezeigt werden, ist es möglich, die historischen Werte mit in das Diagramm aufzunehmen und so einen noch weiteren Blick in die Vergangenheit zu liefern und einen besseren Überblick zu ermöglichen.

Das zweite Diagramm zur Visualisierung ist ein Streudiagramm. In diesem wird mit einer wöchentlichen Auflösung die durchschnittliche Außentemperatur gegenüber dem durchschnittlichen Verbrauch angezeigt. Hierbei wird die Temperatur auf der waagerechten Achse und der Verbrauch auf der senkrechten Achse eingetragen. Diese Art der Darstellung ermöglicht es, Korrelationen einfach zu erkennen. Im Falle einer großen Korrelation, durch eine Klimaanlage etwa, kann der starke Anstieg des Verbrauchs ab einer bestimmten Temperatur leicht erkannt werden. Dies eröffnet den Nutzern dann einen Rückschluss auf potentielle Einsparungen. Es könnte etwa auch sein, dass ein Nutzer erkennt, dass es

bereits ab einer Temperatur von 16 Grad Celsius zu einem Anstieg kommt, die Klimaanlage also schon bei dieser Temperatur angeschaltet wird, obwohl es den Mitarbeitern genügen würde, sie erst ab 20 Grad zu einzuschalten. Eine andere Situation, in der diese Darstellung besonders nützlich ist, ist der Fall, dass sowohl bei niedrigen als auch bei hohen Temperaturen ein Anstieg des Stromverbrauchs zu messen ist. Würde nur die einfache, über den gesamten Zeitraum berechnete Korrelation angezeigt werden, würden sich die beiden Extreme verkürzen und keine Korrelation könnte erkannt werden. In dem Diagramm könnte in diesem Fall jedoch leicht erkannt werden, dass zwei unterschiedliche Arten von Korrelationen vorliegen.

Für diesen zweiten Graphen wurde die Auflösung von einer Woche gewählt. Dies hat mehrere Vorteile gegenüber der monatlichen oder täglichen Ansicht. Das Problem bei einer monatlichen Darstellung ist, dass diese unterschiedliche Längen haben und daher, um wirklich vergleichbar zu sein, um die Anzahl der Tage bereinigt werden müssten. Des Weiteren ist es für ein Streudiagramm gut, viele Datenpunkte zu haben, wofür die monatlichen Daten nicht ausreichen würden. Gegen eine tägliche Darstellung wurde sich entschieden, da es in dieser Auflösung zu starken Schwankungen sowohl in Temperatur als auch im Verbrauch kommt und dies ein zu starkes Rauschen erzeugt.

6.2 Korrelationsanalyse

Der zweite Teil des neu entwickelten Moduls ist ein System zur automatischen Korrelationsanalyse. Nachdem im ersten Teil, der Visualisierung, die Nutzer eigenständig einen intuitiven Einblick in die Korrelation der Daten finden konnten, versucht das System in diesem Schritt, automatisiert eine Korrelation zu finden.

Die Ergebnisse aus Kapitel 5 zeigen, dass auch bei Stromdaten eine Korrelation vorliegen kann, dass die Korrelation aber oft nicht so eindeutig ist wie bei Heizsystemen. Es gibt zwei große Unterschiede, die betrachtet werden müssen. So kann es bei Stromdaten nicht nur zu einer Korrelation in eine Richtung kommen, sondern es ist auch ein Szenario denkbar, bei dem aufgrund von einer elektrischen Heizung und einer Klimaanlage sowohl eine positive

als auch eine negative Korrelation auftritt. Um dieses Szenario behandeln zu können, wurde eine Methode zur Korrelationsanalyse entwickelt. Ein weiterer Unterschied zum Heizsystem ist, dass nicht der gesamte Verbrauch von der Außentemperatur abhängig ist. Bei einer Heizung ist dieser Zusammenhang direkt gegeben - wird es draußen kälter, steigt auch der Verbrauch der Heizung. Der Stromverbrauch wird jedoch nicht nur von Geräten beeinflusst, deren Verbrauch direkt von der Außentemperatur abhängig ist, weshalb die Korrelationen, die hier auftreten, geringer sind.

Um das erste beschriebene Problem der positiven und negativen Korrelation in einem Datensatz zu umgehen, wird zunächst die Korrelation über die Gesamtheit der Daten berechnet und angezeigt. Danach werden Korrelationen für eine Reihe von Werten berechnet, die über oder unter bestimmten Temperaturen liegen. Dies ermöglicht es, nur die Reihe an Werten zu untersuchen, die für eine Korrelation mit Klimatisierung oder Heizung in Frage kommt. Durch das Betrachten nur eines Ausschnitts aus der Gesamtheit der Werte kann somit das Problem der sich verkürzenden positiven und negativen Korrelationen umgangen werden.

Nachdem die Korrelationen für die Ausschnitte berechnet wurden, wird zusätzlich noch der Anteil des Stromverbrauchs für den betrachteten Teil am Gesamtverbrauch berechnet. Dies ermöglicht später Aussagen wie etwa die folgende: „Die Werte über 18° Celsius korrelieren zu 82 Prozent mit der Außentemperatur und beschreiben insgesamt 67 Prozent des Gesamtverbrauchs“.

6.3 Witterungsbereinigung

Als letzter Teil des Systems wurde ein Modul zur Witterungsbereinigung entwickelt. Dieses orientiert sich an dem Vorgehen, das auch zur Witterungsbereinigung von Heizungsdaten genutzt wird, und nutzt die Heiz- bzw. Kühlgradtage. Das Modul sollte wie die anderen auch für Experten und Laien leicht nutzbar sein. Dieses Modul unterscheidet sich von den beiden anderen bereits beschriebenen dadurch, dass hier Nutzerinteraktion nötig ist.

Da das System von smartB auf dem globalen Markt angeboten wird, muss das neu entwickelte Modul auf unterschiedliche Länder anwendbar sein. In den USA etwa tragen Klimaanlage einen großen Teil zum Gesamtverbrauch von Gebäuden bei.¹ In Frankreich hingegen werden häufig elektrische Heizungen genutzt die zu etwa einem Viertel des Stromverbrauchs der Haushalte beitragen.² Es ist also denkbar, dass ein Nutzer eine Witterungsbereinigung sowohl um Heizgradtage als auch um Kühlgradtage durchführen möchte. Das System muss die Möglichkeit bieten, zwischen den entsprechenden Optionen wählen zu können.

6.3.1 Definition Gradtagzahl

Um zu verstehen, wie die Witterungsbereinigung durchgeführt wird, muss zuerst geklärt werden, wie Heiz- bzw. Kühlgradtage definiert sind. Bei beiden handelt es sich um eine Version der Gradtagzahl, die im Folgenden erklärt wird.

Als Beispiel seien hier die in VDI-Norm 3807³ definierten Heizgradtage erwähnt. Das Vorgehen zur Berechnung der Kühlgradtage erfolgt analog.

Die Gradtagzahl ist die Summe der Differenzen zwischen einer gewählten Grenztemperatur (genannt Heizgrenze) und der durchschnittlichen täglichen Außentemperatur über einen gewählten Zeitraum.⁴ In Tabelle 6.1 ist die Berechnung der Gradtagzahl für eine Woche bei einer Heizgrenze von 15 Grad Celsius exemplarisch dargestellt.

¹So sind Klimaanlage in Privathaushalten allein für insgesamt 5% des Gesamtstromverbrauchs der USA verantwortlich (vgl. <http://energy.gov/energysaver/articles/air-conditioning>, <http://aceee.org/topics/air-conditioners>, abgerufen 2015-08-30

²vgl. *Bilan prévisionnel* 2014, S. 24.

³ Die VDI Norm 3807 („Verbrauchskennwerte für Gebäude“) stellt Regeln zur Berechnung von Verbrauchskennwerten auf. Dies ermöglicht es, Gebäude zu vergleichen und Einsparpotentiale zu entdecken. Vgl. <https://www.vdi.de/technik/fachthemen/bauen-und-gebaeudetechnik/fachbereiche/facility-management/richtlinien/vdi-3807/>, abgerufen 2015-08-31

⁴vgl. Wosnitza und Hilgers 2012, 274f.

Tag	Ø Außentemperatur °C	Gradtagzahl
1	12	3
2	16	0
3	17	0
4	14	1
5	12	3
6	16	0
7	12	3
Summe		10

Tabelle 6.1: Beispiel Heizgradtage (Basis 15° Celsius)

6.3.2 Bereinigung über Gradtagzahl

Nachdem die Gradtagzahl berechnet wurde, kann mit ihrer Hilfe eine Witterungsbereinigung durchgeführt werden. Dies kann auf zwei unterschiedliche Arten geschehen: als Bereinigung um die Heizgradtage eines bestimmten Jahres oder als Bereinigung um das langjährige Mittel. Die erste Variante ermöglicht eine Antwort auf die Frage *Wie wäre mein Stromverbrauch in diesem Jahr mit dem Wetter vom letzten Jahr gewesen?*, die zweite Variante ermöglicht eine grundlegendere Aussage über den erwarteten durchschnittlichen Verbrauch.

Um eine Bereinigung durchzuführen, wird zuerst eine Referenzgradtagzahl (zum Beispiel das langjährige Mittel) ausgewählt. Anschließend wird für jedes Jahr die Referenzgradtagzahl durch die für dieses Jahr ermittelte Gradtagzahl dividiert und mit dem Stromverbrauch multipliziert. In Tabelle 6.2 ist eine exemplarische Rechnung abgebildet. In diesem Beispiel unterscheiden sich die Jahresverbräuche erheblich, doch auch die Heizgradtage unterscheiden sich deutlich. In den bereinigten Werten wird sichtbar, dass der höhere Stromverbrauch in 2014 vor allem durch den kälteren Winter (und somit die höhere Gradtagzahl) bedingt ist.

Jahr	Heizgradtage	Verbrauch	Korrekturfaktor	bereinigter Verbrauch
2012	1.300	10.000 kWh	0,923	9.230 kWh
2013	1.100	8.000 kWh	1,09	8.720 kWh
2014	1.500	12.000 kWh	0,80	9,600 kWh

Tabelle 6.2: Beispiel Witterungsbereinigung (Bereinigt um das langjährige Mittel 1200)

In Deutschland werden Heizgradtage vom DWD für 500 Stationen bereitgestellt.⁵ Diese können in unterschiedlichen Auflösungen kostenlos auf der Website heruntergeladen werden. Des Weiteren können die Gradtagzahlen für die in Kapitel 4.2.3 beschriebenen TRY-Regionen heruntergeladen und für einen Vergleich genutzt werden. Alternativ erlaubt es die EnEV aber auch, die Gradtagzahlen anhand von eigenen Messungen zu berechnen.

6.3.3 Umsetzung der Witterungsbereinigung

Die Witterungsbereinigung, die in dem neu entwickelten System umgesetzt wurde, soll mehreren Ansprüchen genügen: Die Nutzer sollen auswählen können, ob sie eine Witterungsbereinigung um Heiz- oder Kühlgradtage durchführen, sie sollen entscheiden, auf Basis welcher Temperatur die Gradtagzahlen berechnet werden und sie sollen entscheiden, auf Basis welches Referenzzeitraums die Bereinigung durchgeführt werden soll. Dabei können sie zwischen einem der bemessenen Jahre oder dem langjährigen Mittel wählen. Auf eine Witterungsbereinigung um den Standort Würzburg oder anhand der TRY-Klimaregionen wurde verzichtet, da das System auch außerhalb von Deutschland zur Anwendung kommt und dieser Vergleich daher nicht sinnvoll wäre.

Bei der Benutzung des Systems muss der Nutzer zuerst auswählen, ob er eine Witterungsbereinigung um Heiz- oder Kühlgradtage durchführen möchte. Er hat in den vorherigen

⁵http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=_dwdww_w_klima_umwelt_klimadaten_deutschland&T17400910631149670782053gsbDocumentPath=Navigation%2F0effentlichkeit%2FKlima__Umwelt%2FKlimadaten%2Fspezialdaten%2FGTZ__start__node.html__nnn%3Dtrue&_state=maximized&_windowLabel=T17400910631149670782053,aufgerufen 2015-08-23

Teilen des Systems schon sehen können, welche Korrelation in seinem Gebäude vorliegt und anhand dieser Daten hoffentlich eine Idee bekommen, ob eine - und wenn ja welche - Witterungsbereinigung für sein Gebäude sinnvoll sein könnte. Nachdem er die Art der Bereinigung ausgewählt hat, muss er die Heiz-/Kühlgrenze angeben. Wenn ein Nutzer weiß, dass seine Klimaanlage erst bei einer Temperatur von 21 Grad Celsius eingeschaltet wird, dann bietet es sich an, die Kühlgradtage auf Basis dieser Temperatur zu berechnen. Dieses Wissen könnte er entweder schon zu Beginn der Analyse gehabt oder aus den vorhergehenden Modulen gewonnen haben. Nachdem er die Temperatur ausgewählt hat, wählt er schließlich das Jahr (oder das langjährige Mittel) aus, auf dessen Grundlage die Bereinigung durchgeführt werden soll.

Nachdem er alle Angaben getroffen hat, kann er in einer Tabelle, die sich unter der Eingabemaske befindet, die witterungsbereinigten Werte pro Jahr sehen. Außerdem werden in der Tabelle auch die Gradtagzahlen der Jahre angegeben, um eine einfachere Nachprüfbarkeit für den Nutzer zu erreichen. Die angezeigten Ergebnisse sind dabei sofort nach Änderung der Parameter verfügbar, dies ermöglicht eine eigenständige Analyse der Witterungsbereinigung.

Es ist an dieser Stelle denkbar, dass der Nutzer eine Witterungsbereinigung durchführt, die für sein Gebäude unpassend ist. So könnte er zum Beispiel eine Bereinigung um Heizgradtage durchführen, obwohl er keinerlei Verbraucher hat, die von steigenden Temperaturen beeinflusst werden. An dieser Stelle hätte das System so entwickelt werden können, dass solch eine Auswahl in diesem Fall gar nicht möglich gewesen wäre. Es ist aber davon auszugehen, dass die Nutzer des Systems eine gewisse Fachkenntnis haben und damit Aussagen treffen können, die besser sind als Aussagen, die sich ausschließlich aus den Daten ergeben. Um Nutzer auf eventuelle Fehleingaben hinzuweisen, wird aber ein kleines Symbol mit Erklärung gezeigt, wenn eine Witterungsbereinigung mit unerwarteten Werten durchgeführt wurde. Dies lässt den Nutzern ihre Freiheit, reduziert aber die Gefahr, dass sie unwissentlich Fehler machen, die zu falschen Aussagen führen könnten. Dies folgt dem Motto *Help users recognize, diagnose, and recover from errors*, das der Usability-Experte Jakob Nielsen

in seiner Sammlung *10 Usability Heuristics for User Interface Design* beschreibt.⁶

Für die zeitliche Auflösung der Witterungsbereinigung wurde ein Jahr gewählt. Es ist durchaus auch denkbar, eine Witterungsbereinigung auf Basis von monatlichen oder gar wöchentlichen Werten durchzuführen, dies führt jedoch zu einigen Problemen. Wie bereits erwähnt wurde, basiert die Berechnung der Korrekturfaktoren auf der Division des Referenzwerts durch den zu bereinigenden Wert. Betrachtet man Werte mit einer Auflösung von Wochen oder Monaten, kommt es häufig vor, dass einer der Werte (also entweder der Wert im Referenzjahr oder im zu bereinigenden Jahr) Null ist. Dies führt zu Problemen bei der Berechnung. Wenn der Wert des zu bereinigenden Jahres (also der Divisor) Null ist, ist das Ergebnis der Berechnung ($\frac{x}{0}$) undefiniert und es kann keine Bereinigung anhand dieses Korrekturfaktors durchgeführt werden. Wenn hingegen der Referenzwert (also der Divident) Null ist, ergibt sich als Korrekturfaktor 0 ($\frac{0}{x} = 0$) was dazu führt, dass der bereinigte Wert 0 kWh beträgt. Das Problem des Referenzjahres lässt sich dadurch einschränken, dass man langjährig gemittelte Werte betrachtet (da so die Wahrscheinlichkeit höher ist, dass ein Wert in einem der Jahre eine Temperatur über bzw. unter der Referenztemperatur hatte und damit der Durchschnitt dieser Werte über Null liegt), eine noch einfachere Lösung des Problem ist es jedoch, eine Auflösung von einem Jahr zu wählen, da so mit sehr großer Wahrscheinlichkeit die Gradtagzahlen für die Jahreswerte ungleich Null sind (hierfür muss schließlich nur ein Tag im Jahr die Referenztemperatur unter- bzw. überschreiten). Aus diesen Gründen wird dem Nutzer bei der Witterungsbereinigung nur die Anzeige der jährlichen Werte ermöglicht.

6.4 Nonfunktionale Anforderungen

Nachdem das System nun aus funktionaler Sicht konzipiert wurde, sollen einige weitere organisatorische Anforderungen aufgestellt werden, die das entwickelte System erfüllen muss. In größeren Betrieben würde dieser Schritt durch ein Pflichten- bzw. Lastenheft

⁶<http://www.nngroup.com/articles/ten-usability-heuristics/>, aufgerufen 2015-08-20.

Ursprünglich veröffentlicht in Nielsen 1994

spezifiziert werden, in agilen Startups wie smartB eins ist, ist dies jedoch unüblich. Die Anforderungen werden zwischen dem Product Owner und den Entwicklern kommuniziert und durch einen iterativen Entwicklungsprozess ständig abgeglichen. Dies bietet den Vorteil, dass nicht am Schluss ein System entwickelt wurde, das zwar den niedergeschriebenen, nicht aber den tatsächlichen Anforderungen entspricht, da diese sich immer ändern können. Durch eine stetige Kommunikation in den kleinen Teams zwischen Entwicklern und Product Owner kann sichergestellt werden, dass ein den Wünschen entsprechendes System entwickelt wurde.

Die nonfunktionalen Anforderungen, die mit dem Product Owner besprochen wurden, sollen nun geklärt werden. Wie bereits in den vorherigen Abschnitten ausgedrückt wurde, soll das Modul sowohl für erfahrene als auch für Gelegenheitsnutzer Einblicke in die bestehenden Zusammenhänge liefern und eine Witterungsbereinigung ermöglichen. Dabei soll das Modul, wie die bestehenden Teile des Systems auch, sofort auf Nutzereingaben reagieren, ohne dass es zu einem Neuladen der Seite kommen muss. Weiterhin sollen die neu entwickelten Teile sich graphisch in das bestehende System integrieren und einem einheitlichen Nutzerkonzept folgen.

7 Entwicklung des Systems

In diesem Kapitel wird die Entwicklung des neuen Systems beschrieben. Dafür werden zunächst die genutzten Technologien genannt und anschließend die Entwicklung der einzelnen Komponenten des neuen Moduls erläutert. Abbildung 7.1 gibt einen Überblick über die neue grobe Architektur des Systems. Die grau hinterlegten Teile wurden in der Entwicklung des neuen Moduls hinzugefügt und werden in den folgenden Abschnitten erläutert.

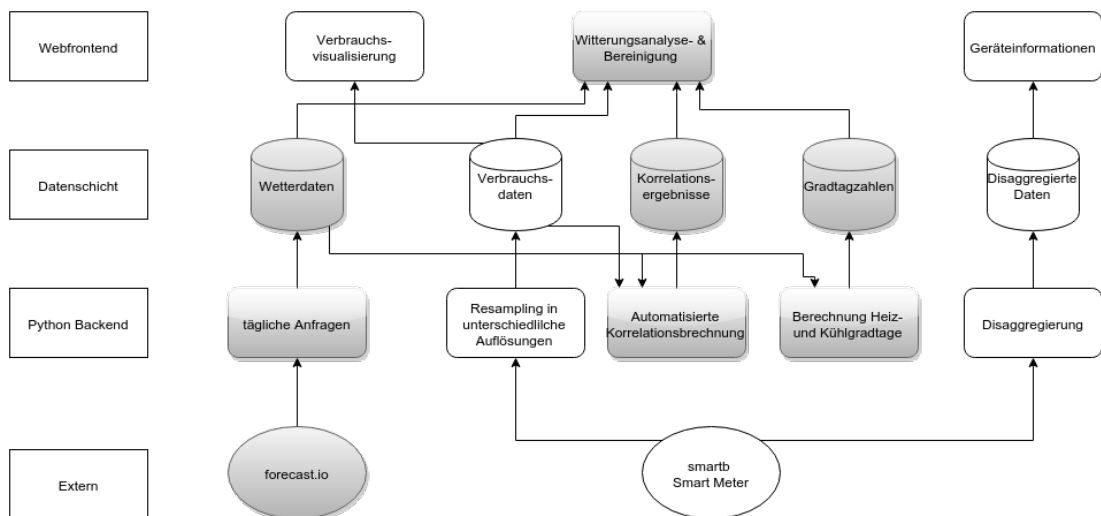


Abbildung 7.1: Architekturdiagramm

7.1 Genutzte Technologien

Der folgende Abschnitt gibt einen Überblick über die genutzten Technologien. Diese lassen sich dabei grob in zwei Kategorien aufteilen: Python¹-Bibliotheken inklusive Tools, die zur Datenanalyse und zur Entwicklung der Funktionen im Backend dienen, und Javascript-Bibliotheken, die zum Anzeigen der Daten im Frontend genutzt werden.

7.1.1 Datenanalyse

Zur in Kapitel 5 beschriebenen Untersuchung der Daten auf Korrelation wurde die Programmiersprache Python genutzt. Der Großteil der Datenanalyse wurde mit Bibliotheken und Tools aus der sciPy-Suite² durchgeführt. Die sciPy Suite umfasst die Bibliotheken NumPy, SciPy, Matplotlib, IPython, Sympy und Pandas. Ein Teil dieser Programme kam zur Datenanalyse und im Backend zum Einsatz.

Zur explorativen Datenanalyse wurde die Programmierumgebung jupyter³ genutzt, die aus dem iPython-Projekt⁴ hervorgegangen ist. Die Umgebung ermöglicht es, Python Code im Browser darzustellen und zu editieren. Der Code selber wird wie gewohnt in purem Python auf dem Rechner ausgeführt (es erfolgt also keine Cross-Compilierung nach Javascript wie etwa bei Skulpt⁵). iPython ermöglicht es, sogenannte *cells* zu erstellen. Diese können entweder Code oder beschreibende, mit markdown⁶ formatierte Texte enthalten. Des Weiteren können Graphiken, die zum Beispiel mit der weiter unten beschriebenen Bibliothek matplotlib erzeugt wurden, direkt angezeigt werden. Ein Beispiel für diese Funktionalität findet sich in Abbildung 7.2. Die mit jupyter erstellten sogenannten *Notebooks* befinden sich in einer Datei und lassen sich somit einfach weitergeben, dies ermöglicht auch ein gemeinsames Arbeiten an Projekten.

¹<http://python.org>

²Eric Jones, Travis Oliphant, Pearu Peterson u. a. 2001, vgl.

³<https://jupyter.org/>

⁴Pérez und Granger 2007.

⁵<http://www.skulpt.org/>

⁶vgl. <http://commonmark.org/>

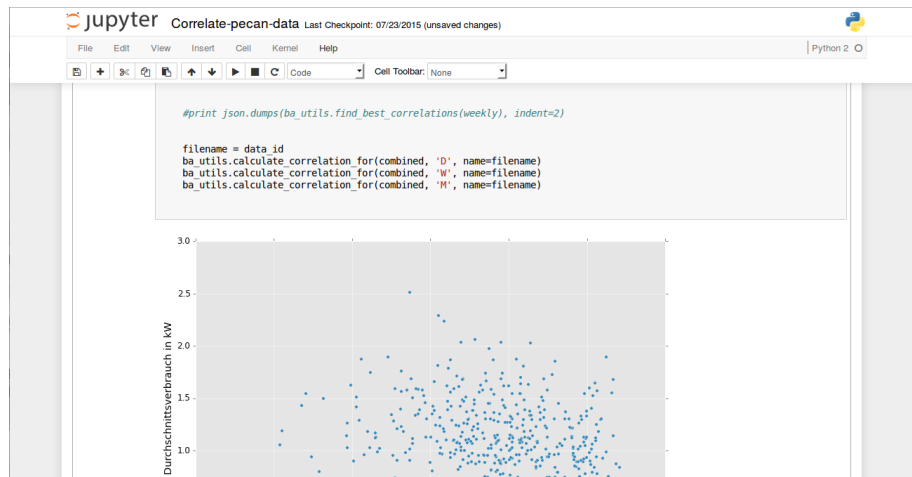


Abbildung 7.2: Die jupyter Programmierumgebung

Um Daten zu laden und zu analysieren, wurde die Bibliothek Pandas^{7,8} genutzt. Pandas wurde entwickelt, um Datenanalysen (insbesondere von Zeitreihendaten) zu vereinfachen. Pandas ermöglicht es, übliche Formate wie beispielsweise Javascript Object Notation (JSON), Comma Separated Values (CSV) oder Excel einzulesen und in einen sogenannten *Dataframe* umzuwandeln. Auf diesen Dataframe können dann Operationen wie etwa das Resampeln in eine andere Auflösung oder die Untersuchung auf lineare Regression schnell durchgeführt werden. Ein Beispiel für einen Workflow, in dem ein CSV Datensatz geladen, auf eine wöchentliche Auflösung umgewandelt und dann auf Korrelation zwischen zwei Reihen untersucht wird, findet sich in Beispiel 2.

Um Daten auf Korrelationen zu untersuchen, ist es oft hilfreich, sich diese anzuschauen. Um ansprechende Graphen zu erzeugen, wurde die Bibliothek matplotlib^{9,10} genutzt. Mit matplotlib können Visualisierungen (wie etwa die in Kapitel 5) erstellt werden. Dabei ist gerade das Zusammenspiel zwischen pandas und matplotlib herauszustellen. Beispiel 3 zeigt, wie die Daten aus Beispiel 2 in einem Liniendiagramm mit zwei Achsen dargestellt werden können. Hier wird in den ersten Zeilen iPython (bzw. Jupyter) so konfiguriert, dass die mit matplotlib erzeugten Graphen direkt in die Seite eingebunden werden. Anschlie-

⁷<http://pandas.pydata.org/>

⁸McKinney 2010.

⁹<http://matplotlib.org/>

¹⁰Hunter 2007.

```
dataframe = pd.read_csv("/path/to/data.csv",
                        header=0,
                        names=['datetime', 'consumption', 'temperature_c'],
                        sep=";",
                        parse_dates=[0],
                        index_col=[0]
                    )
weekly_data = dataframe.resample('W')
correlation = weekly_data.temeperature_c.corr(weekly_data.consumption)
print correaltion
```

Listing 2: Beispiel zur Datenanalyse mit Pandas

end wird auf die Temperatur und auf den Stromverbrauch aus dem vorherigen Beispiel zugegriffen und diese werden als Graph dargestellt. Nach dem Ausführen der Zelle, die diesen Code enthält erscheint der generierte Graph direkt unter dem Code.

```
# "magisches" iPython-Kommando das dafuer sorgt, dass die erzeugten
# Graphen direkt in das jupyter Notebook eingebunden werden
%matplotlib inline

weekly_data.temperature_c.plot()
weekly_data.consumption.plot(secondary_y=True)
```

Listing 3: Zwei-Achsen Liniendiagramm mit matplotlib

Die Grundlage für alle Analysen mit den sciPy-Bibliotheken liefert numPy^{11,12}. So sind zum Beispiel alle Spalten eines Pandas-Dataframes intern numpy Arrays. numPy stellt eine Möglichkeit bereit, n-dimensionale Arrays zu halten und auf diesen hochperformante Berechnungen auszuführen. Aus Performance-Gründen ist ein Großteil der numPy-Bibliothek

¹¹<http://www.numpy.org/>

¹²Walt, Colbert und Varoquaux 2011.

in C implementiert. Diese Module können normal aus Python Code aufgerufen werden, können aber deutlich schneller und ressourceneffizienter ausgeführt werden.

Zur einfacheren Erstellung der in 7.2 erwähnten Kommandozeilenprogramme wurde die Python-Bibliothek `click`¹³ benutzt, die die Interaktion mit Kommandozeilenparametern deutlich vereinfacht und zu einer besseren Benutzbarkeit der Programme führt. So ermöglicht sie es, mithilfe von Python Dekoratoren Parameter zu definieren, die dem Programm übergeben werden können. In dieser Definition können auch Hilfetexte oder Standardwerte gesetzt werden. Eine weitere Möglichkeit ist es, Unix-Umgebungsvariablen zu definieren, die genutzt werden sollen, wenn ein Parameter nicht übergeben wurde. Dies vereinfacht vor allem die spätere Benutzung als cronjob, da hier die Einstellungen für alle Tools einmalig in den Umgebungsvariablen gemacht werden können und nicht beim Aufruf jedes Tools geändert werden müssen.

7.1.2 Frontend

Bei dem Web-Frontend handelt es sich um eine Single Page Application (SPA). Die angeforderten Seiten werden also nicht auf dem Server generiert und fertig an den Browser geschickt, sondern die Anwendung besteht aus Templates (sogenannten *views*), die anschließend mit Daten gefüllt werden, indem die Daten als JSON nachgeladen werden. Dies ermöglicht es, dynamisch Daten nachzuladen, ohne dass der Browser die Seite neu laden muss. Dies führt zu einer schnelleren Nutzerinteraktion.

Die Anwendung basiert neben den grundlegenden Webtechnologien wie HTML, CSS und Javascript auf dem Javascript-Framework `angular.js`.¹⁴ Angular ermöglicht es auf einfache Art, dynamische Anwendungen zu entwickeln. Eine Angular App wird in *Services*, *Views* und *Controller* aufgeteilt. Die *Services* kommunizieren mit dem Server, um Daten nachzuladen, die *Controller* halten die Daten und stellen Interaktionsmöglichkeiten auf Programmebene dar und die *Views* sorgen für die Anzeige der Daten. Eins der Kernfeatures ist das

¹³<http://click.pocoo.org/4/>

¹⁴<http://angularjs.org/>

Two-Way Data Binding. Mit Hilfe dieser Technologie können Daten aus einem Controller einfach in einem View angezeigt werden. Weiterhin werden aber auch Änderungen, die im View an den Daten vorgenommen werden (etwa durch eine Interaktion vom Nutzer), direkt in den Controller übernommen, ohne dass hierfür weitere Programmierschritte notwendig sind. Dies führt zu einer deutlich schnelleren Anwendungsentwicklung.

Zur Darstellung der Visualisierungen wird die Javascript Bibliothek d3.js¹⁵ unter Zuhilfenahme des Angular-Wrappers angular-nvd3¹⁶ genutzt. Mit d3 können Scalable Vector Graphics (SVG) Graphiken im Browser erzeugt und direkt in die Website eingebunden werden. Die Nutzung von SVG hat gegenüber statischen Bildern, die auf dem Server generiert werden, die Vorteile, dass die Graphen beliebig skalierbar sind und dass eine Nutzerinteraktion ermöglicht wird. So können Nutzer zum Beispiel in der im nächsten Abschnitt beschriebenen Visualisierung mit dem Mauszeiger über Datenpunkte im Graphen gehen, um weitere Informationen (wie zum Beispiel den genauen Verbrauch) zu erhalten. Solch eine Interaktion wäre mit auf dem Server vorgenerierten Bildern nicht möglich.

7.2 Automatisierte Datenbeschaffung

Schon im vorhergehenden Kapitel zur Analyse der Korrelationen wurden Wetterdaten heruntergeladen, um diese mit den Energiedaten zu vergleichen. Die Daten, die hierfür genutzt wurden, kamen vom DWD, der NOAA oder direkt vom Pecan Street Projekt. Alle Datensätze folgten unterschiedlichen Formaten und mussten auf unterschiedliche Arten umgewandelt werden, bevor sie genutzt werden konnten. Die Aufbereitung dieser Daten erfordert das manuelle Suchen einer passenden Datenquelle, die Analyse der Datenstruktur und das Schreiben eines Adapters, der die Daten in das gewünschte Format umwandelt. Dieses Vorgehen lässt sich nicht auf das zu entwickelnde System übertragen, da hier die Daten automatisiert vom System bezogen werden müssen.

¹⁵<http://d3js.org/>

¹⁶<http://krispo.github.io/angular-nvd3/>

Um den Prozess der Beschaffung der Wetterdaten automatisieren zu können, wurden mehrere Anbieter analysiert und (in Bezug auf Kosten und Datenqualität) verglichen. Bei dem Anbieter, der schließlich gewählt wurde, handelt es sich um `forecast.io`.¹⁷ Forecast.io aggregiert bestehende Datenquellen wie beispielsweise die METAR und ISH Daten und fügt sie in einer einheitlichen Schnittstelle mit globaler Abdeckung zusammen.¹⁸

Die Daten stehen als REST-Schnittstelle bereit und lassen sich per HTTP-Get-Aufruf mit den Parametern Zeitpunkt und Koordinaten aufrufen. Bei der Antwort vom Server handelt es sich dann um JSON-Dokumente, die für jeweils einen Tag Daten (zum Beispiel zu Windgeschwindigkeit, Regenmenge und Lufttemperatur) mit einer stündlichen Auflösung beinhalten.

Um diese Daten automatisiert in das System zu integrieren, wurde ein Python Kommandozeilenprogramm geschrieben. Diesem Programm wird ein Startzeitpunkt, ein Endzeitpunkt und eine Liste von Gebäudenamen übergeben. Es orientiert sich in der Signatur an einem Kommandozeilenprogramm zum Download der Smart Meter-Daten, das schon im System existierte und ist exemplarisch in Beispiel 4 zu sehen.

```
download_weather --start_date 2014-01-01
                  --end_date 2015-01-01
                  --buildings pecan
```

Listing 4: Aufruf der Funktion zum Download der Wetterdaten

Das Skript findet in der Datenbank die ausgewählten Gebäude und extrahiert ihre Koordinaten (hierfür musste zuerst noch das Datenbank-Modell um Felder für die Koordinaten erweitert und diese im Frontend hinzugefügt werden. Siehe Kapitel 7.3). Anschließend wird für jedes Gebäude und jeden Tag, der im ausgewählten Zeitraum liegt, eine Anfrage an `forecast.io` gestellt und die Daten werden in einem zu dem Gebäude gehörenden Ordner gesammelt.

¹⁷<http://developer.forecast.io/>

¹⁸<https://developer.forecast.io/docs/v2#data-sources>, abgerufen 2015-08-27

Anschließend wird ein zweites Skript aufgerufen. Diesem Skript wird nur eine Liste von Gebäuden übergeben. Diese Liste wird genutzt, um die Rohdaten, die im vorherigen Schritt heruntergeladen wurden, einzulesen und auf wöchentliche und monatliche Auflösung zu resampeln. So erhält man zwei Dateien pro Gebäude: `weekly.json` und `monthly.json`. Diese beinhalten die Aggregate über den gesamten Zeitraum, für den Rohdaten vorhanden sind. Eine Übersicht über die hierbei erstellte Ordnerstruktur findet sich in Abbildung 7.3.

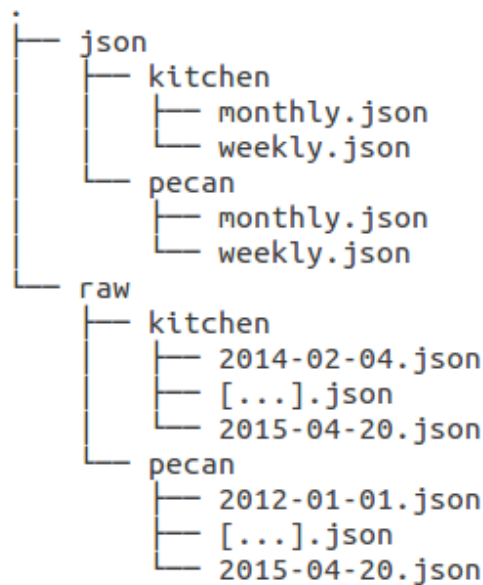


Abbildung 7.3: Ordnerstruktur der Wetterdaten

Beide Skripte werden schließlich auf dem Server per cronjob für jedes der Gebäude einmal pro Tag ausgeführt. Jeden Tag werden die neuesten Wetterdaten heruntergeladen und die aggregierten Werte berechnet. Von dort sind sie im Frontend zur Analyse nutzbar. Wird ein neues Gebäude hinzugefügt, kann das Skript einmalig manuell aufgerufen werden, um auch Daten herunterzuladen, die weiter in die Vergangenheit zurückreichen.

7.3 Einbindung der Gebäudekoordinaten

Damit die Daten wie im vorherigen Abschnitt beschrieben über den Dienst `forecast.io` heruntergeladen werden konnten, mussten die Koordinaten der Gebäude bekannt sein.

Dies ist wichtig, da `forecast.io` die Koordinaten für den Ort übergeben werden, für den Wetterdaten angefordert werden. Zu Beginn der Arbeit wurde der Standort der Gebäude als Adresse gespeichert. Zu dieser Adresse mussten nun die Geokoordinaten hinzugefügt werden. Dafür wurde zuerst das Datenbankschema angepasst, eine Karte in den Gebäudeeinstellungen im Frontend hinzugefügt und ein automatisierter Dienst zur Bestimmung der Koordinaten genutzt.

Zuerst musste das Datenbankschema so angepasst werden, dass die Koordinaten gespeichert werden konnten. Dafür wurde die Tabelle, die die Gebäudeinformationen enthält (genannt `buildings`) um die Spalten `latitude` und `longitude` erweitert. Die Daten werden dabei in einer PostgreSQL-Datenbank¹⁹ gespeichert, die mit dem Objekt Relationaler Mapper (ORM) `SQLAlchemy`²⁰ aus Python heraus genutzt wird. Um die Spalten der Datenbank hinzuzufügen, wurde die Python `Building` Klasse um die beiden Felder erweitert. Anschließend wurde eine Datenbankmigration mit `alembic`²¹ erstellt. Mit Hilfe von `alembic` können für Projekte, die `SQLAlchemy` benutzen, automatisch Migrationen erzeugt werden. Hierfür gleicht `alembic` das aktuelle Schema der Datenbank mit den Python-Klassen (bzw. `Models`) ab, erkennt Unterschiede und erzeugt gut lesbare Python-Skripte, die eine Migration der Datenbank auf das neue Schema durchführen.

Nachdem die Felder nun zur Datenbank hinzugefügt wurden, mussten sie mit Inhalt gefüllt werden. Dafür wurde ein Ansatz gewählt, der initial die Koordinaten des Gebäudes automatisiert herausfindet, diese dann im System anzeigt, um Nutzern die Möglichkeit zu geben, Korrekturen vorzunehmen.

Um diese Funktionalität zum Frontend hinzuzufügen, wurde die Seite mit den Gebäudeeinstellungen um eine Kartenkomponente erweitert. Diese basiert auf der Javascript Bibliothek `leaflet`,²² die es ermöglicht, `Openstreetmap`²³-Karten einzubinden. Zur besseren

¹⁹<http://www.postgresql.org/>

²⁰<http://www.sqlalchemy.org/>

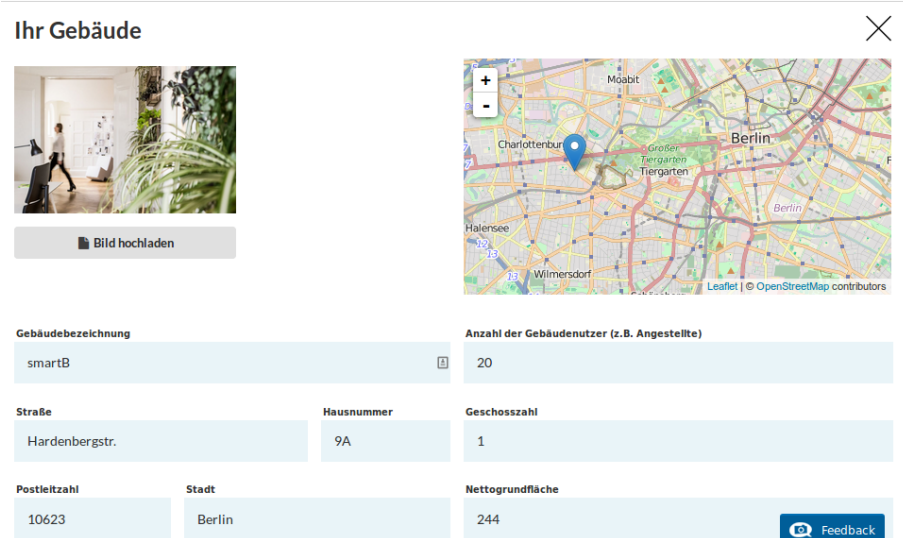
²¹<https://alembic.readthedocs.org/en/latest/>

²²<http://leafletjs.com/>

²³Bei `Openstreetmap` handelt es sich um das „Wikipedia für Karten“. Freiwillige erstellen gemeinsam eine Karte, die die gesamte Erde abdeckt. Die Daten stehen dabei unter einer freien Lizenz. Dies ermöglicht es sowohl Karten zu erstellen, als auch direkt mit den Daten zu arbeiten um zum Beispiel Geocoding-

Integration in das bestehende, auf angular.js basierende System wurde eine MIT-lizenzierte Direktive²⁴ benutzt.

Der neue Teil der Gebäudeeinstellungsseite prüft, ob schon Koordinaten für das Gebäude eingetragen sind. Wenn dies nicht der Fall ist, wird eine Anfrage an den auf Openstreet-map basierenden Web-Dienst Nominatim²⁵ geschickt. Die Anfrage ist strukturiert und enthält Hausnummer, Straße, Stadt und Land. Die Antwort wird per URL-Parameter auf einen Eintrag begrenzt und dieser kann im Frontend genutzt werden. Bei dem Format, mit dem kommuniziert wird, handelt es sich um JSON. Die zur Verfügung gestellten Koordinaten werden dem Nutzer auf der Karte angezeigt. Sollten sie fehlerhaft sein, kann er den Marker, der die Position des Gebäudes anzeigt, per Drag and Drop an die richtige Stelle verschieben. Mit einem Klick auf *speichern* können die korrigierten Koordinaten in die Datenbank übernommen werden. Beim nächsten Aufrufen der Seite stellt das System fest, dass die Koordinaten schon vorhanden sind, und führt keine erneute Anfrage an Nominatim durch.



The screenshot shows a form titled "Ihr Gebäude" with a close button (X) in the top right corner. On the left, there is a placeholder image of a modern office interior with a person and a "Bild hochladen" button. On the right, there is a map of Berlin with a blue location pin. Below the map and image, there are several input fields for building information:

Gebäudebezeichnung smartB	Anzahl der Gebäudenutzer (z.B. Angestellte) 20	
Straße Hardenbergstr.	Hausnummer 9A	Geschosszahl 1
Postleitzahl 10623	Stadt Berlin	Nettogrundfläche 244

A "Feedback" button is located at the bottom right of the form.

Abbildung 7.4: Die neue Kartenkomponente in den Gebäudeeinstellungen

Services zu erstellen. Genauere Informationen finden sich unter <http://www.openstreetmap.org>

²⁴<https://github.com/tombatossals/angular-leaflet-directive>

²⁵Clemens 2015, vgl.

7.4 Visualisierung von Korrelationen

Um das neue Modul zur Analyse und Bereinigung in das Frontend zu integrieren, wurde eine neue Seite erstellt. Dabei wurde der Methode des von Facebook definierten *Dark Launch*²⁶ gefolgt. Die neue Seite wurde in das System integriert, aber noch nicht in der Hauptnavigation aufgeführt. Dies bot mehrere Vorteile. Einerseits konnte im Backend getestet werden, ob der Server mit der Menge an Anfragen (etwa zum Download der Wetterdaten) überfordert oder ob alles in Ordnung war. Die Tatsache, dass die neue Seite nicht in der Hauptnavigation aufgeführt, sondern nur über die direkte Eingabe der Uniform Resource Locator (URL) erreichbar war, ermöglichte es andererseits, die neuen Features schon im bestehenden System zu testen, ohne ein unausgereiftes Produkt für Kunden zugänglich zu machen. Des Weiteren war es möglich, den Kunden die neue Seite bei Präsentationen zu zeigen und Feedback einzuholen.

Die Visualisierungskomponente ermöglicht es dem Nutzer, die wöchentlichen Verbrauchs- und Temperaturdaten in einem Streudiagramm und die monatlichen Daten in einem Linien-/Balkendiagramm zu betrachten. Beide Darstellungsformen sollten dem Nutzer die Möglichkeit geben, erste Erkenntnisse über eventuelle Zusammenhänge und Korrelationen zu erhalten.

Wie in Kapitel 6 beschrieben, wurde sowohl ein Balken-/Liniendiagramm als auch ein Streudiagramm erstellt. Das Streudiagramm zeigt wie beschrieben wöchentliche Daten. Die wöchentlichen Wetterdaten wurden, wie im Abschnitt zur Datenbeschaffung aufgeführt, erstellt, doch die Stromdaten lagen zu Beginn nicht in einer wöchentlichen Auflösung vor. Im System wurden bisher nie wöchentliche Daten angezeigt, weshalb das bestehende Modul zum Resamplen der Daten im Backend angepasst werden musste. Da auch hier alle Daten als Pandas-Dataframe vorlagen und die Funktionalität zum Schreiben der Daten auf die Festplatte schon entwickelt war, handelte es sich um eine einfache Änderung.

Zur Nutzung kommt hier die *angular-nvd3*-Bibliothek zur Erstellung der Graphen. Ein Angular-Service wurde entwickelt, um die Kommunikation mit dem Server zu übernehmen.

²⁶*Hammering Usernames* 2015.

Bei jedem Aufruf der Seite wird eine Anfrage für die Wetterdaten gemacht, da diese als Dateien vorliegen, die den gesamten bemessenen Zeitraum umfassen. Die Stromdaten, die schon an anderen Stellen des Systems genutzt wurden, lagen jedoch in Jahren gequantelt vor. Um auch diese Daten nutzen und im Frontend anzeigen zu können, muss eine Anfrage pro Jahr gestellt werden. Es wird bei dem Jahr begonnen, das in der Datenbank als Installationsdatum gespeichert ist und die Anfragen werden bis zum aktuellen Jahr gestellt. Die Daten werden von dem Service zusammengefasst und können dann im Controller genutzt werden. Für die Anzeige der Daten müssen der d3-Bibliothek lediglich die Daten und die gewünschten Optionen übergeben werden. Die Bibliothek stellt eine Angular-Direktive bereit, die dann direkt in das Template eingebunden werden kann.

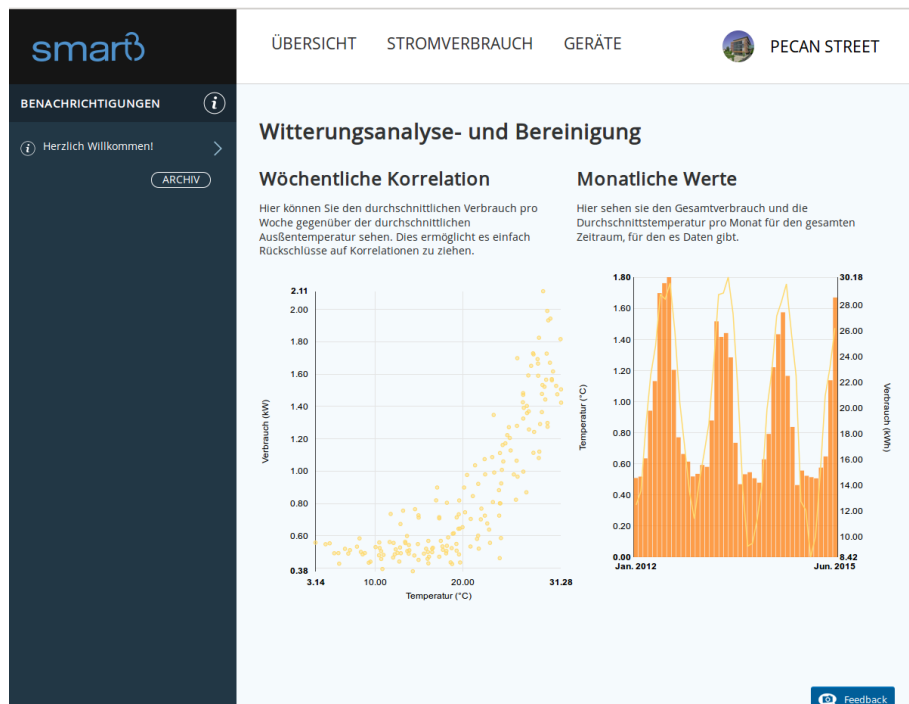


Abbildung 7.5: Die Datenvisualisierung im Frontend

7.5 Automatisierte Korrelationsanalyse

Mit der vorhergehenden Datenvisualisierung war es Nutzern möglich, sich selbstständig über eventuelle Korrelationen in den Daten zu informieren. Zur Vereinfachung dieses Pro-

zesses wurde im nächsten Schritt die Möglichkeit hinzugefügt, automatisiert vom System gefundene Korrelationen zu betrachten. Hierfür wurde die in 5.2.4 erwähnte Methode genutzt. Um die Daten auf Korrelationen zu untersuchen, werden mehrere Schritte unternommen. Um nun noch einen besseren Eindruck davon erlangen zu können, ob es sich um eine positive oder eine negative Korrelation handelt, werden schrittweise nur Teilmengen der Daten betrachtet. Das System versucht dabei, sowohl positive als auch negative Korrelationen (also Klimatisierung und Heizung) zu finden.

Das Vorgehen soll hier beispielhaft für Klimatisierung erläutert werden, ist aber analog auf die Heizung anzuwenden. Wenn die Vermutung besteht, dass es sich um eine Korrelation aufgrund einer Klimaanlage handelt, werden nur die Daten betrachtet, die über 15°C liegen. Für diese Daten wird die Korrelation berechnet. Ebenso wird für die Werte bis 25°C vorgegangen, wobei in jedem Schritt die Temperatur um ein Grad erhöht und dann die Korrelation berechnet wird. Die betrachteten Ausschnitte der Daten werden also immer kleiner. Wenn die gefundene Korrelation größer als die bisher größte Korrelation ist, wird sie zusammen mit der aktuellen Temperatur gespeichert. Außerdem wird der Anteil der betrachteten Daten an den Gesamtdaten und der Anteil des betrachteten Stromverbrauchs am gesamten Stromverbrauch berechnet. Um sicherzustellen, dass der Anteil der betrachteten Daten nicht unterrepräsentativ klein wird, wird dem Skript ein Parameter mit dem Anteil gegeben, den die betrachteten Daten mindestens im Verhältnis zur Gesamtheit der Daten beschreiben müssen. Standardmäßig handelt es sich hierbei um 30%. Das Ergebnis dieser Berechnung wird als JSON ausgeliefert und ist in Beispiel 5 zu sehen. Eine graphische Darstellung des Prozesses zur Berechnung findet sich in Abbildung 7.6.

Als Schnittstelle zu dem Skript wurde ein weiteres Kommandozeilenprogramm erstellt. Diesem wird als Parameter die Liste der Gebäude übergeben, für die die Korrelation berechnet werden soll. Sollten dem Skript keine Gebäude übergeben werden, findet die Berechnung für alle Gebäude statt, die in der Datenbank vorhanden sind. Alternativ kann die Liste der Gebäude auch, wie in Kapitel 7.1.1 beschrieben, über Umgebungsvariablen gesteuert werden.

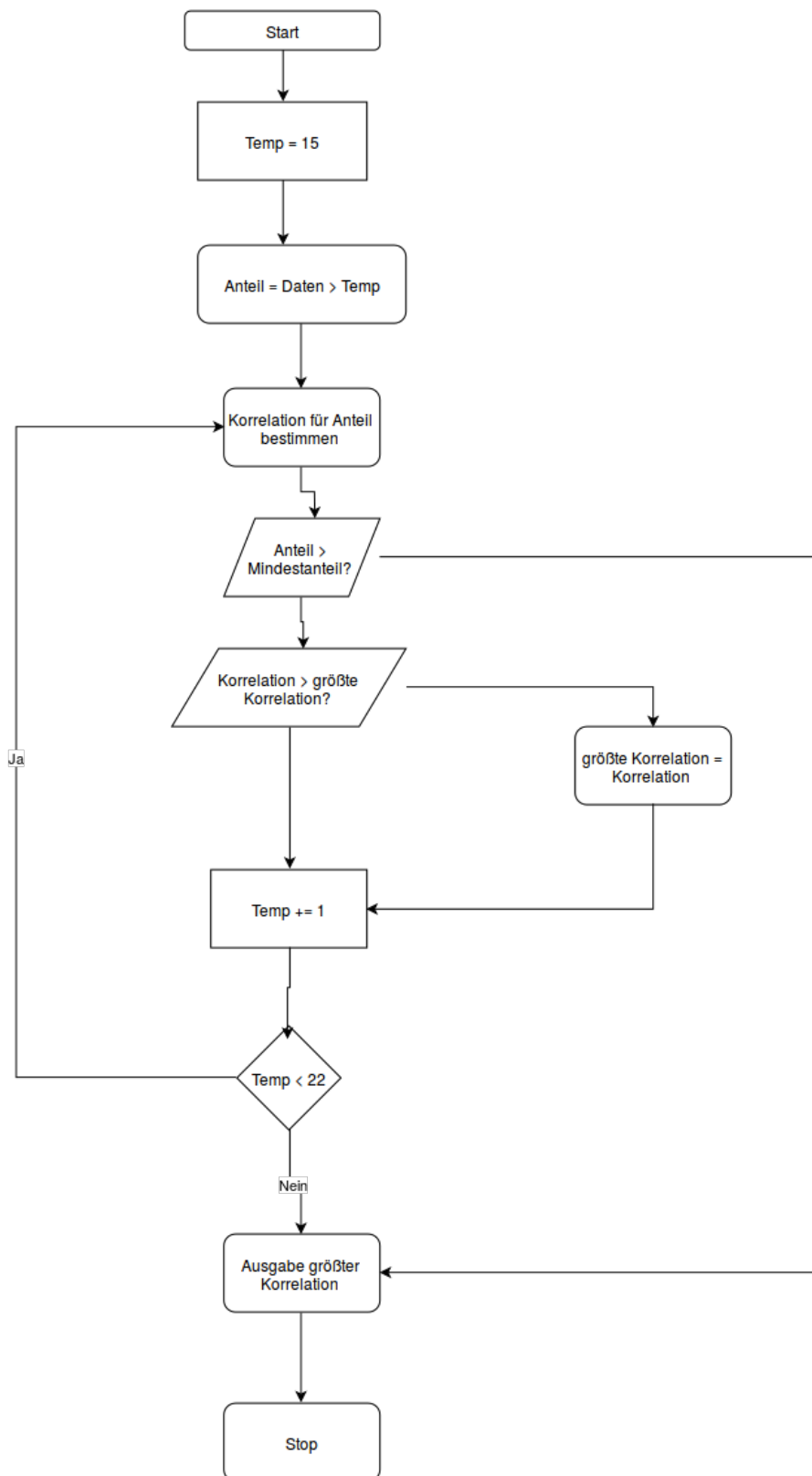


Abbildung 7.6: Vorgehen zur Korrelationsanalyse

```
{
  all: {
    corr: 0.8177745083200059
  },
  cooling: {
    percentage: 0.7065217391304348,
    consumption_percentage: 0.8229986865364779,
    temp: 15,
    corr: 0.8671574064376415
  },
  heating: {
    percentage: null,
    consumption_percentage: null,
    temp: null,
    corr: null
  }
}
```

Listing 5: JSON Beispiel für vom System gefundene Korrelation

7.6 Witterungsbereinigung über Gradtagzahlen

Das letzte Modul des neuen Systems sollte es den Nutzern ermöglichen, eine Witterungsberreinigung auf die Stromdaten auszuführen. Um auf die individuellen Umstände der Nutzer einzugehen, berechnet das System sowohl Kühlgradtage als auch Heizgradtage auf Basis unterschiedlicher Temperaturen. Hierfür werden erneut die Wetterdaten von forecast.io genutzt. Die zwei Nutzungsszenarien, die das System abbilden soll, sind die Witterungsberreinigung um ein bestimmtes Jahr („Wie wäre unser Stromverbrauch mit dem Wetter vom letzten Jahr gewesen?“) und der Vergleich zu den Durchschnittswerten der letzten

sieben Jahre.

Zwischenzeitlich bestand die Idee, die Option zur Witterungsbereinigung nur anzubieten, wenn auch automatisiert eine Korrelation gefunden wurde. Da es sich bei den Nutzern der Witterungsbereinigung aber um Experten handelt, die nicht von technischen System bevormundet werden sollen wurde auf diese Option verzichtet. Um Verwirrungen zu vermeiden wurden hier jedoch ausführlicherer Hilfestellungen in Form von erklärenden Texten integriert.

Auch für dieses Modul mussten Anpassungen im Backend durchgeführt werden, um es den Nutzern später zu ermöglichen, mehrere unterschiedliche Bereinigungen durchzuführen, ohne dass sie dafür Wartezeiten in Kauf nehmen müssen. Auf dem Server wurde ein neues Skript erstellt, das aus dem bestehenden Skript zum Berechnen der Korrelationen aufgerufen wird. Dies hat den Vorteil, dass die Daten, die bereits im Korrelationsskript geladen wurden, auch zur Berechnung der Gradtagzahlen genutzt werden können, was zu einer schnelleren Ausführung führt. Das neue Skript kann sowohl Heiz- als auch Kühlgradtage auf der Basis mehrerer Referenztemperaturen vorberechnen. Die Ergebnisse der Berechnungen werden - wie bei den anderen Skripten auch - als JSON auf der Festplatte gespeichert. Von dort aus können sie aus dem Frontend aufgerufen und dort genutzt werden. Ein Beispiel hierfür findet sich in Listing 6.

Im Frontend kann der Nutzer zunächst auswählen, ob durch Heiz- oder durch Kühlgradtage bereinigt werden soll. Anschließend wählt er die Temperatur aus, auf deren Basis die Gradtagzahlen genutzt werden sollen. Alle auswählbaren Gradtagzahlen liegen dabei vorberechnet im JSON-Format vor, so dass der Nutzer eine sofortige Rückmeldung vom System erhält, ohne Wartezeiten in Kauf nehmen zu müssen. Unter der Auswahl der Grundlage zur Bereinigung befindet sich eine Tabelle, die alle gemessenen Jahre umfasst (vergleiche Abbildung 7.7). Zusätzlich zum Jahr werden der Gesamtverbrauch in diesem Jahr und die Gradtagzahlen auf Basis der ausgewählten Temperatur angezeigt. Die Tabelle wird ergänzt um eine Spalte, in der der bereinigte Verbrauch angezeigt wird.

Bereinigte Werte

Hier können Sie die errechneten Gradtagzahlen sehen und ein Jahr auswählen, um das die anderen Werte bereinigt werden sollen. Dies ermöglicht es Ihnen einen Vergleich anzustellen, und zwar so als wäre die Temperatur in allen Jahren so gewesen wie im ausgewählten Referenzjahr.

Bereinige um:

Grundtemperatur Gradtagzahlen: °C

Referenzjahr:

Jahr	Kühlgradtage	orig. Wert	bereinigter Wert
2012	1.565	12,265 kWh	11,761 kWh
2013	1.501	10,516 kWh	10,516 kWh
2014	1.406	10,219 kWh	10,907 kWh
2015	891	5,582 kWh	9,398 kWh

Abbildung 7.7: Das Modul zur Witterungsbereinigung

```
{
  "cooling_degree_days": {
    "15": {
      "2012": 2680.944370289852,
      "2013": 2506.025942028985,
      "2014": 2343.9430500000003,
      "seven_year_average": 2360.518482210145
    },
    "16": {
      "2012": 2399.1986999999997,
      "2013": 2249.9395833333338,
      "2014": 2092.7729166666663,
      "seven_year_average": 2107.932069166667
    }
  },
  ...
},
  "heating_degree_days": {
    "8": {
      ...
    },
    "9": {
      ...
    }
  },
  ...
}
```

Listing 6: Beispiel für Gradtagzahl JSON

8 Test des Systems

Eine der Leitfragen, unter denen diese Arbeit steht, lautet „Wie kann der Zusammenhang zwischen Temperatur und Stromverbrauch für (unerfahrene) Nutzer sichtbar gemacht werden?“. Um zu überprüfen, ob das entwickelte Modul für Experten und Laien verständlich ist, wurden im Anschluss an die Entwicklung zwei Tests durchgeführt. Bei dem ersten Test handelte es sich um ein Gespräch mit einem Energiemanager. Dabei war das Ziel, zu sehen, ob ein Experte das System für einfach nutzbar und funktional sinnvoll hält. Der zweite Test zielte vermehrt auf die Usability des Systems. Er wurde mit einem Mitarbeiter von smartB durchgeführt, der kein Energiemanager, aber grundsätzlich mit dem Thema vertraut ist. Hier wurden Aufgaben gestellt, die der Nutzer lösen sollte. Sein Umgang mit dem System wurde dabei analysiert.

Da hier nur Tests mit zwei Personen durchgeführt wurden, kann natürlich nicht von belastbaren Ergebnissen gesprochen werden. Die Tests sollten lediglich dazu dienen, einen Einblick zu geben, ob das System in seiner jetzigen Form für Nutzer einfach verständlich und gut zu bedienen ist. Inwiefern sich die Ergebnisse dieser Tests auch auf spätere Nutzer übertragen lassen, bleibt dabei offen. Hier können Technologien wie etwa Google Analytics¹ genutzt werden, um den Umgang der Nutzer mit dem Live-System mitzuschneiden und dann zu analysieren. Aufgrund der Ergebnisse der Analyse können dann eventuell Änderungen getätigt werden.

¹<http://www.google.com/analytics/>

8.1 Expertengespräch

Nach der Entwicklung des Systems wurde einem erfahrenen Energiemanager einer Partnerfirma von smartB ein vorläufiger Zugang zu dem System ermöglicht. Er bestätigte in einem Telefonat, dass das Streudiagramm und das Balken-/Liniendiagramm einen einfachen Überblick über den Zusammenhang der Daten aufzeigen. Die automatisch gefundenen Korrelationswerte waren für ihn nicht sofort verständlich, woraufhin die beschreibenden Texte geändert wurden, um besser zu erklären, was für eine Analyse durchgeführt wurde. Das letzte Modul der Anwendung - die Witterungsbereinigung - konnte er sofort verstehen und nutzen. Er schlug einige Änderungen vor - so war in der ihm vorgestellten Version der Vergleich mit dem langjährigen Mittel noch nicht möglich, wurde aber auf seinen Wunsch hinzugefügt. Außerdem wurden in dem Gespräch einige kleinere Usability-Probleme erkannt, die im Folgenden behoben wurden.²

8.2 Nutzertest mit unerfahrenem Nutzer

Ein Ziel dieser Arbeit ist es, ein System zu entwickeln, das sowohl von Experten als auch von Gelegenheitsnutzern einfach zu nutzen ist. Um zu testen, wie einfach das System für Laien zu benutzen ist, wurde ein weiterer Test mit einem Mitarbeiter aus der Verkaufsabteilung von smartB, der bisher nicht mit dem System vertraut war, durchgeführt. Hierbei handelte es sich im Vergleich zu dem vorhergehenden Gespräch um einen strukturierten Usability-Test. Dem Nutzer wurde Zugang zum System gegeben und er sollte zwei Aktionen durchführen. Dabei war er angehalten, seine Gedanken laut auszusprechen, damit etwaige Unklarheiten dokumentiert werden konnten.

Die Untersuchung begann damit, dass dem Nutzer Zugang auf die neu entwickelte Seite gegeben wurde. Ein Gebäude war bereits vorausgewählt, da die Gebäudeauswahl nicht

²Zum Zeitpunkt des Tests war etwa die Auswahl des Referenzjahres zur Bereinigung noch in die Tabelle eingegliedert. Sie wurde im Folgenden über die Tabelle zu den anderen Drop-Down-Menüs verschoben, um ein einheitlicheres Benutzungskonzept zu bilden.

Teil des Tests sein sollte. Bei dem Gebäude handelte es sich um eins der Gebäude aus dem Pecan Street Project. Dieses wurde gewählt, da hier ausreichend Daten vorlagen und eine Korrelation vorhanden war. Zuerst wurden dem Nutzer einige Minuten Zeit gegeben, um sich mit dem System vertraut zu machen. Schon hierbei fielen einige Kleinigkeiten auf; so waren zum Beispiel die Punkte im Streudiagramm zu klein und der Kontrast zum Hintergrund zu gering.

Die erste Aufgabe, die dem Nutzer gestellt wurde, lautete „Sage mir bitte, ob du für dieses Gebäude eine Korrelation zwischen Außentemperatur und Stromverbrauch erkennen kannst“. Die Aufgabe konnte sofort gelöst werden, und eine Korrelation wurde erkannt. Die Frage, die dieser Aufgabe zu Grunde lag, war die Überlegung, in welchem der Module die Korrelation am deutlichsten wurde. Die Testperson hatte das Verständnis der Korrelation durch Betrachten des Streudiagramms erlangt und stellte sogar selbstständig fest, dass bei einem Szenario wie in 5.2.4 beschrieben, hier ein talförmiger Zusammenhang sichtbar sein müsste.

Die zweite Aufgabe, die dem Nutzer gegeben wurde, war „Nenne mir bitte den um Kühlgradtage auf einer Basis von 19 Grad Celsius bereinigten Verbrauch von 2013, wenn das Jahr 2014 als Referenzjahr für den Verbrauch genutzt wird.“ Diese Aufgabe bereitete dem Nutzer etwas mehr Probleme, da er mit dem Prinzip der Witterungsbereinigung und der Gradtagzahlen nicht vollständig vertraut war. Es gelang ihm zwar schnell, die geforderten Eingaben im Eingabeformular zu machen, ihm war aber nicht klar, wie die Ergebnisse zu deuten sind. Im Regelfall wird die Witterungsbereinigung nur von erfahreneren Nutzern durchgeführt, denen bewusst ist, wie sie funktioniert. Sollte sie von Nutzern durchgeführt werden, denen das Konzept nicht klar ist, sind diese dazu angehalten, sich eigenständig zu informieren. Nachdem der Testperson das Konzept der Witterungsbereinigung erklärt worden war, war auch für sie die Nutzung des Systems einfach verständlich, und das richtige Ergebnis konnte genannt werden.

8.3 Erfüllung der funktionalen Anforderungen

Die funktionalen Anforderungen, die an das System gestellt wurden, lassen sich grob in zwei Teile unterteilen - ein Modul zur Witterungsanalyse und ein Modul zur Bereinigung. Mit dem Modul zur Analyse sollte es möglich sein, Zusammenhänge zwischen Außentemperatur und Stromverbrauch zu erkennen. Das Modul, das hierfür entwickelt wurde, besteht aus zwei unterschiedlichen Visualisierungen. Die erste Art der Visualisierung, stellt die Daten in einer für Nutzer intuitiv erkennbaren Form des Balken- und Liniendiagramms dar. Die zweite Visualisierung widmet sich genauer dem Zusammenhang der beiden Faktoren, in dem diese in einem Streudiagramm abgebildet werden. Diese Art der Darstellung ermöglicht es, einfach zu sehen, ob lineare - eventuell auch zweigeteilte - Korrelationen zwischen den Daten bestehen. Abschließend wurde diese Anforderung durch die automatisierte Korrelationsanalyse abgedeckt. Diese hilft, zusätzlich zu den graphischen Daten auch noch eine konkrete Zahl zu liefern, mit der der Grad der Korrelation beschrieben werden kann.

Der zweite Teil des neuen System sollte es den Benutzern ermöglichen, ihre Jahresverbräuche zur besseren Vergleichbarkeit zu bereinigen. Hierfür wurde das Modul zur Witterungsbereinigung entwickelt. Das Modul ermöglicht es, die beiden möglichen Arten der Bereinigung - um Kühl- und um Heizgradtage - durchzuführen.

8.4 Erfüllung der nonfunktionalen Anforderungen

In Kapitel 6.4 wurden nonfunktionale Anforderungen an das System aufgestellt. Ihre Erfüllung soll in diesem Abschnitt untersucht werden. In den in diesem Kapitel beschriebenen Nutzertests zeigte sich, dass die erste Anforderung - ein Modul, das einfach zu verstehen und zu benutzen ist - erfüllt wurde. Die Visualisierungen ermöglichen es den Nutzern, ein eigenständiges Verständnis der Zusammenhänge zwischen den Daten zu entwickeln.

Durch die Vorberechnung der relevanten Daten wie etwa der Gradtagzahlen (vgl. Kapitel 7.6) ist es möglich, den Nutzern sofort nach Änderungen der Parameter im System eine

Rückmeldung zu geben. Dies trägt zur Erfüllung der Anforderung an ein System bei, das schnell und ohne Neuladen der Seite betrieben werden kann.

Zuletzt sollte das neue Modul sich in Bezug auf Aussehen und Nutzungskonzept in das bestehende System integrieren. Durch die Nutzung der Farben für die Visualisierung, die auch in den bestehenden Graphen schon genutzt wurden, und durch Wiederaufgreifen der Tooltips in den Graphen können Nutzer alle Graphen einheitlich benutzen und verstehen. Durch die Nutzung der bestehenden Stile für die Komponenten der Benutzeroberfläche gliedert sich das neue Modul gut in das bestehende System ein.

8.5 Diskussion der Ergebnisse

Das Ziel der beiden Benutzertests war es, zu sehen, ob das System einfach nutzbar ist. Dabei zeigte sich, dass der erfahrene Energiemanager keine Probleme bei der Benutzung des Systems hatte. In dem zweiten Test wurde deutlich, dass die Visualisierung auch für unerfahrenere Nutzer einfach zu verstehen war. In dem Teil des Tests, der sich mit der Witterungsbereinigung beschäftigte, lagen die Schwierigkeiten nicht in der Nutzung des Systems selbst, sondern im Verständnis der dabei gefundenen Ergebnisse. Hier fehlten erklärende Hinweistexte im Programm. Dies ist aber ein Problem, dessen Lösung gegen das Grundprinzip des Systems verstößt, da auch in keinem anderen Teil ausgiebige Informationstexte vorhanden sind. Das Einbinden eines erklärenden Textes würde somit zu einem uneinheitlichen Nutzererlebnis führen. Es muss vorausgesetzt werden können, dass Nutzer, die ein Interesse daran haben, eine Witterungsbereinigung durchzuführen, mit dem zugrundeliegenden Konzept vertraut sind.

Zusammenfassend zeigt sich also, dass insbesondere die Analyse der Daten für Nutzer einfach verständlich ist und dass auch die Nutzung des Systems keine Probleme bereitet. Ein Verständnis der Bereinigung über Gradtagzahlen ist nötig, wenn sie durchgeführt werden soll.

Die Überprüfung der funktionalen und nonfunktionalen Anforderungen zeigt, dass das

neu entwickelte Modul das in Kapitel 6 entwickelte Konzept funktional umgesetzt hat und dass das neue Modul sich auch auf den anderen Ebenen - wie etwa Nutzungskonzept und Design - gut in das bestehende System integriert.

9 Fazit und Ausblick

In den vorherigen Kapiteln wurde gezeigt, welche Art der Witterungsanalyse und -bereinigung mit dem heutigen Stand der Technik möglich ist. Dabei wurde nur die Korrelation zwischen Außentemperatur und dem Gesamtverbrauch von Gebäuden analysiert. Für die Zukunft wären zwei weitere Ansätze von Interesse: Mit dem von smartB entwickelten System wird es in Zukunft noch besser möglich sein, einzelne Verbraucher durch die Disaggregation der Stromdaten zu erkennen. Wenn Informationen über einzelne Verbraucher vorhanden sind, bietet es sich an, nur die Verbraucher zu analysieren und zu bereinigen, die auch tatsächlich in Verbindung mit der Außentemperatur stehen. So werden zum Beispiel in dem aktuell entwickelten System zwar Klimaanlage und elektrische Heizungen betrachtet, die sicherlich mit der Außentemperatur korrelieren, es werden jedoch auch andere Geräte wie beispielsweise Durchlauferhitzer und Bürogeräte in die Korrelationsanalyse und -bereinigung mit einbezogen. Es wäre noch sinnvoller, diese Geräte aus der Analyse und Bereinigung auszuschließen und nur die wirklich korrelierenden Verbraucher zu betrachten. Die Disaggregation ist noch nicht ganz auf dem Stand, bei dem ein so detaillierter Vergleich möglich ist. In den nächsten Jahren wird jedoch voraussichtlich die Genauigkeit deutlich zunehmen und einen solchen Vergleich ermöglichen.

Eine weitere Möglichkeit wäre die Korrelationsanalyse für Beleuchtung zum Beispiel anhand von Sonnenscheinstunden. Dies wurde in dieser Arbeit aufgrund des Umfangs und der nicht ausreichend vorhandenen Daten nicht behandelt, würde aber einen weiteren Einblick in den Verbrauch von Gebäuden ermöglichen. Da es jedoch eine recht große Korrelation zwischen der in dieser Arbeit betrachteten Außentemperatur und den äußeren Lichtver-

hältnissen gibt,¹ kann auch dieser Ansatz schon als Annäherung betrachtet werden. Eine noch genauere Analyse wäre jedoch interessant. Hierfür müsste jedoch erst ein passender Faktor gefunden werden, so könnte man sowohl Sonnenscheinstunden als auch Bedeckung betrachten.

Darüber hinaus werden in Zukunft auch andere Module, die das System von smartB ergänzen, auf dieser Arbeit aufbauen können. Bei einem Feature, das in naher Zukunft umgesetzt werden soll, handelt es sich um den Vergleich zweier unterschiedlicher Gebäude. Damit diese miteinander verglichen werden können, müssen sie zuerst auf einen Faktor (wie etwa kWh pro Quadratmeter) genormt und dann um regionale Wetterschwankungen bereinigt werden. Hierfür können die in dieser Arbeit berechneten Gradtagzahlen genutzt werden. Diese müssen anschließend nur mit einem beliebigen gemeinsamen Referenzwert bereinigt werden, um Vergleichbarkeit herstellen zu können.

Doch die vergangenen Kapitel zeigen auch: Witterungsbereinigung von Stromdaten ist aufgrund der Vielfalt der elektrischen Verbraucher in einem Stromkreis bisher ein wenig betrachtetes Feld. Es ist jedoch auch hier möglich, Korrelationen zwischen Außentemperatur und Stromverbrauch zu finden. Wenn es eine Korrelation in den Daten gibt, dann bietet es sich an, eine Witterungsbereinigung durchzuführen. Diese kann mit dem entwickelten Programm einfach verständlich für die Nutzer umgesetzt werden.

Doch auch schon der erste Schritt, die Visualisierung und Analyse der Daten, kann Nutzern hilfreiche Einblicke in den Stromverbrauch ihrer Gebäude liefern. Die automatisierte Korrelationsanalyse ist ein sinnvolles Werkzeug zur Analyse von Stromdaten, da die Zusammenhänge zwischen Stromverbrauch und Außentemperatur hier nicht so direkt und eindeutig sind wie die zwischen der Heizungsnutzung und der Außentemperatur. Mit Hilfe der automatisierten, schrittweisen Analyse können hier weitere Einblicke in die Zusammenhänge errungen werden.

Nun sollen noch die drei Leitfragen aus dem Beginn der Arbeit beantwortet werden. Die Frage, inwiefern es eine Korrelation von Strom- und Temperaturdaten gibt und welche Un-

¹Hier ist anzumerken, dass die Korrelation nur auf die Jahreszeiten zurückzuführen ist und die beiden Faktoren sich nicht gegenseitig beeinflussen. vgl. Peter Hofer 2013, S. 10

terschiede zur Korrelation bei Heizdaten bestehen, wurde in Kapitel 5 beantwortet. Dort wurde gezeigt, dass durchaus eine Korrelation zwischen Außentemperatur und Stromverbrauch von Gebäuden vorliegen kann. Dabei ist je nach Land und Gebäude die Korrelation unterschiedlich stark. Anhand der Daten aus dem Pecan Street Projekt aus den USA konnte gut eine starke positive Korrelation zwischen den beiden Faktoren erkannt werden. Grund hierfür ist der Einsatz von Klimaanlage, der in den USA weit verbreitet ist. Das Beispiel aus Frankreich zeigt, dass die Korrelation aber auch negativ sein kann. Dies tritt auf, wenn - wie in Frankreich verbreitet - mit Strom geheizt wird oder zum Beispiel die Beleuchtung im Winter mehr genutzt wird. Auch der Sonderfall der sowohl positiven als auch negativen Korrelation in einem Datensatz wurde betrachtet und eine Methode zur Untersuchung auf Korrelationen entwickelt.

Die zweite Leitfrage lautete „Wie kann der Zusammenhang zwischen Temperatur und Stromverbrauch für (unerfahrene) Nutzer sichtbar gemacht werden?“. Hier zeigte sich insbesondere durch die Nutzertests in Kapitel 8, dass das Streudiagramm eine geeignete Art der Darstellung von Korrelationen ist, da aus diesem besonders schnell ein Zusammenhang ersichtlich ist und auch hier der Sonderfall der zweiteiligen Korrelation erkannt werden kann. Die automatisierte Korrelationsanalyse ist ein weiteres Mittel, um es für Nutzer einfacher zu machen, präzisere Aussagen über ihren Stromverbrauch und die Korrelation zur Außentemperatur treffen zu können.

Zuletzt sollte noch die Frage beantwortet werden, wie eine Witterungsbereinigung für Stromdaten umgesetzt werden kann. Hier konnte zuerst einmal gezeigt werden, dass eine Witterungsbereinigung auch für Stromdaten sinnvoll ist, da auch hier oft eine Korrelation besteht. Anhand des für erfahrene Nutzer vertrauten Vorgehens zur Witterungsbereinigung von Heizdaten über die Gradtagzahlen wurde ein System entwickelt, das es Nutzern ermöglicht, einfach und dynamisch eine Witterungsbereinigung sowohl um Heiz- als auch um Kühlgradtage anhand unterschiedlicher Referenztemperaturen und -jahre durchzuführen.

Abschließend lässt sich sagen, dass im Rahmen dieser Arbeit das bisher wenig betrachtete Feld der Witterungsanalyse und -bereinigung von Stromdaten untersucht und eine

für Nutzer ansprechende und interaktive Möglichkeit geschaffen wurde, eine Analyse und Bereinigung vorzunehmen. Die Arbeit bietet weiterhin die Grundlage für viele andere Möglichkeiten, die in diesem Kapitel vorgestellt wurden, und wird somit auch in Zukunft mehr Energietransparenz für Nutzer ermöglichen machen.

A Übersicht über die veröffentlichten NILM-Datensätze

Haushalte

Bei den meisten der öffentlich zugänglichen Datensätze, die den Stromverbrauch beschreiben, handelt es sich um Datensätze von Wohnhäusern oder Wohnungen. Dieser Anhang soll einen Überblick über diese Datensätze geben und darstellen, welche Datensätze für die Analyse in dieser Arbeit besonders interessant waren.

AMPds

Der AMPds Datensatz ^{1,2} wird von Stephen Makonin veröffentlicht und beschreibt seinen privaten Haushalt in Kanada für etwa zwei Jahre. Dabei wurde sowohl der gesamte Stromkreis als auch einzelne Verbraucher separat bemessen. Zusätzliche zu den Stromdaten sind auch Messwerte zum Gas- und Wasserverbrauch vorhanden. Alle Werte liegen in einer minütlichen Auflösung vor. Besonders interessant an dem Haushalt ist, dass ein sogenanntes *dual-fuel system* zur Heizung vorliegt. Dies heizt bis zu einer gewissen Temperatur (2°C) mit Strom, bevor auf Gas umgestellt wird. Aus diesem Grund ist die Korrelation im Stromverbrauch geringer als bei einer rein elektrischen Heizung und der Datensatz wurde daher nicht für die Untersuchung genutzt.

¹<http://ampds.org/>, aufgerufen 2015-08-21

²Makonin u. a. 2013.

ECO

Für den ECO Datensatz³ wurden sechs Haushalte in der Schweiz für acht Monate untersucht. Dabei wurde sowohl der Gesamtverbrauch als auch einzelne speziell ausgewählte Geräte in den Haushalten mit einer Auflösung von 1 Hz gemessen. Zusätzlich zu den Stromdaten enthält der Datensatz außerdem Informationen dazu, wann und wie viele Menschen in dem Haus waren. Dies soll es Wissenschaftlern ermöglichen, aus dem Verbrauch von Haushaltsgeräten Rückschlüsse auf die Belegung von Gebäuden zu ziehen und diese mit den gegebenen Daten zu verifizieren.

GREEND

Der GREEND Datensatz⁴ stammt aus Österreich und Italien. Es wurden insgesamt in acht Häusern gemessen, jeweils die Hälfte in Österreich und Italien. In den Unterlagen finden sich Informationen zu den Bewohnern der Häuser. Dies ermöglicht es, Rückschlüsse auf ihr Verbraucherverhalten zu ziehen. In manchen der Haushalte wurden nur einzelne Geräte bzw. Gerätegruppen bemessen, andere beinhalten auch Daten über den Gesamtverbrauch. Die Geräte, die bemessen wurden, unterscheiden sich zwischen den Haushalten, um eine möglichst große Menge an unterschiedlichen Daten bereitstellen zu können. Dies erschwert aber die Vergleichbarkeit zwischen den Haushalten. Der Zeitraum, für den Daten vorliegen, beträgt zwischen 3 und 12 Monaten.

iAWE

Ziel des iAWE Projektes^{5,6} war es, zusätzlich zu den schon veröffentlichten Datensätzen aus Industrienationen wie den USA, Österreich oder Italien auch einen Datensatz aus einem Entwicklungsland - in diesem Fall Indien - bereitzustellen. Dies ist gerade in Indien

³Beckel u. a. 2014.

⁴Monacchi u. a. 2014.

⁵<http://iawe.github.io/>

⁶Batra, Gulati u. a. 2013.

interessant, da dort 93 Prozent des gesamten Stromverbrauchs aus Haushalten stammt⁷ - in Deutschland sind es zum Vergleich nur 27 Prozent.⁸ Gemessen wurde in einem Haus über einen Zeitraum von 73 Tagen. Zusätzlich zum Stromverbrauch werden in diesem Datensatz auch der Wasserverbrauch und Umweltdaten bereitgestellt. Die Stromdaten umfassen sowohl den Gesamtverbrauch als auch einzelne ausgewählte Verbraucher. Die Autoren stellen besonders heraus, dass das Messsystem als Open Source System entwickelt wurde und zusätzlich genaue Metadaten zu den bemessenen Geräten veröffentlicht wurden.

IHEPCDS

Der IHEPCDS Datensatz^{9,10} aus Frankreich (Region Sceaux) ist mit einer Laufzeit von 47 Monaten der langfristige veröffentlichte Datensatz im NILM Bereich. Zusätzlich zum Gesamtverbrauch wurden drei Substromkreise bemessen. Diese umfassen die Küche, den Wäscheraum und die Klimaanlage und Durchlauferhitzer. Aufgrund der langen Laufzeit (und der Nähe zu Deutschland, dem derzeitigen Haupteinsatzort des smartB Systems) bot sich dieser Datensatz für die Untersuchung besonders an.

Pecan Street

Das Pecan Street Projekt¹¹ aus den USA stellt mit dem Dataport¹² eine sehr große Anzahl an Messwerten für Wissenschaftler kostenlos zur Verfügung. Über einen Zeitraum von mehreren Jahren wurden mehrere hundert Gebäude in Texas, Kalifornien und Colorado bemessen. Je nach Gebäude unterscheidet sich die Menge an Daten, die verfügbar sind. Für viele der Haushalte stehen neben den Aggregaten auch Informationen zu einzelnen

⁷Batra, Gulati u. a. 2013, S. 1.

⁸vgl. Wilke 2013.

⁹<http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption>,
aufgerufen 2014-08-21

¹⁰Hébrail u. a. 2010.

¹¹<http://pecanstreet.org/>

¹²<https://dataport.pecanstreet.org/>

Geräten und zum Wasserverbrauch bereit. Zusätzlich wurden einige Teilnehmende darum gebeten, in regelmäßigen Abständen Fragebögen auszufüllen, die Wissenschaftlern noch mehr Informationen liefern konnten.

Als weitere Besonderheit an diesem Datensatz lässt sich herausstellen, dass manche der Haushalte an besonderen Studien teilgenommen haben, bei denen etwa ihre Haushalte mit *intelligenten* Geräten ausgestattet wurden, die sich automatisch dann einschalteten, wenn der Strompreis gering war oder bei denen die Bewohner SMS-Nachrichten bekamen, wenn der Gesamtverbrauch auf dem Netz gerade besonders hoch war, damit sie ihren Stromverbrauch in diesen Fällen reduzieren konnten. Diese Informationen stehen in einer großen Metadatenbank zur Verfügung und ermöglichen es so zum Beispiel auch, zu sehen, welche Methoden sich in Zukunft nutzen lassen, um den Gesamtverbrauch besser zu verteilen und Spitzenlastzeiten zu entlasten.

Da bei dem Pecan Street Projekt eine sehr große Menge von Daten gesammelt wurde, stehen diese nicht gemeinsam zum Download bereit, sondern die Wissenschaftler erhalten direkten Zugang auf die Datenbank. Dies ermöglicht es, in der Metadatenbank gezielt nach Gebäuden zu suchen, die den eigenen Ansprüchen entsprechen (zum Beispiel eine besonders lange Messdauer), und dann gezielt diese Daten abzurufen.

REDD

Der REDD Datensatz^{13, 14} umfasst 10 Gebäude. Der längste gemessene Zeitraum umfasst jedoch nur 19 Tage und war daher nicht für den Anspruch dieser Arbeit geeignet. Die Daten werden für jedes Gebäude in einer hochfrequenten Auflösung als Rohdaten und in einer geringeren Auflösung zur direkten Nutzung angeboten. Zusätzlich zu den Daten werden Software Tools bereit gestellt, um mit den Daten arbeiten zu können.

¹³Kolter und Johnson 2011.

¹⁴<http://redd.csail.mit.edu/>

Smart*

Im Smart* Projekt^{15,16} finden sich drei Haushalte, deren Daten über einen Zeitraum von bis zu 90 Tagen erfasst wurden, und aggregierte Daten von über 400 Haushalten, mit einer Erfassungsdauer von jeweils 24 Stunden. In den drei Haushalten wurden unterschiedliche Dinge gemessen. Alle enthalten Daten über Temperatur sowohl innerhalb als auch außerhalb der Gebäude, zwei der Datensätze erhalten aggregierte Werte und einer enthält einzeln bemessene Geräte. Zusätzlich beinhaltet einer der Datensätze Informationen darüber, ob Türen geöffnet oder geschlossen wurden und ein anderer Daten über aus Wind- und Solarenergie generierten Strom.

SustData

Auf der Insel Maderia wurden 50 Haushalte über eine Dauer von bis zu 511 Tagen im SustData Projekt bemessen.^{17,18} Das eingesetzte System war in der Lage, den Gesamtstromverbrauch zu messen und den Nutzern Feedback über ihren Stromverbrauch zu geben. Ziel des Projektes war es, die Nutzer zu animieren, weniger Strom zu verbrauchen und nachhaltiger mit den Ressourcen umzugehen. Die Datensätze, die für diese Arbeit gesucht wurden, sollten jedoch so nah an einem durchschnittlichen (Büro-)Gebäude sein wie möglich. Hier ist davon auszugehen, dass die meisten Gebäude nicht mit Technologien versehen sind, die zu einem nachhaltigeren Handeln auffordern. Aus diesem Grund wurde der Datensatz - trotz der langen Zeitspanne - nicht genutzt.

¹⁵<http://traces.cs.umass.edu/index.php/Smart/Smart>

¹⁶Barker u. a. 2012.

¹⁷Pereira u. a. 2014.

¹⁸<http://aveiro.m-iti.org:3000/>

UK Dale

Der UK Dale Datensatz^{19,20} beinhaltet Daten von fünf Gebäuden mit bis zu 655 Tagen Laufzeit. Die Daten wurden dabei in einer - im Vergleich zu den anderen Datensätzen sehr hohen - Auflösung von 16 kHz als Aggregat und 1/6 Hz für einzelne Verbraucher gemessen. Auch das im UK Dale Projekt eingesetzte Messsystem wurde als Open Source Lösung veröffentlicht. Zu den fünf Gebäuden, für die Daten vorliegen, steht eine Reihe an Metadaten wie zum Beispiel Baujahr und Heizungsart bereit, die weitere Rückschlüsse auf die Energieeffizienz zulässt.²¹

Universitäten

Die Anzahl an öffentlichen Datensätzen, die Stromdaten beschreiben, ist in den vergangenen Jahren (etwa seit 2011) stets gestiegen, die meisten der Datensätze beschreiben jedoch Wohngebäude. Es gibt jedoch auch zwei öffentliche Datensätze, die kommerzielle Gebäude beschreiben. Bei beiden handelt es sich um Universitätsgebäude. Die Datensätze werden im folgenden Abschnitt vorgestellt.

BERDS

Das BERkeley EnerGy Disaggregation Data Set (BERDS)^{22, 23} umfasst drei Stromkreise aus einer Universität, die über eine Dauer von 7 Tagen erfasst wurden. Laut der Untersuchung stammen die Daten aus dem openBMS System,²⁴ zum Zeitpunkt dieser Arbeit

¹⁹<http://data.ukedc.rl.ac.uk/simplebrowse/edc/efficiency/residential/EnergyConsumption/Domestic>, aufgerufen 2015-08-21

²⁰Kelly und Knottenbelt 2015.

²¹<http://www.nature.com/articles/sdata20157/tables/1>, aufgerufen 2015-08-21

²²Maasoumy u. a. 2013.

²³<http://www.eecs.berkeley.edu/~maasoumy/>, aufgerufen 2015-08-21, verfügbar über den Link *Download BERDS*

²⁴<http://berkeley.openbms.org/map/>

war es aber leider nicht möglich, aus dieser Quelle mehr Daten herunterzuladen, und die Zeitspanne von 7 Tagen war nicht ausreichend für eine Witterungsanalyse.

COMBED

Im COMBED Projekt^{25, 26} wurde eine Universität in Neu Delhi in Indien untersucht. Dabei wurden mit einer Auflösung von 30 Sekunden mehrere Gebäude bemessen. Die Daten für die Gebäude sind nach Etagen unterteilt, enthalten aber auch Aggregate für die einzelnen Gebäude und Informationen über Aufzüge - ein Detail, das sonst in keinem anderen öffentlichen Datensatz verfügbar ist. Die Daten stehen als CSV oder als hdf Datei bereit, die direkt in das von den Autoren entwickelte NILM Analyse Toolkit *nilmtk*²⁷ integriert werden kann.

²⁵Batra, Parson u. a. 2014.

²⁶<http://combed.github.io/>

²⁷<http://nilmtk.github.io/>

Literatur

- Barker, Sean, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy und Jeannie Albrecht (2012). „Smart*: An open data set and tools for enabling research in sustainable homes“. In: *SustKDD, August*. URL: <http://lass.cs.umass.edu/~sbarker/research/pdf/sustkdd12-smart.pdf> (besucht am 14.07.2015).
- Batra, Nipun, Manoj Gulati, Amarjeet Singh und Mani B. Srivastava (2013). „It’s Different: Insights into Home Energy Consumption in India“. In: *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. BuildSys’13. New York, NY, USA: ACM, 3:1–3:8. ISBN: 978-1-4503-2431-1. DOI: 10.1145/2528282.2528293. URL: <http://doi.acm.org/10.1145/2528282.2528293>.
- Batra, Nipun, Oliver Parson, Mario Berges, Amarjeet Singh und Alex Rogers (2014). „A comparison of non-intrusive load monitoring methods for commercial and residential buildings“. In: *arXiv:1408.6595 [cs]*. arXiv: 1408.6595. URL: <http://arxiv.org/abs/1408.6595> (besucht am 14.07.2015).
- Beckel, Christian, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake und Silvia Santini (2014). „The ECO Data Set and the Performance of Non-intrusive Load Monitoring Algorithms“. In: *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. BuildSys ’14. New York, NY, USA: ACM, S. 80–89. ISBN: 978-1-4503-3144-9. DOI: 10.1145/2674061.2674064. URL: <http://doi.acm.org/10.1145/2674061.2674064> (besucht am 14.07.2015).
- Bilan prévisionnel* (2014). Französisch. URL: <http://www.rte-france.com/fr/article/bilan-previsionnel> (besucht am 30.08.2015).

- Bundesregierung | Energie"tatort" Büro – Wo kann ich am Schreibtisch Energie sparen? (2015). URL: <http://www.bundesregierung.de/Content/DE/Magazine/MagazinVerbraucher/002/t4-energietatort-buero.html> (besucht am 02.07.2015).
- Clemens, Konstantin (2015). „Geocoding with OpenStreetMap Data“. In: *GEOProcessing 2015*, S. 10. URL: http://www.researchgate.net/profile/Bruno_M_Meneses/publication/280575974_Water_Quality_Impact_Assessment_of_Land_Use_and_Land_Cover_Changes._A_dynamic_IT_model_for_territorial_integrated_management/links/55bb739208aed621de0d9692.pdf#page=11 (besucht am 20.08.2015).
- Energieeffizienz, Nationaler Aktionsplan (2014). „Mehr aus Energie machen“. In: URL: https://ssl.vdb-info.de/media/file/3977.Nationaler_Aktionsplan_Energieeffizienz.pdf (besucht am 03.08.2015).
- Eric Jones, Travis Oliphant, Pearu Peterson u. a. (2001). *SciPy.org — SciPy.org*. URL: <http://www.scipy.org/index.html> (besucht am 20.08.2015).
- Fuchs-Kittowski, Klaus und Volker Wohlgemuth (2011). „Umweltinformatik und Umweltforschung in ihrer Institutionalisierung und Interdisziplinarität“. ger. In: *Wissenschaftsforschung 2010*. Berlin: wvb, Wiss. Verl, S. 301. ISBN: 978-3-86573-590-4. URL: http://www.wissenschaftsforschung.de/JB10_99-152.pdf (besucht am 16.08.2015).
- Hammering Usernames* (2015). URL: https://www.facebook.com/note.php?note_id=96390263919 (besucht am 20.08.2015).
- Hébrail, Georges, Bernard Huguency, Yves Lechevallier und Fabrice Rossi (2010). „Exploratory Analysis of Functional Data via Clustering and Optimal Segmentation“. In: *Neurocomputing* 73.7-9. arXiv: 1004.0456, S. 1125–1141. ISSN: 09252312. DOI: 10.1016/j.neucom.2009.11.022. URL: <http://arxiv.org/abs/1004.0456> (besucht am 15.07.2015).
- Hunter, J.D. (2007). „Matplotlib: A 2D Graphics Environment“. In: *Computing in Science Engineering* 9.3, S. 90–95. ISSN: 1521-9615. DOI: 10.1109/MCSE.2007.55.
- Kelly, Jack und William Knottenbelt (2015). „The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes“. In: *Scientific Data* 2, S. 150007. ISSN: 2052-4463. DOI: 10.1038/sdata.2015.7. URL: <http://www.nature.com/articles/sdata20157> (besucht am 14.07.2015).

- Kolter, J. Zico und Matthew J. Johnson (2011). „REDD: A public data set for energy disaggregation research“. In: *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, San Diego, CA. Bd. 25. Citeseer, S. 59–62. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.5796&rep=rep1&type=pdf> (besucht am 14.07.2015).
- Maasoumy, Mehdi, B. Sanandaji, Kameshwar Poolla und Alberto Sangiovanni Vincentelli (2013). „Berds-berkeley energy disaggregation data set“. In: *Proceedings of the Workshop on Big Learning at the Conference on Neural Information Processing Systems (NIPS)*.
- Makonin, Stephen, Fred Popowich, Lyn Bartram, Bob Gill und Ivan V. Bajic (2013). „AMPds: A Public Dataset for Load Disaggregation and Eco-Feedback Research“. In: *Proceedings of the 2013 IEEE Electrical Power and Energy Conference (EPEC)*.
- McKinney, Wes (2010). „Data Structures for Statistical Computing in Python“. In: S. 51–56. URL: <http://conference.scipy.org/proceedings/scipy2010/mckinney.html> (besucht am 20.08.2015).
- Mittag, Hans-Joachim (2014). *Statistik*. Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-54386-9 978-3-642-54387-6. URL: <http://link.springer.com/10.1007/978-3-642-54387-6> (besucht am 23.08.2015).
- Monacchi, Andrea, Dominik Egarter, Wilfried Elmenreich, Salvatore D’Alessandro und Andrea M. Tonello (2014). „GREEND: An Energy Consumption Dataset of Households in Italy and Austria“. In: *arXiv:1405.3100 [cs]*. arXiv: 1405.3100. URL: <http://arxiv.org/abs/1405.3100> (besucht am 14.07.2015).
- Neal Lott, Russell Vose (2008). „The Integrated Surface Database: Partnerships and Progress“. In:
- Nielsen, Jakob (1994). „Usability Inspection Methods“. In: *Conference Companion on Human Factors in Computing Systems*. CHI ’94. New York, NY, USA: ACM, S. 413–414. ISBN: 978-0-89791-651-6. DOI: 10.1145/259963.260531. URL: <http://doi.acm.org/10.1145/259963.260531> (besucht am 20.08.2015).
- Page, Bernd und Lorenz M. Hilty (1995). „Umweltinformatik“. In: *Informatikmethoden für Umweltschutz und Umweltforschung*. München/Wien 2. URL: <http://www.ulb.tu-darmstadt.de/tocs/46524487.pdf> (besucht am 18.08.2015).

- Paris, James (2006). „A framework for non-intrusive load monitoring and diagnostics“. eng. Thesis. Massachusetts Institute of Technology. URL: <http://dspace.mit.edu/handle/1721.1/37081> (besucht am 31.08.2015).
- Pereira, Lucas, Filipe Quintal, Rodolfo Goncalves und Nuno Jardim Nunes (2014). „Sust-Data: A Public Dataset for ICT4S Electric Energy Research“. en. In: Atlantis Press. ISBN: 978-94-6252-022-6. DOI: 10.2991/ict4s-14.2014.44. URL: <http://www.atlantis-press.com/php/paper-details.php?id=13462> (besucht am 14.07.2015).
- Pérez, F. und B.E. Granger (2007). „IPython: A System for Interactive Scientific Computing“. In: *Computing in Science Engineering* 9.3, S. 21–29. ISSN: 1521-9615. DOI: 10.1109/MCSE.2007.53.
- Peter Hofer (2013). *Einfluss von Temperatur- und Globalstrahlungsschwankungen auf den Energieverbrauch der Gebäude*. Techn. Ber.
- Verordnung über energiesparenden Wärmeschutz und energiesparende Anlagentechnik bei Gebäuden (Energieeinsparverordnung - EnEV)* (2007).
- Walt, S. van der, S.C. Colbert und G. Varoquaux (2011). „The NumPy Array: A Structure for Efficient Numerical Computation“. In: *Computing in Science Engineering* 13.2, S. 22–30. ISSN: 1521-9615. DOI: 10.1109/MCSE.2011.37.
- Weber, Lukas (2002). *Energie in Bürogebäuden: Verbrauch und energierelevante Entscheidungen*. vdf Hochschulverlag AG. URL: https://books.google.de/books?hl=de&lr=&id=S30bG6MRf4C&oi=fnd&pg=PA13&dq=Energie+in+B%C3%83%C2%BCrogeb%C3%83%C2%A4uden&ots=qDEt8n0017&sig=Pq4ZDfTrssx7w0_Pun3XVP2ICmA (besucht am 31.08.2015).
- Weglage, Andreas, Hrsg. (2010). *Energieausweis - das große Kompendium: Grundlagen - Erstellung - Haftung ; mit 17 Tabellen*. ger. 3. Aufl. Praxis Bauwesen. Wiesbaden: Vieweg + Teubner. ISBN: 978-3-8348-0862-2.
- Wilke, Sibylle (2013). *Energieverbrauch der privaten Haushalte*. de. Text. URL: <http://www.umweltbundesamt.de/daten/private-haushalte-konsum/energieverbrauch-der-privaten-haushalte> (besucht am 31.08.2015).
- Wosnitza, Franz und Hans Gerd Hilgers (2012). *Energieeffizienz und Energiemanagement*. de. Wiesbaden: Vieweg+Teubner Verlag. ISBN: 978-3-8348-1941-3 978-3-8348-8671-2.

URL: <http://link.springer.com/10.1007/978-3-8348-8671-2> (besucht am 16.08.2015).