

Übersicht

Simon Waczowicz*, Stefan Klaiber, Peter Bretschneider, Irina Konotop, Dirk Westermann, Markus Reischl und Ralf Mikut

Data Mining zur Analyse der Auswirkungen von Preissignalen auf Haushaltsstromkunden

Data mining to analyse the effects of price signals on household electricity customers

Zusammenfassung: Dieser Beitrag widmet sich der Analyse der Auswirkungen von Preissignalen auf das Verbrauchsverhalten von Haushaltsstromkunden. Er schlägt einen systematischen Auswerteprozess vor, der die Datenvorverarbeitung, verschiedene Aggregationsschritte, die Analyse mit Clusterverfahren (für preisbeeinflusste Typtage bei einzelnen oder aggregierten Haushalten) und die Analyse einzelner Abtastzeitpunkte umfasst. Dieser Auswerteprozess wird auf den Olympic Peninsula Project Datensatz angewendet.

Schlüsselwörter: Data Mining, Demand Response, Anreizsignal, Haushaltsstromkunden.

Abstract: The article is dedicated to the analysis of the effects of different price signals on the consumption behavior of household electricity customers. It proposes a systematic analysis process consisting of data preprocessing, different aggregation steps, the analysis with clustering methods and the analysis of time samples. This analysis process was applied to the Olympic Peninsula Project database.

Keywords: Data mining, demand response, price incentives, household electricity customers.

***Korrespondenzautor:** Simon Waczowicz, Institut für Angewandte Informatik (IAI), Karlsruher Institut für Technologie (KIT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, E-Mail: simon.waczowicz@kit.edu

Markus Reischl, Ralf Mikut: Institut für Angewandte Informatik (IAI), Karlsruher Institut für Technologie (KIT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen

Stefan Klaiber, Peter Bretschneider: Institutsteil Angewandte Systemtechnik (AST) des Fraunhofer IOSB, Am Vogelherd 50, D-98693 Ilmenau

Irina Konotop, Dirk Westermann: Fachgebiet Elektrische Energieversorgung, Technische Universität Ilmenau, Gustav-Kirchhoff-Straße 1, D-98684 Ilmenau

DOI 10.1515/auto-2014-1108

Eingang 17. April 2014; angenommen 8. August 2014

1 Einleitung

Spätestens mit dem Energiekonzept der Bundesregierung, beschlossen im September 2010, findet in Deutschland ein gesellschaftliches und politisches Umdenken statt, welches dazu führt, die Stromerzeugung in konventionellen Kraftwerken zunächst durch erneuerbare Energiequellen zu ergänzen, später komplett durch sie zu ersetzen. Der Transformationsprozess, der unter dem Begriff Energiewende zusammengefasst ist, macht umfangreiche und weitreichende Veränderungen im deutschen Energieversorgungssystem notwendig [1]. Die zentrale Rolle nimmt dabei die Integration großer Strommengen aus erneuerbaren Energiequellen („Windstrom“, „Solarstrom“) ein. Der Anteil der erneuerbaren Energien an der Brutto-Stromerzeugung betrug im Jahr 2013 bereits 23.9% und soll bis zum Jahr 2030 auf 50% erhöht werden [2, 3]. Für die Integration der erneuerbaren Energien in das existierende Stromversorgungssystem Deutschlands bedarf es vielfältiger Lösungsansätze, welche sich gegenseitig nicht zwangsläufig ausschließen und deshalb kombiniert betrachtet werden müssen.

Eine charakteristische Eigenschaft der erneuerbaren Energien ist die schwankende und nicht bedarfsorientierte Stromproduktion. Stromspeicher sind eine Möglichkeit, einen Ausgleich zwischen Stromproduktion und Stromverbrauch herzustellen. Allerdings sind zukünftig deutlich höhere Speicherkapazitäten vonnöten [4]. Ein ergänzender Ansatz ist die Flexibilisierung des Stromverbrauchs mittels Steuer- (Demand Side Management, DSM) und Preissignalen (Demand Response, DR) [5]. Voraussetzung hierfür ist die IT-basierte kommunikative Vernetzung, Steuerung und Interaktion von Stromerzeugern, direkt und indirekt steuerbaren Stromverbrauchern und In-

frastrukturkomponenten [6, 7]. In dem sogenannten *Smart Grid* werden Leistungen einzelner Verbraucher mit intelligenten Verbrauchszählern (*Smart Meter*) erfasst, meist mit Abtastzeiten von 15 Minuten.

Zur Modellierung von *Smart Meter* Datensätzen werden oftmals dynamische Regressionsmodelle angelehrt, siehe z.B. [8–10]. Erste Ansätze zur Untersuchung mit Data-Mining-Methoden sind in [11, 12] gezeigt. In [13] erfolgt eine Klassifikation von individuellen Haushaltslastkurven mit künstlichen neuronalen Netzen (Kohonenetze, engl.: *self-organizing maps* (SOM)). In [14–18] kommen diverse Clusterverfahren zur Analyse von *Smart Meter* Messdaten zum Einsatz. Die Anwendung verschiedener Data-Mining-Methoden zur Datenanalyse und Modellierung bei *Smart Meter* Messdaten unter Berücksichtigung zeitvariabler Stromtarife wurde im Rahmen des RESIDENS-Projektes durchgeführt [5]. In [19, 20] sind die Auswirkungen einer aktiven Verbrauchsbeeinflussung durch Echtzeit-Verbrauchsmonitoring inkl. anschließender Datenanalyse demonstriert.

Allerdings fehlt bislang eine Anleitung zur systematischen Vorgehensweise zur Untersuchung solcher Datensätze, jede der bisherigen Arbeiten fokussiert lediglich auf einzelne Aspekte. Zudem ist die Frage ungeklärt, wie Lastverschiebungen einzelner heterogener Verbraucher oder von heterogenen Verbrauchergruppen miteinander verglichen werden können.

Der vorliegende Beitrag schlägt ein Data-Mining-Verfahren mit sieben Teilschritten vor, das die Auswirkungen von unterschiedlichen Preissignalen auf das Verbrauchsverhalten von Haushaltsstromkunden in einem *Smart Grid* systematisch analysiert. Die Anwendung der Analysemethoden findet exemplarisch an einem Datensatz (*Olympic Peninsula Project* [21]) statt.

Der Beitrag ist wie folgt gegliedert. Eine systemtechnische Betrachtung und Einordnung des DR- bzw. DSM-Prozesses erfolgt in Kapitel 2. Kapitel 3 befasst sich mit dem vorgeschlagenen Data-Mining-Verfahren zur Analyse von preisbeeinflussten Stromverbrauchsdaten auf Haushaltsebene. In Kapitel 4 wird der untersuchte Datensatz beschrieben. Dabei werden die Rahmenbedingungen der Feldstudie sowie die Haushaltstechnik skizziert. Wesentliche Analyseergebnisse sowie Ergebnisse der Modellierung mit dynamischen Regressionsmodellen stehen in Kapitel 5. Der Beitrag schließt mit einer Zusammenfassung und Bewertung der gefundenen Ergebnisse.

2 DSM-Konzept

Die Flexibilisierung des Stromverbrauchs von Industrie- und Haushaltsstromkunden wird als ein Mittel zur Synchronisation von Stromerzeugung und -verbrauch gesehen. Alle Maßnahmen zur Beeinflussung des elektrischen Bedarfs werden als DSM bezeichnet [22]. Eine gezielte Schaltung von Verbraucherlasten, beispielsweise die Abschaltung einer Kühlanlage bei einem Industriekunden durch den Netzbetreiber, fällt unter den Begriff des direkten DSM, das im vorliegenden Beitrag unberücksichtigt bleibt. Wird durch ein zeitvariables Anreizsignal versucht, eine Änderung des Verbrauchsmusters zu bewirken, handelt es sich um indirektes DSM. Im Folgenden wird indirektes DSM als Demand Response bezeichnet. DR-Maßnahmen finden i.d.R. in Verteilungsnetzen der Niederspannungsebene statt und unterliegen verschiedenen Interessen und Einflüssen. Dort agierende Akteure (Bilanzkreisverantwortliche, Lieferanten und Netzbetreiber) haben z.T. sehr unterschiedliche Zielstellungen (netzgeführte, ökonomische), was sich wiederum in der Beschaffenheit des Anreizsignals zeigt [23]. Die Vermeidung von Spitzen- (*Peak Clipping*) oder Niedriglasten (*Valley Filling*) kann als Ziel einer Lastbeeinflussung gesehen werden. Bei einer Lastverschiebung (*Load Shifting*) kommt es zur Verlagerung der Lasten von Spitzen- zu Niedriglastzeiten. Alle drei genannten Maßnahmen führen zu einer Glättung des Lastprofils.

Für Datenanalyseverfahren, Prognose- und Modellierungsmethoden ist eine systemtechnische Einordnung des DR-Prozesses ratsam [23]. Abbildung 1 verdeutlicht die komplexen Zusammenhänge des preisbeeinflussten Verbrauchsprozesses. Für die Planung einer DR-Maßnahme spielen sowohl Prognosen exogener Einflussgrößen des elektrischen Energiesystems $\hat{u}_{EES}[k]$, als auch die Prognose der unbeeinflussten Verbraucherlast $\hat{P}_U[k]$ eine Rolle (siehe Kasten „Planung“). Der Einfluss der modifizierten Verbraucherlasten auf die Preis-/Tariffbildung führt zu einem rückgekoppelten System. Die preisbeeinflusste Verbraucherlast $P[k]$ entsteht durch die Addition von zwei Signalanteilen (Abbildung 1):

- einer nicht preisbeeinflussten Verbraucherlast $P_U[k]$, welche nur durch meteorologische, tageszeitabhängige, saisonale und singuläre Faktoren ($u[k]$) beeinflusst ist (siehe Kasten „Prozess Verbraucher“) und
- einem durch Preissignale beeinflussten Anteil der Verbraucherlast $P_R[k]$ (siehe Kasten „DSM-Prozess“).

Die systemtheoretische Modellierung des preisbeeinflussten Verbrauchsverhaltens, z.B. durch ein nichtlineares Differentialgleichungssystem wie in [10], vervollständigt den

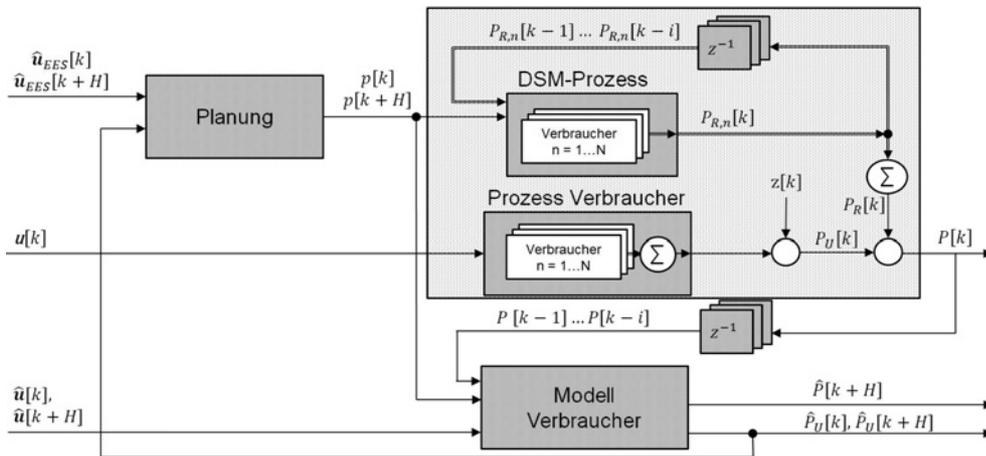


Abbildung 1: Beeinflusster Verbrauchsprozess mit Verbrauchermodell und Planung (modifiziert nach [23]): H – Vorhersagehorizont, n – Haushaltsindex, $p[k]$ – Preis-/Anreizsignal, \hat{x} – Schätzung für x , $u[k]$ – Eingangsgrößen, $u_{EES}[k]$ – exogene Einflussgrößen des elektrischen Energiesystems, $P[k]$ – (preis-) beeinflusste Verbraucherlast, $P_R[k]$ – (preis-) beeinflusster Anteil der Verbraucherlast, $P_U[k]$ – beeinflusster Anteil der Verbraucherlast, $z[k]$ – Störgröße.

beeinflussten Verbrauchsprozess (siehe Kasten „Modell Verbraucher“), soll aber nicht Bestandteil dieses Beitrages sein.

3 Data-Mining-Methoden

Für das beschriebene DSM-Konzept ist eine genaue Kenntnis des Verbrauchsverhaltens in Abhängigkeit verschiedener Anreizsignale (insb. Tarife) wünschenswert. Konkret umfasst das die folgenden Fragestellungen:

1. Inwiefern verändern zeitvariable Stromtarife ein nicht durch Strompreise beeinflusstes Verbrauchsverhalten?
2. Welche Unterschiede bezüglich einer zeitlichen Lastverschiebung ergeben sich bei verschiedenen Stromtarifstrukturen?
3. Gibt es saisonale, wochentags- oder situationsbedingte Schwankungen für eine DR-Maßnahme, z.B. für Fälle mit ungewöhnlichen Anreizsignalen?
4. Reagieren alle Haushalte ähnlich auf Preissignale oder gibt es wesentliche Unterschiede?

Bislang werden bei der Modellierung des preisbeeinflussten Verbrauchsverhaltens meist dynamische Regressionsmodelle eingesetzt. Laut [8] eignen sich $ARX(n_a, n_b, n_k)$ -Modelle¹ zur Schätzung des Zusammenhangs zwischen Preissignal $p[k]$ als Eingangsgröße und preisbeeinfluss-

Tabelle 1: Unterschiedliche Datensätze mit Anzahl der Datentupel (N_D) für unterschiedliche Fälle, siehe auch Schritt 2, 3 und 7 des Auswerteprozesses.

	einzelne Haushalte (H)	alle Haushalte einer Tarifgruppe (T_i)
Gesamtzeitreihen (G)	Fall GH ($N_D = N$)	Fall GT ($N_D = N_{T_i}$)
Variable Zeitreihenlänge (V)	Fall VH ($N_D = N$)	Fall VT ($N_D = N_{T_i}$)
Tageszeitreihen (D)	Fall DH ($N_D = N \cdot D$)	Fall DT ($N_D = N_{T_i} \cdot D$)
Abtastzeitpunkte (S)	Fall SH ($N_D = N \cdot D \cdot K$)	Fall ST ($N_D = N_{T_i} \cdot D \cdot K$)

tem Anteil der Verbraucherlast $P_R[k]$ als Ausgangsgröße

$$P_R[k] + \sum_{i_a=1}^{n_a} a_{i_a} P_R[k-i_a] = \sum_{i_b=1}^{n_b} b_{i_b} p[k-n_k-i_b+1] + e[k], \quad (1)$$

wobei n_a , n_b die Ordnungen des Regressionsmodells, n_k die Totzeit und $e[k]$ ein gaußverteiltes weißes Rauschen bezeichnen. [9] schätzt ein *Finite Impulse Response* (FIR) Modell für die Laständerung als Reaktion auf Preisänderungen und externe Größen (Wetter). Die Parameter dieser Modelle werden in der Regel für komplette Tarifgruppen und längere Zeiträume geschätzt, die Modelle sind folglich hauptsächlich zur Beantwortung der Fragen 1-2 geeignet.

Als Ergänzung sollen weitere Data-Mining-Methoden [24, 25] eingesetzt werden, um auch für die Fragen 3-4 zusätzliche Einsichten zu gewinnen. Die Herausforderung besteht darin, passende Data-Mining-Verfahren und Bewertungsmaße auszuwählen sowie diese geeignet mitein-

¹ *AutoRegressive Model with eXogeneous Inputs*

ander zu verknüpfen [26]. Im Folgenden wird davon ausgegangen, dass Rohdaten in Form von Zeitreihen mit Zeitstempeln, elektrischer Leistung und aktuellen Preisen auf Basis einzelner Haushalte vorliegen (Fall **GH** in Tabelle 1).

Um die oben formulierten Fragestellungen zu beantworten, wird in diesem Beitrag ein systematischer Auswerteprozess mit sieben Schritten vorgeschlagen, der auf bekannten Data-Mining-Verfahren basiert:

Schritt 1: Datenvorverarbeitung. Fehlende und nicht plausible Werte (Ausreißer) werden durch Marker für fehlende Werte (z.B. NaN – not a number) oder über zu große Anstiege detektiert. Fehlende Werte weniger als 30 Minuten werden interpoliert (Fall **GH** in Tabelle 1). Bei größeren Bereichen mit fehlenden Werten ist es ratsam, die Gesamtzeitreihen **GH** und **GT** in kleinere Zeitreihen **VH** bzw. **VT** mit variabler Zeitreihenlänge zu unterteilen. Die beiden Datensätze **VH** und **VT** stellen dabei die Datensätze mit der größtmöglichen Zeitreihenlänge ohne Bereiche mit fehlenden Messwerten dar.

Schritt 2: Tageszerlegung. Mithilfe des Zeitstempels werden aus den Jahreszeitreihen **GH** und **GT** Tageskurven generiert. Z.B. wird die Jahreslastkurve $P_n[k]$ eines Haushalts n in D ($d = 1, \dots, D$) Tageszeitreihen $P_{d,n}[k]$ zerlegt. Tageskurven mit mehr als drei Prozent fehlenden Werten werden gelöscht (Fall **DH** in Tabelle 1).

Schritt 3: Zeitreihenaggregation. Die Ergebnisse von Schritt 1 und 2 werden jeweils über alle N_{T_i} Haushalte mit gleichem Tarif T_i durch Mittelung aggregiert, z.B. mit

$$\bar{P}_d^{T_i}[k] = \frac{1}{N_{T_i}} \sum_{n=1}^{N_{T_i}} P_{d,n}[k], \quad (2)$$

für den Fall **DT** in Tabelle 1.

Schritt 4: Tagesnormierung. Für die Erkennung von Lastverschiebungen ist die Tagesnormierung der Zeitreihen durch Division mit dem Tagesmittelwert sinnvoll:

$$P_{\text{norm},d,n}[k] = \frac{P_{d,n}[k]}{\frac{1}{K} \sum_{k=1}^K P_{d,n}[k]}. \quad (3)$$

Die daraus resultierenden Fälle werden mit dem Index "norm" gekennzeichnet. Z.B. wird damit aus Fall **DH** der Fall **DH_{norm}**.

Schritt 5: Clusteranalyse. Zur Ermittlung von verschiedenen Typtagen bietet sich die Verwendung von Clusterverfahren an, um ähnliche Tagesverläufe zu gruppieren. Dieses unüberwachte Verfahren kommt ohne die Vorgabe von Wochentagen, Jahreszeiten, Tarifen und Haushaltsarten aus und ist somit in der Lage, Typtage unvoreingenommen zu identifizieren. Aufgrund der hohen Variabilität der Energiedaten ist nicht von der Existenz scharfer Cluster, sondern eher von graduellen Übergängen zwi-

schen verschiedenen Clustern auszugehen. Deshalb wird unter der Vielzahl von Clusterverfahren (Übersicht siehe [27]) anstelle von scharfen Clusterverfahren (z.B. k-means) wie in [28] das Fuzzy-C-Means-Verfahren verwendet [29, 30], um Clusterzentren

$$\mathbf{V}_{K \times C} = (\mathbf{v}_1 \quad \dots \quad \mathbf{v}_C) = \begin{pmatrix} v_1[1] & \dots & v_C[1] \\ \vdots & \ddots & \vdots \\ v_1[K] & \dots & v_C[K] \end{pmatrix} \quad (4)$$

(K Länge Zeitreihe, C Anzahl Cluster) durch Lösen des Optimierungsproblems

$$J_q(\boldsymbol{\mu}, \mathbf{V}) = \sum_{n=1}^N \sum_{c=1}^C (\mu_{n,c})^q \cdot d_c^2(\mathbf{x}_n, \mathbf{v}_c) \rightarrow \min_{\boldsymbol{\mu}, \mathbf{V}}, \quad (5a)$$

$$\sum_{c=1}^C \mu_{n,c} = 1, \forall n = 1 \dots N, \text{ mit } \mu_{n,c} \geq 0, \quad (5b)$$

$$\sum_{n=1}^N \mu_{n,c} > 0, \forall c = 1 \dots C \quad (5c)$$

zu finden. Als \mathbf{x}_n werden die Leistungszeitreihe eines Haushalts (Fall **DH**), deren Tagesnormierung (**DH_{norm}**), die mittlere Leistung einer Tarifgruppe (**DT**) oder deren Tagesnormierung (**DT_{norm}**) eingesetzt. Für eine große Zugehörigkeit $\mu_{n,c}$ eines Datentupels $n = 1, \dots, N$ zu einem Clusterzentrum $c = 1, \dots, C$ ist eine kleine Distanz d_c optimal und umgekehrt. Der Fuzzifizierungsgrad q sorgt für „harte“ ($q \rightarrow 1$) oder weichere Zugehörigkeiten $q \rightarrow \infty$. Die Literatur empfiehlt oftmals einen Fuzzifizierungsgrad von $q = 2$ [30], allerdings führt das zu nahezu gleichverteilten Zugehörigkeiten. Deswegen werden hier, wie in [31] empfohlen, Werte zwischen $q = 1.05$ und 1.3 getestet und schließlich $q = 1.1$ gewählt.

Die Zeitreihen werden mit der Euklidischen Distanz

$$d_c(\mathbf{x}_n, \mathbf{v}_c) = \sqrt{\sum_{k=1}^K (x_n[k] - v_c[k])^2} \quad (6)$$

verglichen, die die Wertedifferenzen zum jeweils gleichen Zeitpunkt aggregiert. Die Ergebnisse mit und ohne Normierung in Schritt 4 unterscheiden sich, weil ohne die Normierung ein variierender Tagesstromverbrauch (sowohl durch eine Verschiebung um einen Gleichanteil als auch durch unterschiedliche Amplituden der Tagesschwankungen) einen großen Einfluss ausübt. Dieser Einfluss wird durch die Tagesnormierung fast vollständig unterdrückt.

Anschließend folgt eine iterative Berechnung von

$$\mu_{n,c} = \frac{[d_c^2(\mathbf{x}_n, \mathbf{v}_c)]^{\frac{1}{1-q}}}{\sum_{i=1}^C [d_i^2(\mathbf{x}_n, \mathbf{v}_i)]^{\frac{1}{1-q}}} \quad (7)$$

$$\mathbf{v}_c = \frac{\sum_{n=1}^N (\mu_{n,c})^q \mathbf{x}_n}{\sum_{n=1}^N (\mu_{n,c})^q}. \quad (8)$$

Die Clusterzentren \mathbf{v}_c werden dann als Zeitreihen für Typ-tage dargestellt und bleiben somit interpretierbar.

Für die Bestimmung der Clusteranzahl existieren verschiedene Maße, deren Vor- und Nachteile stark umstritten sind [32]. Exemplarisch wird hier der Trennungsgrad

$$S(\boldsymbol{\mu}, C) = \frac{J_q(\boldsymbol{\mu}, \mathbf{V})}{C \cdot \min_{c,i=1 \dots C, c \neq i} (d^2(\mathbf{v}_c, \mathbf{v}_i))} \quad (9)$$

verwendet. Die Literatur empfiehlt eine Berechnung mit steigender Clusteranzahl und die Wahl des ersten lokalen Minimums [30] für die Wahl von C . Allerdings zeigte sich in Voruntersuchungen, dass so tendenziell $C = 2$ bevorzugt wird. Deshalb wurde das erste lokale Minimum von $S(\boldsymbol{\mu}, C)$, $C > 2$ gewählt. Je nach Anwendung ist ein systematischer Vergleich verschiedener Maße zu empfehlen.

Je nach untersuchten Zeitreihen ergeben sich somit Typ-tage für den Energieverbrauch einzelner Haushalte (Fall **DH**), ganzer Tarifgruppen (**DT**) bzw. die jeweiligen tagenormierten Kurven $\mathbf{DH}_{\text{norm}}$ bzw. $\mathbf{DT}_{\text{norm}}$.

Schritt 6: Bewertung der Zusammenhänge von Clustern zu Werktagen, Monaten und Tarifen. Die Clusteranalyse des vorherigen Schrittes sagt zunächst nichts darüber aus, wie stark die Zusammenhänge zwischen Clustern und Tarifen sind. Eine manuelle Analyse ist hier aufwändig und schwer quantifizierbar. Für die Analyse des Zusammenhangs zwischen Clustern und anderen Ausgangsgrößen werden deshalb Tabellen für die Verbundverteilungen (z.B. von Clustern und Tarifen) sowie informationstheoretische Maße [33] eingesetzt. Letztere basieren auf relativen Häufigkeiten als Schätzungen für Wahrscheinlichkeiten \hat{P} und stellen Maße für Zusammenhänge zwischen wertediskreten Größen dar. Die Schätzungen für die Eingangsentropie $\hat{H}(x)$ für die Cluster, die Ausgangsentropie $\hat{H}(y)$ und die Gesamtentropie $\hat{H}(x, y)$ ergeben sich aus

$$\hat{H}(x) = - \sum_{c=1}^C \hat{P}(x=c) \cdot \log_2 \hat{P}(x=c), \quad (10)$$

$$\hat{H}(y) = - \sum_{l=1}^{m_y} \hat{P}(y=l) \cdot \log_2 \hat{P}(y=l), \quad (11)$$

$$\hat{H}(x, y) = - \sum_{c=1}^C \sum_{l=1}^{m_y} \hat{P}(x=c \cap y=l) \cdot \log_2 \hat{P}(x=c \cap y=l). \quad (12)$$

Als Eingang x wird immer das Cluster mit der maximalen Clusterzugehörigkeit $x=c = \arg \max \mu_c$ verwendet. Für

den Ausgang y werden drei separate Untersuchungen für unterschiedliche Monate ($m_y = 12$), Werk-tage ja/nein ($m_y = 2$) oder Tarife durchgeführt. Die maximale Entropie tritt auf, wenn die jeweiligen wertediskreten Merkmale für x, y gleichverteilt sind und x, y stochastisch unabhängig sind. Die Entropie ist hingegen Null, wenn genau ein wertediskreter Wert die Häufigkeit Eins und alle anderen die Häufigkeit Null aufweisen.

Den Informationsfluss zwischen Ein- und Ausgangsgröße beschreibt die Transinformation

$$\hat{H}(x; y) = \hat{H}(x) + \hat{H}(y) - \hat{H}(x, y). \quad (13)$$

Die relative Transinformation pro Ausgangsentropie (bekannt aus dem ID3-Verfahren bei Entscheidungsbäumen)

$$Q = \frac{\hat{H}(x; y)}{\hat{H}(y)} \quad (14)$$

ergibt ein relatives Maß zwischen 0 (kein Zusammenhang) und 1 (aus Clustern folgt eine eindeutige Ausgangsgröße), mit dem sich die jeweilige Ausgangsgröße durch die Cluster erklären lässt. Somit können unterschiedliche Fälle und Ausgangsgrößen miteinander quantitativ verglichen werden, die jeweiligen Details können dann den x - y -Verbundverteilungen entnommen werden.

Schritt 7: Analyse einzelner Abtastzeitpunkte bezüglich der Tarifeinflüsse auf den Verbrauch. Für eine Analyse der prozentualen Energieeinsparungen eignen sich die Fälle **ST** und **SH**, für eine zeitliche Verschiebung eher die Fälle **ST_{norm}** und **SH_{norm}**, weil hier die prozentuale Einsparung bereits herausgerechnet ist. Diese Fälle resultieren jeweils aus einer Konvertierung jedes Abtastzeitpunktes der Zeitreihen von $\mathbf{DH}_{\text{norm}}$ und $\mathbf{DT}_{\text{norm}}$ in einzelne Datentupel. Danach können unterschiedliche Preise eines Tarifs in Klassen eingeteilt und die korrespondierenden Verbrauchswerte untersucht werden, also z.B. Mittelwerte und Streuungen aller Datentupel, die zu einer bestimmten Preisklasse gehören.

Die Vergleichbarkeit der Ergebnisse setzt allerdings voraus, dass entweder die gleichen Haushalte bzgl. ihres Verhaltens bei unterschiedlichen Tarifen über einen längeren Zeitraum (z.B. ein Jahr) beobachtet werden oder dass die Tarifgruppen durch eine geeignete Studienplanung direkt vergleichbar sind. Anderenfalls kann ein kausaler Zusammenhang zwischen Stromtarifen und Verbrauchsunterschieden nur eingeschränkt bewertet werden.

Für die Auswertungen kann jedes beliebige Data-Mining-Tool verwendet werden (Übersicht siehe [34]), hier wird die MATLAB-Toolbox Gait-CAD [35] eingesetzt.

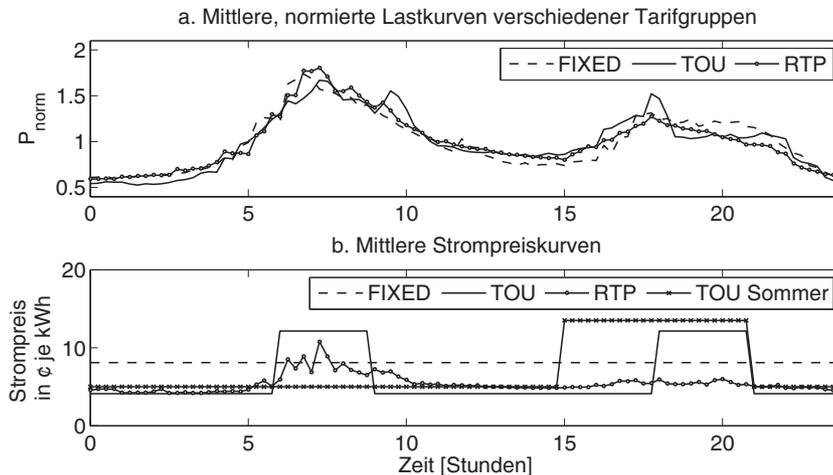


Abbildung 2: Zeitraum Januar bis März 2007: a.) (oben) Mittlere, normierte Tageskurven des beeinflussten und unbeeinflussten Stromverbrauchs, b.) (unten) mittlere Strompreise inkl. TOU-Sommertarif.

4 Olympic Peninsula Project Datensatz

Der verwendete Datensatz stammt aus der *Pacific Northwest GridWise Testbed Demonstration* Feldstudie. Im Teilprojekt *Olympic Peninsula Project*² wurde das Verbrauchsverhalten von Haushaltsstromkunden bei variablen Stromtarifen und geeigneter Technik bzw. bidirektionaler Kommunikation untersucht [21]. Die Messdatenerhebung fand im Zeitraum von April 2006 bis März 2007 statt, wobei die 112 teilnehmenden Haushalte zu gleichen Teilen drei Gruppen zugeordnet wurden:

1. Haushalte der FIXED-Gruppe erhielten einen konstanten Stromtarif von 8.1 ¢/kWh.
2. Die Stromkunden der TOU-Gruppe (engl. *time-of-use*) wurden durch einen Dreistufentarif motiviert, ihren Stromverbrauch tageszeitlich von Hochtarif- (*on-peak tariff*, 12.2 ¢/kWh bzw. 13.5 ¢/kWh) zu Niedertarifzeiten (*off-peak tariff*, 4.1 ¢/kWh bzw. 5 ¢/kWh) zu verschieben. Die dritte Tarifstufe des TOU-Tarifs (*critical-peak tariff* (CPP), 35 ¢/kWh) wurde lediglich an einem Tag für vier Stunden (1. Nov. 2006) ausgerufen. Sie dient der Netzentlastung in netzkritischen Situationen. Die Tagesmuster des TOU-Tarifs waren saisonal konstant und den Teilnehmern bekannt, wobei der Strompreis im Sommer höher lag als im Frühling, Herbst oder Winter.

3. Den Studienteilnehmern der RTP-Gruppe (engl. *real-time price*) wurde ein zeitvariabler Stromtarif übermittelt, der dynamisch und in Echtzeit an einem virtuellen Strommarkt gebildet wurde und die Gegebenheiten am traditionellen Strommarkt gemäß Angebot und Nachfrage widerspiegelte. Diese Gruppe existiert aber nicht im April, weil hier die gleichen Haushalte der CONTROL-Gruppe mit festen Tarifen angehörten. Die CONTROL-Gruppe hat eine sehr schlechte Datenqualität und wird im Folgenden vernachlässigt.

Abbildungen 2a und 2b verdeutlichen das differenzierte Verbrauchsverhalten der unterschiedlichen Tarifgruppen.

Eine wichtige Bedingung für die Projektteilnahme war der Einbau von Geräten zur elektrischen Warmwasserbereitung und Luftheizungsanlagen [21]. Die Haushalte konnten den Raumtemperaturregler und den Warmwasserbereiter gemäß ihrer Komfortbedürfnisse so voreinstellen, dass die Haushaltsgeräte in unterschiedlicher Stärke auf Preisanreize reagieren. Die aufgezeichneten *Smart Meter* Daten repräsentieren nicht nur die Last der Klima- und Lüftungsgeräte, sondern die Last aller im Haushalt befindlichen Geräte wie TV, Küchengeräte, Kühlschrank oder Waschmaschine. An bestimmten Haushaltsgeräten wie der Waschmaschine war zudem eine Anzeige angebracht, die den Nutzer über den aktuellen Strompreis informiert. Der Stromverbrauch setzt sich demnach aus zwei Anteilen zusammen. Einerseits aus der Veränderung der Verbraucherlast, hervorgerufen durch eine automatische Lastverlagerung. Andererseits aus einer manuellen Lastverlagerung, bewirkt durch das manuelle Ein- bzw. Ausschalten bestimmter Haushaltsgeräte, wie Waschmaschine oder Wäschetrockner. Die kumulierte Last aller im

² Download des Datensatzes mit vorheriger Anmeldung unter <https://svn.pnl.gov/olyphen>

Haushalt befindlichen programmier- und steuerbaren Geräte belief sich im Schnitt auf ca. 1.5 kW je Haushalt [21]. Für diesen Datensatz wurden bislang nur Regressionsmodelle für ganze Gruppen aufgestellt [8–10], es fehlen Untersuchungen für die Fragen 3-4 aus Kapitel 3.

5 Ergebnisse

Dynamische Regressionsmodelle: Der in diesem Zusammenhang untersuchte aggregierte Datensatz umfasst 60 Tage (10. Januar 2007 - 10. März 2007) für alle Haushalte der TOU-Tarifgruppe (Fall VT). Es ist zu beachten, dass vor der Modellbildung der lineare Trend in der Preis- und Verbrauchszeitreihe beseitigt wurde (Index “detrrend”). Abbildung 3 stellt die gemessene und den für die Modellstruktur in Gl. (1) geschätzten preisbeeinflussten Anteil der Verbraucherlast inkl. einer FIR-Modellschätzung gegenüber. Sowohl das ARX- als auch das FIR-Modell erkennen nur den wesentlichen Trend einer Preisreaktion (Senkung nach Preissprung), können aber Details und Variationen nicht erklären. Die unbekanntenen Messwerte für $P_R[k] = P[k] - P_V[k]$ müssen mit der Annahme $P_V[k] = P[k]$ für die FIXED-Gruppe berechnet werden, was ebenfalls zu Fehlern beiträgt. Zur Beschreibung des unterschiedlichen Verhaltens einzelner Haushalte (mit großen Bedarfs-

schwankungen) müssten haushaltsindividuelle Modelle angelern und ausgewertet werden. Deshalb sind dynamische Regressionsmodelle für die Identifikation von Typtagen nur bedingt geeignet.

Clusteranalysen: Die Einteilung des OPP-Datensatzes in die Fälle aus Tabelle 1 kann bei der Klärung unterschiedlicher Fragestellungen helfen. So können Typtage für den elektrischen Energieverbrauch einzelner Haushalte am ehesten mit Clusteranalysen der Fälle **DH**, **DT** bzw. DH_{norm} , DT_{norm} identifiziert werden. Als Startcluster wurden zufällige Datentupel aus der FIXED-Gruppe gewählt, um verschiedene Jahreszeiten und Wochentage zu repräsentieren. Der nächste Schritt ist die Bestimmung der optimalen Clusteranzahl (siehe Tabelle 2).

Wie erwartet empfiehlt das erste lokale Minimum der Separation zwei Cluster, außer im Fall DH_{norm} . Allerdings geht so viel Information verloren, was eine spätere Analyse der Tarifauswirkungen erschwert. Mit dem Ausschluss von $C = 2$ ergeben sich vier bzw. sechs Cluster. Clustereinteilungen für höhere lokale Minima der Separation wurden ebenfalls manuell inspiziert, allerdings neigen diese tendenziell zum Abspalten sehr kleiner Gruppen und wurden deshalb verworfen.

Für alle Analysen zum Vergleich der Tarife muss der Monat April ausgeschlossen werden, um Verzerrungen

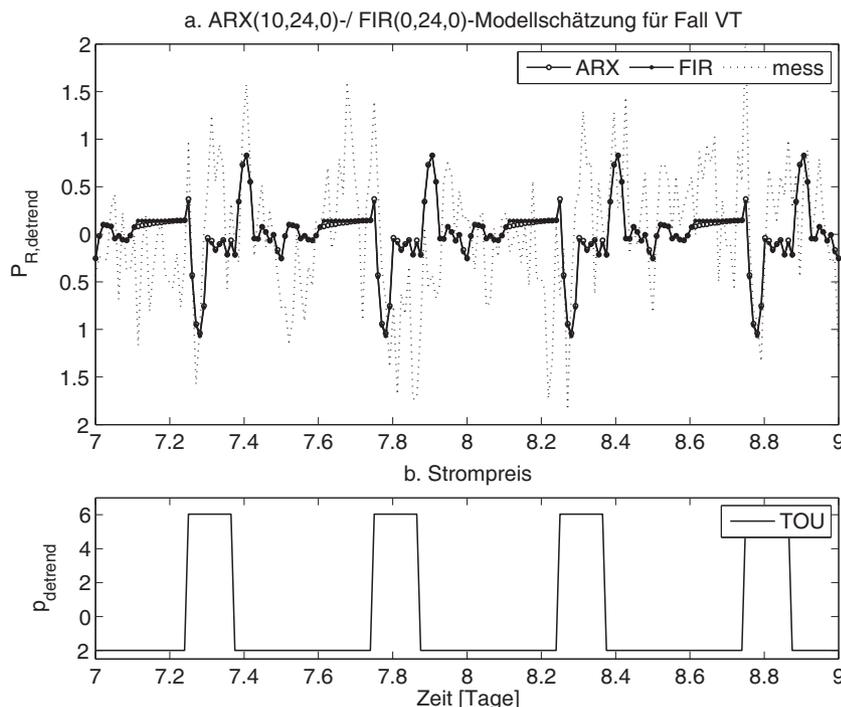


Abbildung 3: ARX-/FIR-Modellschätzung für Fall VT für TOU-Tarifgruppe: Gegenüberstellung des gemessenen preisbeeinflussten Anteils der Last und der ARX(10,24,0)-/FIR(0,24,0)-Modellschätzung; *mean-square error* (MSE): 52.77 % bzw. 52.81 %.

Tabelle 2: Bestimmung der optimalen Clusteranzahl C_{opt} , N_D : Anzahl der Datentupel.

Fälle	N_D	Lokale Minima Separation	$S(C)$	C_{opt}
DH	21894	7850 (2), 9852 (4) , 11898 (6)		4
DH _{norm}	21894	17604 (6) , 13825 (12)		6
DT	931	76 (2), 140 (6) , 181 (9)		6
DT _{norm}	931	623 (2), 299 (6) , 253 (8)		6

durch das Fehlen der RTP-Gruppe auszuschließen. Die entsprechenden Bemerkungen finden sich an den jeweiligen Tabellen und Abbildungen.

DH-/DH_{norm}-Cluster: Bei der Betrachtung haushaltsindividueller Tageskurven sind gut interpretierbare Cluster zu erkennen. Die Verläufe der **DH**-Cluster in Abbildung 4a unterscheiden sich hauptsächlich über die Gesamtenergiemenge, während die **DH_{norm}**-Cluster in Abbildung 4b Lastverschiebungen über den Tag zeigen. Tabelle 3 verdeutlicht für **DH_{norm}**, dass die Cluster 1 (nach hinten verschobene Morgenspitze) und 4 (nach vorne verlagerte Abendspitze) mehr TOU-Haushalte enthalten. Cluster 3 ist über den Tag sehr homogen. Die ausgeprägte Morgenspitze von Cluster 5 in Spitzenpreisen ist hauptsächlich für die FIXED-Gruppe akzeptabel. Die TOU- und RTP-Haushalte in diesem Cluster reagieren kaum auf Preisanreize, was Frage 4 aus Kapitel 3 beantwortet. Abbildung 4c macht die verbleibende Varianz innerhalb der Cluster für die einzelnen Haushalte deutlich.

Signifikante Unterschiede von Mittelwerten können durch den t-Test bzw. Unterschiede von Verteilungsformen binärer Variablen durch den Vierfelder- χ^2 -Test nachgewiesen werden. Da die Daten jedoch stellenweise die Voraussetzungen an diese Tests verletzen (z.B. bezüglich Daten-

Tabelle 3: Verteilung der Typtage für verschiedene Tarife im Fall **DH_{norm}** für sechs Cluster (siehe Abbildung 4b), Werktage ohne Monat April; Exakter Test nach Fisher, zweiseitig: FIXED vs. TOU, FIXED vs. RTP. Kennzeichnung der Signifikanzniveaus: $\alpha \geq 0.05 \rightarrow$ nicht signifikant, $\alpha < 0.05 \rightarrow$ signifikant (*), $\alpha < 0.01 \rightarrow$ hoch signifikant (**), $\alpha < 0.001 \rightarrow$ höchst signifikant (***)

C	FIXED	TOU	RTP
1	405 (22.1%)	808 (44.1%)***	620 (33.8%)***
2	664 (34.8%)	587 (30.8%)	657 (34.4%)
3	1825 (37.0%)	1396 (28.3%)***	1716 (34.8%)
4	644 (27.5%)	953 (40.8%)***	741 (31.7%)*
5	774 (47.6%)	386 (23.7%)***	467 (28.7%)***
6	347 (24.8%)	509 (36.3%)***	546 (38.9%)***

Tabelle 4: Verteilung der Typtage für verschiedene Tarife im Fall **DT_{norm}** für sechs Cluster (siehe Abbildung 5b), Werktage ohne Monat April; Exakter Test nach Fisher, zweiseitig: siehe Beschriftung in Tabelle 3.

C	FIXED	TOU	RTP
1	18 (12.6%)	67 (46.9%)***	58 (40.6%)
2	0 (0.0%)	19 (90.5%)***	2 (9.5%)
3	21 (23.3%)	25 (27.8%)	44 (48.9%)*
4	95 (79.2%)	13 (10.8%)***	12 (10.0%)***
5	3 (6.3%)	11 (22.9%)	34 (70.8%)***
6	60 (35.5%)	62 (36.7%)	47 (27.8%)

umfang, Homoskedastizität), kommen in den Tabellen 3, 4 und 6 die robusteren Sonderformen Welch-Test und exakter Test nach Fisher zum Einsatz. Die Verschlechterung des Signifikanzniveaus wird in Kauf genommen, um im Gegenzug verlässlichere Ergebnisse zu erhalten.

DT-/DT_{norm}-Clustern: Die Zeitreihenaggregation verringert die Varianz innerhalb der Cluster (Abbildung 5). Die **DT_{norm}**-Cluster zeigen wiederum die Lastverschiebung besser als die **DT**-Cluster. Im Fall **DT_{norm}** ist Cluster 1 mit dem ausgeglicheneren Verlauf eine Reaktion von TOU und RTP auf Anreizsignale. Cluster 2 ist tendenziell ein TOU-Cluster mit einer nach hinten verschobenen Morgenspitze und einer nach vorn verlagerten Abendspitze. Cluster 4 zeigt den unbeeinflussten Verbrauch der FIXED-Gruppe. Tabelle 4 belegt, dass Cluster 5 besonders viele Tage der RTP-Gruppe enthält.

Analyse der Cluster mit informationstheoretischen Maßen: Die Bewertung der Ergebnisse der Clusteranalyse mit informationstheoretischen Maßen (Tabelle 5) zeigt, dass die relative Transinformation pro Ausgangsgröße "Vertrag" im Fall **DT_{norm}** am höchsten ist. Durch die Cluster dieses Datensatzes kann der Einfluss des Tarifmodells am besten bewertet werden. Werktage lassen sich durch Cluster des Falls **DT_{norm}** gut erklären ($Q = 0.35$), Cluster des Falls **DT** erklären wegen der starken Verbrauchsunterschiede zwischen Sommer- und Wintermonaten hauptsächlich die monatlichen Verbrauchsunterschiede. Das Clustern der haushaltsindividuellen Lastkurven (Fälle **DH**, **DH_{norm}**) liefert wegen der größeren Varianz geringere Erkenntnisse über den Zusammenhang der Ausgangsgrößen Monat, Werktag, Vertrag und den gefundenen Clustern.

Die normierten Fälle **DH_{norm}**, **DT_{norm}** eignen sich folglich besser zur Beurteilung der Preisauswirkung als die nicht normierten Fälle. Außerdem sind die Tarifgruppen **DT_{norm}** aussagekräftiger als die Haushalte **DH_{norm}**. Die informationstheoretischen Maße geben somit Hinweise auf

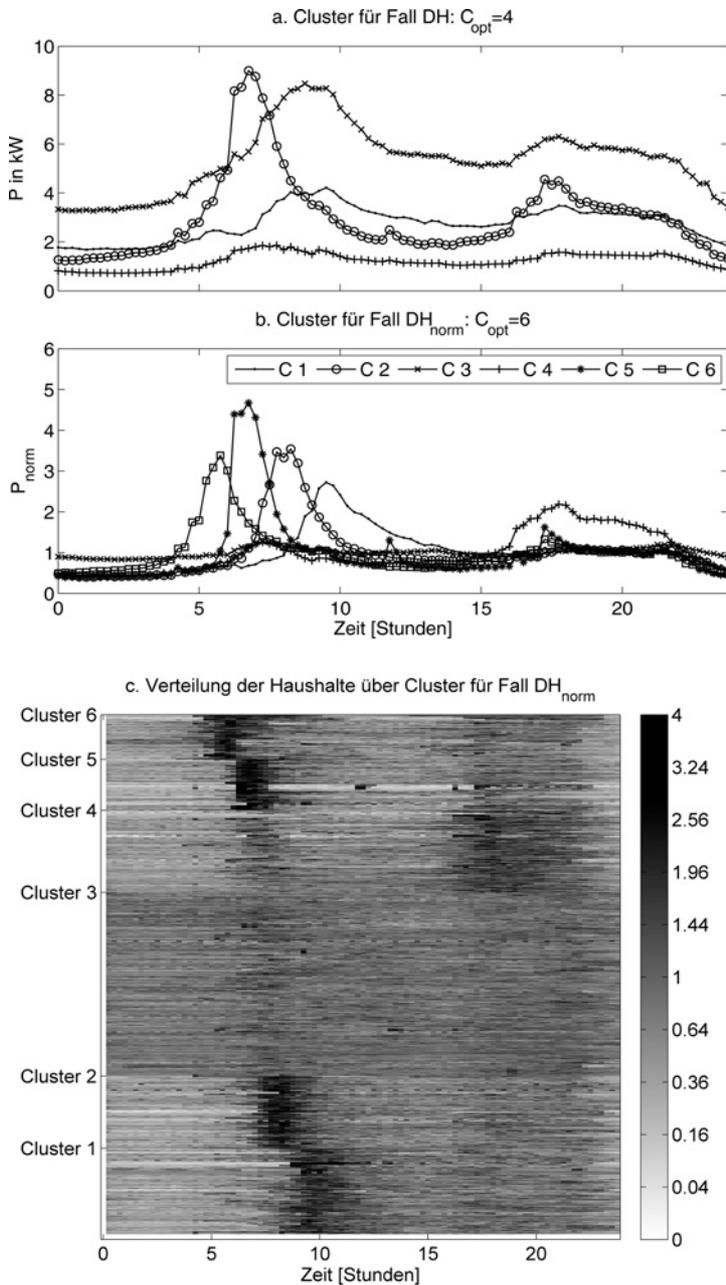


Abbildung 4: Gefundene Cluster: a.) (oben) Leistung für Fall DH, b.) (Mitte) normierte Leistung für Fall DH_{norm} , c.) (unten) Verteilung der Haushalte für Fall DH_{norm} mit grauwertkodiertem normierten Verbrauch.

prinzipielle Zusammenhänge, deren Relevanz dann im Detail zu prüfen ist, siehe z.B. Tabellen 4 und 5.

Analyse einzelner Abtastzeitpunkte im TOU- und RTP-Tarif: Tabelle 6 belegt, dass ein extrem teurer TOU-Tarif von 35 ¢/kWh zu einer drastischen Lastabsenkung von ungefähr 40 % führt, was Frage 3 aus Kapitel 3 beantwortet. Es muss allerdings angemerkt werden, dass dieses Ergebnis auf einer relativ geringen Datenbasis beruht. Bei der Preisklasse 12.15 beträgt diese Lastverschiebung noch ca. 13 % des Tagesdurchschnittes.

Ein ähnliches Verhalten ist bei der Betrachtung des TOU-Sondertarifs im Sommer zu erwarten. Da dieser Sondertarif mit den Preisstufen 5.0 ¢/kWh und 13.5 ¢/kWh lediglich im Sommer gilt, beträgt die Datenmenge ca. ein Viertel der Datenmenge des regulären TOU-Tarifs (4.12 ¢/kWh und 12.15 ¢/kWh). In den Hochtarifzeiten liegt die durchschnittliche Leistung der Haushalte der TOU-Gruppe allerdings um ca. 7 % höher als die durchschnittliche Leistung der Haushalte der FIXED-Gruppe, siehe Frage 3 in Kapitel 3.

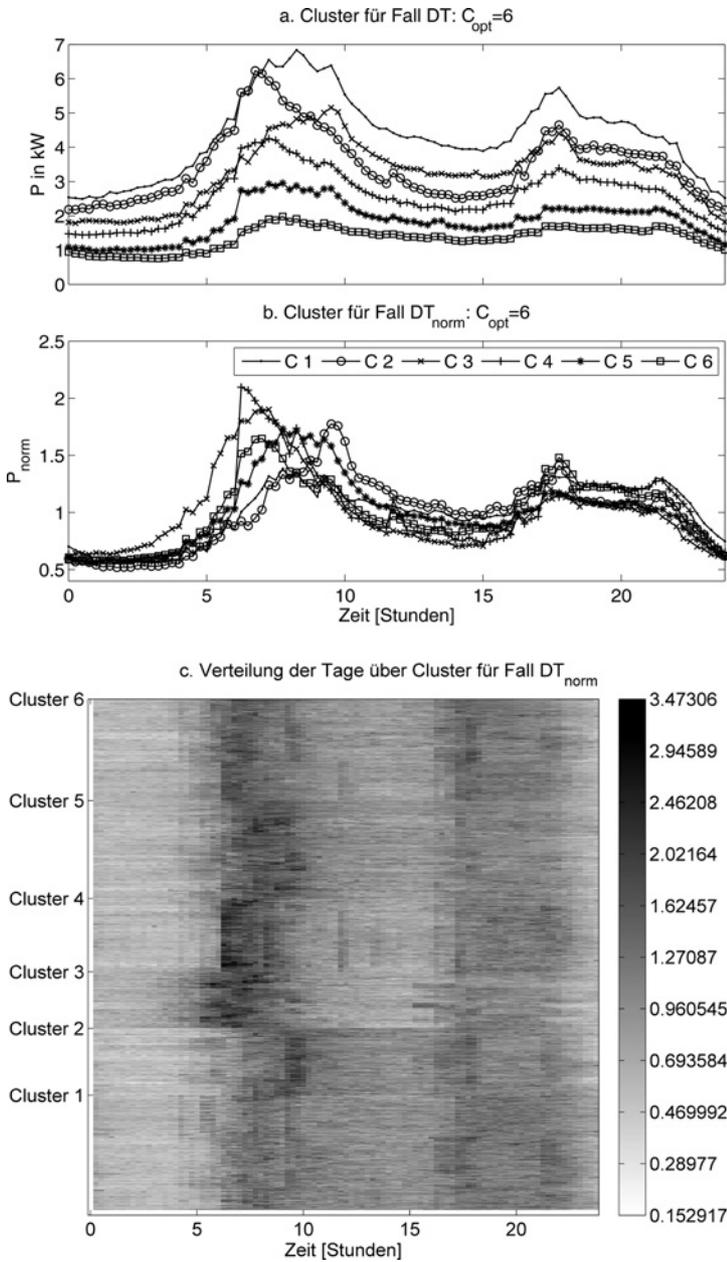


Abbildung 5: Gefundene Cluster: a.) (oben) Leistung für Fall DT, b.) (Mitte) normierte Leistung für Fall DT_{norm} , c.) (unten) Verteilung der Tiptage für Fall DT_{norm} mit grauwertkodiertem normierten Verbrauch.

Eine mögliche Erklärung für diese untypische Reaktion könnte die Dauer des Hochtarifs geben. Im TOU-Sommertarif erstreckt sich die Hochpreisphase über einen Zeitraum von sechs Stunden (15:00 bis 21:00). Die Verdopplung des Hochtarifzeitraums führt scheinbar zu falschen Anreizen zur Lastverlagerung.

Eine Bewertung der Auswirkung von RTP-Preisen auf das Verbrauchsverhalten von Haushaltsstromkunden fällt schwer. Auch hier bewirken extrem hohe Strompreise in der Preisklasse von ca. 36.94 ¢/kWh eine Lastabsenkung

von durchschnittlich 6% im Vergleich zu Haushalten der FIXED-Gruppe. Allerdings dreht sich dieses Verhältnis in der zweithöchsten Preisklasse von ca. 20.39 ¢/kWh um. Es wird mehr Strom verbraucht als zu gleichen Zeiten bei der Tarifgruppe mit konstantem Strompreis.

In allen anderen Preisklassen sind nur geringe prozentuale Unterschiede im Verbrauchsverhalten zwischen der FIXED- und der RTP-Gruppe auszumachen. Echtzeit-Elektrizitätspreise scheinen in diesem Projekt kein geeignetes Anreizsignal für eine Veränderung des Verbrauchs-

Tabelle 5: Auswertung der Clusteranalyse mit informationstheoretischen Maßen für die optimale Clusteranzahl je Fall aus Tabelle 2, ausgenommen: Monat April.

Fall	Tage	N_D	Q		
			Monat	Werktag	Vertrag
DH	alle	21153	0.09	0.01	0.00
DH	Werktage	14045	0.10	0.00	0.00
DH _{norm}	alle	21153	0.01	0.02	0.01
DH _{norm}	Werktage	14045	0.01	0.00	0.02
DT	alle	891	0.34	0.06	0.03
DT	Werktage	591	0.34	0.00	0.04
DT _{norm}	alle	891	0.15	0.35	0.12
DT _{norm}	Werktage	591	0.18	0.00	0.16

Tabelle 6: Fall SH_{norm}: Auswertung der Preisklassen (PK in €/kWh) für den TOU-Tarif (oben) und den RTP-Tarif (unten); N_D : Anzahl der Datentupel pro Gruppe, ausgenommen: Monat April; Angabe der mittleren normierten Leistung, Standardabweichung und Datentupelanzahl je Preisklasse und Gruppe; Zwei-Stichproben Welch-Test, zweiseitig: FIXED vs. TOU bzw. FIXED vs. RTP; Kennzeichnung der Signifikanzniveaus: siehe Beschriftung in Tabelle 3.

PK	N_D	FIXED	TOU
4.12	≥ 398408	0.879±0.821	0.924±0.961***
5.0	≥ 105264	0.966±1.155	0.943±1.116***
12.15	≥ 132624	1.364±1.237	1.230±1.235***
13.5	≥ 35088	1.102±0.981	1.172±1.222***
35.0	400	1.013±0.775	0.602±0.590***
PK	N_D	FIXED	RTP
ca. 2.08	≥ 188263	0.946±1.054	0.954±1.123*
ca. 5.48	≥ 453782	1.010±0.973	1.006±0.994*
ca. 11.02	≥ 12879	1.185±0.970	1.189±1.014
ca. 20.39	≥ 7722	1.112±0.888	1.150±1.038*
ca. 36.94	≥ 6610	1.352±1.138	1.289±1.229**

verhaltens und für eine Lastverlagerung zu sein. Eine mögliche Erklärung ist eine Überforderung der Haushalte durch die schnell veränderlichen Preise, wodurch z.B. getroffene Entscheidungen (z.B. Beginn eines Waschgangs) nicht mehr rückgängig gemacht werden können.

6 Zusammenfassung und Ausblick

Ziel der Arbeit war, eine systematische, datenbasierte Analyse der Auswirkungen verschiedener Preissignale auf das Verbrauchsverhalten von Haushaltsstromkunden vorzustellen. Dazu wurde der zu untersuchenden *Smart Meter* Datensatz für den Auswerteprozess in verschiedene Teildatensätze unterteilt. Für die umfassende Analyse wird ein Schema mit sieben Schritten empfohlen, an dessen En-

de eine qualitative und quantitative Bewertung des preisbeeinflussten Verbrauchsverhaltens steht. Die verschiedenen aggregierten Datensätze und deren Untersuchung mit Data-Mining-Verfahren (Clustermethoden) erlauben es, strukturelle Unterschiede zwischen den Tarifgruppen auszumachen und Typtage zu bestimmen. Hierfür eignen sich im Besonderen die tagesnormierten Leistungskurven. Die TOU-Tarifgruppe zeigt hier klarere Verschiebungen als die RTP-Tarifgruppe. Insgesamt ergänzt das vorgeschlagene Schema die bekannten Regressionsmethoden, indem es unterschiedliche Reaktionen auf Preissignale (z.B. für ungewöhnliche Anreizsignale, für einzelne Haushalte) aufzeigt.

Ungeachtet der Analyseschritte sind allgemeingültige Forderungen an den *Smart Meter* Datensatz zu stellen. Um jahreszeitliche oder wochentagsbedingte Einflüsse zu bewerten, ist es unbedingt ratsam, dass der Datensatz Messwerte über ein ganzes Jahr beinhaltet. Zudem muss bei der Auswahl der Studienteilnehmer sorgsam vorgegangen werden, um nicht systematische Ungleichheiten in den verschiedenen Tarifgruppen zu verursachen.

Langfristig ist ein steigendes Lastverlagerungspotenzial durch „intelligente“ Geräte (Kühl-/Gefrierschrank, Wärmepumpe) zu erwarten, die ein Energiemanagementsystem (EMS) automatisch steuert und optimiert. Bei Geräten wie z.B. Waschmaschinen ist abzuwarten, ob die Bewohner Schaltvorgänge durch EMS akzeptieren. Bei RTP-Tarifen funktioniert eine Optimierung nur bedingt, wenn der Strompreis im Vorfeld nicht bekannt ist.

Das vorgeschlagene Konzept zur systematischen Analyse muss zukünftig auf andere artverwandte Datensätze angewendet werden, um die getroffenen Aussagen zur Methodik und zur Auswirkung von Anreizsignalen zu validieren und zu generalisieren.

Danksagung: Der Dank der Autoren gilt der Deutschen Forschungsgemeinschaft (DFG), die diesen Beitrag im Rahmen eines Forschungsprojektes (MI 1315/2-1) ermöglicht hat.

Literatur

1. *Dena*: dena-Netzstudie II - Integration erneuerbarer Energien in die deutsche Stromversorgung im Zeitraum 2015 - 2020 mit Ausblick 2025. Techn. Ber., Deutsche Energie-Agentur GmbH (dena), 2010.
2. *BDEW*: Erneuerbare Energien und das EEG: Zahlen, Fakten, Grafiken (2014). Techn. Ber., BDEW Bundesverband der Energie- und Wasserwirtschaft e.V., 2014.
3. *BMU; BMWi*: Das Energiekonzept der Bundesregierung 2010 und die Energiewende 2011. Bundesministerien für Wirtschaft

- und Technologie (BMWi) und für Umwelt, Naturschutz und Reaktorsicherheit (BMU), 2011.
4. *Bachem, A.; Buchal, C.*: Energiewende – quo vadis? *Physik Journal*, 12(12), S. 33, 2013.
 5. *Arnoldt, A.; Warweg, O.; Frenzel, S.; Werner, M.; Bretschneider, P.*: Analyse des Einflusses vom Demand Response RESIDENS Feldversuch auf das Kundenlastverhalten. In: VDE-Kongress 2012, 2012.
 6. *Appelrath, H.-J.; Kagermann, H.; Mayer, C.*: Future Energy Grid. Migrationspfade ins Internet der Energie. Acatech Studie. Februar, 2012.
 7. *Fang, X.; Misra, S.; Xue, G.; Yang, D.*: Smart Grid – The New and Improved Power Grid: A Survey. *IEEE Communications Surveys Tutorials*, 14(4), S. 944–980, 2012.
 8. *Corradi, O.; Ochsenfeld, H.; Madsen, H.; Pinson, P.*: Controlling Electricity Consumption by Forecasting its Response to Varying Prices. *IEEE Transactions on Power Systems*, 28(1), S. 421–429, 2013.
 9. *Dorini, G. et al.*: Chance-Constrained Optimization of Demand Response to Price Signals. In: *IEEE Transactions on Smart Grid*, Bd. 4, S. 2072–2080, 2013.
 10. *Waczowicz, S.; Klaiber, S.; Bretschneider, P.; Konotop, I.; Westermann, D.; Mikut, R.*: Virtuelle Speicher als adaptierbare Verbrauchermodelle zur Lastprognose und Betriebsführung in Verteilnetzen. In: Proc., 23. Workshop Computational Intelligence, S. 109–128, KIT Scientific Publishing, 2013.
 11. *Axon, C. J.; Darby, S. J.; Granell, R.; et al.*: Towards an Understanding of Dynamic Energy Pricing and Tariffs. In: Proc., 47th International Universities Power Engineering Conference (IUPERC), S. 1–5, IEEE, 2012.
 12. *Ghosh, S.; Sun, X.; Zhang, X.*: Consumer Profiling for Demand Response Programs in Smart Grids. In: Proc., IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), S. 1–6, IEEE, 2012.
 13. *Llanos, J. et al.*: Load Profile Generator and Load Forecasting for a Renewable based Microgrid using Self Organizing Maps and Neural Networks. In: Proc., Int. Joint Conf. on Neural Networks (IJCNN), S. 1–8, 2012.
 14. *Chicco, G. et al.*: Emergent Electricity Customer Classification. In: *IEE Proceedings Generation, Transmission and Distribution*, Bd. 152, S. 164–172, IET, 2005.
 15. *Chicco, G.; Ilie, I.-S.*: Support Vector Clustering of Electrical Load Pattern Data. *IEEE Transactions on Power Systems*, 24(3), S. 1619–1628, 2009.
 16. *Chicco, G.*: Overview and Performance Assessment of the Clustering Methods for Electrical Load Pattern Grouping. *Energy*, 42(1), S. 68–80, 2012.
 17. *Benítez, I. et al.*: Dynamic Clustering Segmentation Applied to Load Profiles of Energy Consumption from Spanish Customers. *International Journal of Electrical Power & Energy Systems*, 55, S. 437–448, 2014.
 18. *Nizar, A. H.; Dong, Z. Y.; Zhao, J.*: Load Profiling and Data Mining Techniques in Electricity Deregulated Market. In: Proc., IEEE Power Engineering Society General Meeting, S. 1–7, IEEE, 2006.
 19. *Alahmad, M. A. et al.*: A Comparative Study of Three Feedback Devices for Residential Real-time Energy Monitoring. *IEEE Transactions on Industrial Electronics*, 59(4), S. 2002–2013, 2012.
 20. *Paetz, A.-G.; Becker, B.; Fichtner, W.; Schmeck, H.*: Shifting Electricity Demand with Smart Home Technologies—An Experimental Study on User Acceptance. In: Proc., 30th USAEE/IAEE North American Conference, Bd. 19, S. 20, 2011.
 21. *Hammerstrom, D.; Ambrosio, R.; Brous, J.; et al.*: Pacific Northwest GridWise Testbed Demonstration Projects. Part I. Olympic Peninsula Project, 2007.
 22. *Sonnenschein, M.; Rapp, B.; Bremer, J.*: Demand Side Management and Demand Response. *Handbuch Energiemanagement, EW Medien und Kongresse GmbH, Frankfurt am Main*, 2010.
 23. *Klaiber, S.; Bretschneider, P.; Waczowicz, S.; Mikut, R.; Konotop, I.; Westermann, D.*: Intelligente Prognoseverfahren für beeinflusstes Verbraucherverhalten in Energiesystemen. In: Proc., 23. Workshop Computational Intelligence, S. 93–108, KIT Scientific Publishing, 2013.
 24. *Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P.*: From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17, S. 37–54, 1996.
 25. *Mikut, R.*: Data Mining für hochdimensionale Messsysteme. *Technisches Messen*, 77(10), S. 524–529, 2010.
 26. *Mikut, R.*: Data Mining in der Medizin und Medizintechnik. *Universitätsverlag Karlsruhe*, 2008.
 27. *Jain, A. K.*: Data Clustering: 50 Years Beyond K-means. *Pattern Recognition Letters*, 31(8), S. 651–666, 2010.
 28. *Loose, T. et al.*: Automatisierte Interpretation von Zeitreihen am Beispiel von klinischen Bewegungsanalysen. *at - Automatisierungstechnik*, 52, S. 359–369, 2004.
 29. *Bezdek, J. C.*: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
 30. *Höppner, F.; Klawonn, F.; Kruse, R.*: *Fuzzy Cluster Analysis*. John Wiley, New York, 1999.
 31. *Kroll, A.*: *Computational Intelligence: Eine Einführung in Probleme, Methoden und technische Anwendungen*. Oldenbourg Verlag, 2013.
 32. *Wang, W.; Zhang, Y.*: On Fuzzy Cluster Validity Indices. *Fuzzy Sets and Systems*, 158(19), S. 2095–2117, 2007.
 33. *Shannon, C. E.*: A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, S. 379–423, 1948.
 34. *Mikut, R.; Reischl, M.*: *Data Mining Tools*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1 (5), S. 431–443, 2011.
 35. *Mikut, R.; Burmeister, O.; Braun, S.; Reischl, M.*: The Open Source Matlab Toolbox Gait-CAD and its Application to Bioelectric Signal Processing. In: Proc., DGBMT-Workshop Biosignalverarbeitung, Potsdam, S. 109–111, 2008.

Dipl.-Ing. Simon Waczowicz

Institut für Angewandte Informatik (IAI), Karlsruher Institut für Technologie (KIT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Tel.: +49-(0)721-608-25755
simon.waczowicz@kit.edu

Dipl.-Ing. Simon Waczowicz ist wissenschaftlicher Mitarbeiter im Institut für Angewandte Informatik am Karlsruher Institut für Technologie. Hauptarbeitsgebiete: Modellierung von Energiesystemen, Datenanalyse.

Dipl.-Ing. Stefan Klaiber

Institutsteil Angewandte Systemtechnik (AST) des Fraunhofer IOSB, Am Vogelherd 50, D-98693 Ilmenau, Tel.: +49-(0)3677-46-1105
stefan.klaiber@iosb-ast.fraunhofer.de

Dipl.-Ing. Stefan Klaiber ist wissenschaftlicher Mitarbeiter am Institutsteil Angewandte Systemtechnik (AST) des Fraunhofer IOSB in der Arbeitsgruppe Energiewirtschaft und Systemanalysen der Abteilung Energie. Hauptarbeitsgebiete: Energieprognosen, Energiedatenmanagement, Modellbildung und Datenanalysen.

Dr.-Ing. Peter Bretschneider

Institutsteil Angewandte Systemtechnik (AST) des Fraunhofer IOSB, Am Vogelherd 50, D-98693 Ilmenau, Tel.: +49-(0)3677-46-1102
peter.bretschneider@iosb-ast.fraunhofer.de

Dr.-Ing. Peter Bretschneider ist stellvertretender Leiter des Institutsteils Angewandte Systemtechnik (AST) und Leiter der Abteilung Energie des Fraunhofer IOSB. Forschungsschwerpunkte: Energiemanagement, Zeitreihenanalyse und Prognose, Optimierung und Betriebsführung in Stromversorgungsnetzen, Smart Grids und Energiemärkte.

Dipl.-Ing. Irina Konotop

Fachgebiet Elektrische Energieversorgung, Technische Universität Ilmenau, Gustav-Kirchhoff-Straße 1, D-98684 Ilmenau, Tel.: +49-(0)3677-69-1490
irina.konotop@tu-ilmenau.de

Dipl.-Ing. Irina Konotop ist wissenschaftliche Mitarbeiterin an der Fakultät für Elektrotechnik und Informationstechnik im Fachgebiet Elektrische Energieversorgung an der Technischen Universität Ilmenau. Hauptarbeitsgebiete: Netzbetrieb mit steuerbaren Verbrauchern, Optimierung.

Prof. Dr.-Ing. Dirk Westermann

Fachgebiet Elektrische Energieversorgung, Technische Universität Ilmenau, Gustav-Kirchhoff-Straße 1, D-98684 Ilmenau, Tel.: +49-(0)3677-69-2838
dirk.westermann@tu-ilmenau.de

Prof. Dr.-Ing. Dirk Westermann ist Vollprofessor an der Fakultät für Elektrotechnik und Informationstechnik im Fachgebiet Elektrische Energieversorgung an der Technischen Universität Ilmenau. Forschungsschwerpunkte: Betriebsführung und -analyse sowie -design der zukünftigen elektrischen Energiesysteme.

Dr.-Ing. Markus Reischl

Institut für Angewandte Informatik (IAI), Karlsruher Institut für Technologie (KIT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Tel.: +49-(0)721-608-25749
markus.reischl@kit.edu

Dr.-Ing. Markus Reischl ist wissenschaftlicher Mitarbeiter im Institut für Angewandte Informatik am Karlsruher Institut für Technologie. Hauptarbeitsgebiete: Bildverarbeitung, Bioinformatik, Datenauswertung.

Prof. Dr.-Ing. Ralf Mikut

Institut für Angewandte Informatik (IAI), Karlsruher Institut für Technologie (KIT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Tel.: +49-(0)721-608-25731
ralf.mikut@kit.edu

Prof. Dr.-Ing. Ralf Mikut ist Projektleiter am Institut für Angewandte Informatik (IAI) und außerplanmäßiger Professor an der Fakultät für Maschinenbau des Karlsruher Instituts für Technologie (KIT). Forschungsschwerpunkte: Computational Intelligence, Data Mining, Biosignalanalyse, Bildverarbeitung, Regelungstechnik, Robotik, Bewegungsanalyse.