Smart Meter Driven Segmentation: What Your Consumption Says About You

Adrian Albert and Ram Rajagopal

***Abstract*—With the rollout of smart metering infrastructure at scale, demand-response (DR) programs may now be tailored based on users' consumption patterns as mined from sensed data. For issuing DR events it is key to understand the inter-temporal consumption dynamics as to appropriately segment the user population. We propose to infer *occupancy states* from consumption time series data using a hidden Markov model framework. Occupancy is characterized in this model by *1) magnitude, 2) duration, and 3) variability*. We show that users may be grouped according to their consumption patterns into groups that exhibit qualitatively different dynamics that may be exploited for program enrollment purposes. We investigate empirically the information that residential energy consumers' temporal energy demand patterns characterized by these three dimensions may convey about their demographic, household, and appliance stock characteristics. Our analysis shows that temporal patterns in the user's consumption data can predict with good accuracy certain user characteristics. We use this framework to argue that there is a large degree of individual *predictability* in user consumption at a population level.**

***Index Terms*— Classification, smart meter data, state-based modelling.**

I. Introduction

DEMAND response and energy efficiency programs have gained wide acceptance as a mechanism to cope with management of peak load demand in the grid. Such programs benefit from the heterogeneity in the end user consumption goals. For example, consumers whose consumption pattern is more *flexible*, can be compensated to either curtail or defer their consumption during peak energy hours. Other programs target consumers who are inefficient in their consumption due to long term choices, such as utilizing inefficient appliances or having poor home thermal insulation. Typical programs offer rebates in exchange for purchasing more efficient appliances or pursuing energy efficient interventions at their dwelling.

Due to the heterogeneity in consumer behavior and characteristics, finding the right consumers for the right program can be costly: it is not feasible to contact every consumer and obtain demographic, dwelling and behavior information. Furthermore, it is observed that excessive contact with potential consumers can lead to attrition or loss of enrollment (noted, e.g., in [11]). Yet, nowadays utilities have at their disposal a wealth of data about consumers that is not yet extensively used for improving program performance [33]. advanced metering infrastructure (AMI) deployments have enabled acquisition of fine grained energy consumption information which could reveal various preferences, behaviors and characteristics of individual consumers. We define *smart meter predictive program segmentation* as the problem of uncovering predictive relationships between consumption (as measured by smart meters) and consumer preferences, characteristics and behaviors. In this paper we describe this problem in more detail, and propose a first methodology that captures behaviors at multiple time-scales and forms to build a predictive segmentation program.

An initial step in uncovering segments of consumers is building rich, dynamic models of individual consumption; if such models were available at the individual level, one may then use them to compare user behavior and relate it with other available information on socio-demographic characteristics and user preference. We posit that, when external stimuli (e.g., weather patterns [32]) are accounted for, individual temporal consumption depends solely on behavioral choices (e.g., lifestyle) as reflected in consumption. Moreover, we hypothesize that these choices stem from socio-demographic characteristics and attitudes towards energy efficiency. Our interest here is for a *descriptive* analysis employing simple models of consumption. To this end, we 1) assess whether individual users' (weather-normalized) temporal consumption patterns is reasonably described using a simple, generative stochastic model that captures both dynamics and magnitudes of consumption (a hidden Markov model, HMM [29]); 2) compare and group users according to similarities in their consumption patterns (using spectral clustering of HMMs [27]; and 3) utilize the inferred characteristics of consumption to predict exogenous characteristics (demographics, appliance stock, etc.) To the best of our knowledge, this is the first paper to propose a dynamic model of consumption for program segmentation, as well as to predict user attributes and lifestyle using consumption characteristics.

The rest of the paper is structured as follows. Section II defines smart meter predictive program segmentation. Section III reviews the literature on data analytics for smart meters. Section IV outlines the modelling framework and algorithms used and how they relate to our application. Section V describes the data used. Section VI presents a discussion of consumption characteristics mined from the user population. Section VII shows that consumption characteristics may be used to infer

Manuscript received September 23, 2012; revised March 19, 2013 and April 29, 2013; accepted May 02, 2013. Date of publication June 21, 2013; date of current version October 17, 2013. This work was supported in part by Advanced Research Projects Agency-Energy (ARPA-E) and the Tomkat Center for Sustainable Energy. The work of R. Rajagopal was supported by the Powell Foundation Fellowship. Paper no. TPWRS-01072-2012.

A. Albert is with the Electrical Engineering and Management Science and Engineering Departments, Stanford University, Stanford, CA 94305 USA (e-mail: adalbert@stanford.edu).

R. Rajagopal is with the Civil and Environmental Engineering Department, Stanford University, Stanford, CA 94305 USA (e-mail: ramr@stanford.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2013.2266122

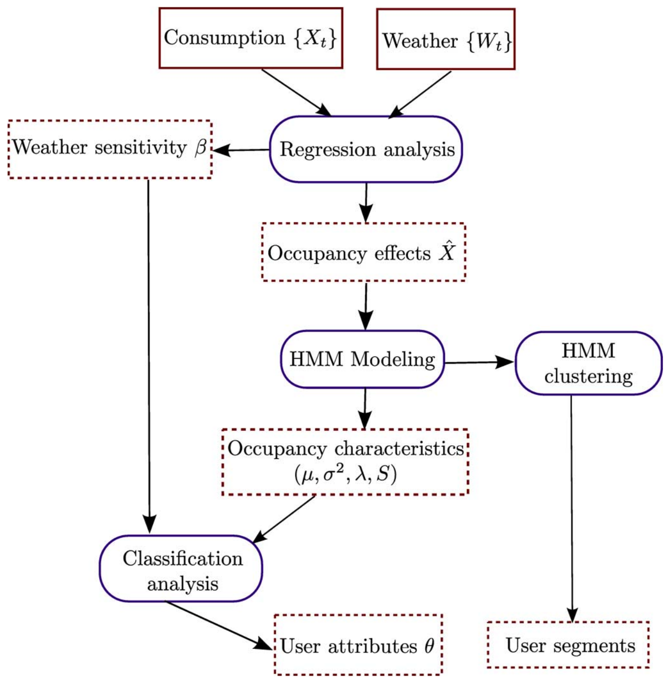

Fig. 1. Schematic of the analysis presented in this paper (see Section II).

user attributes. Section VIII offers a discussion of how this methodology may be used in practice for targeted incentive programs. Section IX concludes the paper.

II. Problem Statement

Predictive Program Segmentation: We observe N time series $\{X\}_n$, $n = 1, \ldots N$, each representing the energy consumption (in kWh) sampled at regular intervals Δt (here 30-min aggregates of 10-min data) of an individual household n (here referred to interchangeably as users or individuals). Our goal is to derive characterizations of individual consumption across a user base as to identify segments in the population whose usage patterns are in some way similar. Users in different groups have different inter-temporal consumption decision dynamics, and should benefit from different consumption curtailment strategies. From a system operator's perspective, the demand of users with less predictable temporal consumption profiles is more expensive to service than for those whose consumption one can predict well [5]. Yet demand-response programs may also benefit from diverse load patterns: with proper aggregation and control mechanisms (differential pricing, incentives for load deferral, etc.), one may achieve load-balancing results building on the relative flexibility in consumption of more random users. Moreover, it is expected that people sharing similarities on certain levels (lifestyle, appliances, house characteristics for the energy context) will generally exhibit similar characteristics in consumption. This hypothesis has been hard to validate in the past because of insufficiently detailed metered data. We propose a data-driven methodology that allows predicting user attributes using only consumption characteristics as inferred from the consumption time series. The steps are outlined below and in Fig. 1.

Dynamic User Consumption Model: For each user, we model energy use as the sum of a deterministic *weather sensitivity* component and an *occupancy* component:

$$X_t = f(W_t) + Y_t, \tag{1}$$

where W_t are exogeneously-specified weather time series, $f(\cdot)$ is a functional form specified by the modeler (here linear-in-parameters, see Section V), and Y is an occupancy-related component. Note that we explicitly adopted the simplification that weather and occupancy are independent, which is a strong assumption to make; subsequent extensions on this basic model will address this issue (see Section IX).

What we here term "occupancy" is in fact a proxy for a combination of 1) number of household members and their presence in the house, and 2) their activity while at home. In a real-world situation we do not observe ground-truth labels of the different activities over time that result in energy consumption decisions; but can we identify certain patterns that are associated with more high-level lifestyle and appliance ownership proxies?

Modeling each consumer separately has several advantages: 1) having a separate model for each user more richly captures the individual characteristics for personalized targeting strategies; 2) the computation is easily parallelizable, as no dependencies across users are taken into account; and 3) new AMI deployments and data storage and processing capabilities are making possible to analyze consumption at the individual level, offering identification properties and rich structure previously only available from panel-type analyses.

The predominant approach to modelling energy consumption has focused on the first component of (1), primarily by specifying a linear form of $f(\cdot)$ on weather $\{W_t\}$. We formulate a typical linear regression model in Section V; however, we argue there that this modelling choice leaves out important structure in the idiosyncratic component of the regression. In particular we observe a large degree of serial correlation and departure from the assumption of normality. Even when (linear) deterministic weather effects are accounted for, some users may display a bursty, relatively predictable consumption behavior, whereas others may appear to consume rather at random. In other words, the linear model predicts consumption well for some users, but fails to do so for others.

Our contribution is to explicitly model the idiosyncratic component Y, which we interpret as encoding occupancy effects on consumption. We posit that consumption follows an inter-temporal sequence of decisions that describe *occupancy states* of a household in addition to the (deterministic) baseline provided by weather. We wish to propose simple characterizations of the occupancy states in terms of *magnitude*, *duration*, and *variability*. Thus consumption may be represented schematically as in Fig. 2, with the magnitude dimensions (μ, σ^2) on the horizontal and vertical axis, the duration dimension represented by the radius of the circles (λ), and variability represented by the grayscale intensity of the transition arrow between the two states. A natural modelling choice for that end is a hidden (semi-)Markov model.

Model-Based User Classes: For the purpose of intervention programs, we want to group our N time series into $K \ll N$ types, so that each group may receive a tailored incentive. In

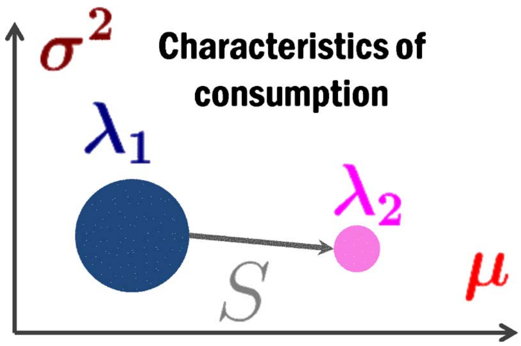

Fig. 2. Consumption characteristics: magnitude μ, duration λ, variability S.

practice it is critical to have a small enough K (to reduce administrative costs and user confusion), yet to capture the temporal dynamics structure in the data well. Given the non-linear fashion in which these patterns interact across users, we employ unsupervised learning techniques to compute groupings starting from the individual models for consumption. These groups are natural targeting segments for marketing demand-response programs.

User and Consumption Characteristics: Having characterized individual consumption according to weather sensitivity, magnitudes, duration, and variability, we wish to relate such structure to user attributes. We are interested in prediction power of model features with respect to lifestyle- and appliance-related attributes of users. This approach is the inverse of the current practice at utilities, which perform "psychographic" segmentation that attempts to infer consumption patterns from questionnaire data administered to customer sample [25].

III. Literature Review

Energy consumption has been of significant interest for a number of literatures, in particular economics and engineering. Econometric studies are typically concerned with user demand elasticity in the presence of nonlinear pricing and deregulation (e.g., [31], [32]). Another topic that has received much attention has been the effect of weather on residential energy use. For example, in [15] the authors argue for a non-linear dependence of energy consumption with temperature, whereas in [12] the authors propose a specification that uses lagged temperature values. In [16], the authors use the same dataset as us in a panel format to estimate econometrically the effects of real-time feedback on consumption. However until recently most such studies could only use daily or monthly data.

Because of the unavailability of high-resolution, hourly or sub-hourly meter data, the literature on energy analytics energy is still in its infancy. Some recent works use low-resolution consumption data for more recent statistical learning algorithms to improve on forecasting performance of early work. For example, in [21] the authors apply Gaussian process regression to monthly energy consumption data for residential buildings and household characteristics obtained from databases to predict energy usage at a city block level. With the recent introduction of *smart meters* at scale, more (mostly application-oriented) studies have been published that address high resolution time series modeling or customer clustering. For example, [33] challenges the assumption in the load analysis practice of morning and evening peaks by using a large time series corpus to identify 40 typical load shapes using `K-Means` clustering. In [10] the authors use 15-minute resolution smart meter data from ~200 customers of an utility company in Germany to perform a `K-Means` clustering analysis to inform cluster-based pricing schemes. They propose time-varying prices that are a function of the "number of peaks" and peak width, which essentially are rough measures of consumption dynamics. In [30] a features-based (`K-Means`) clustering is proposed using quantities such as the mean, standard deviation, maximum Lyapunov coefficient, Fourier mode energy for each of the 11 weeks and the ~1000 customers in their data. A similar approach is taken also in, e.g., [9]. More closely related to our analysis methodology, [2] identifies groups of substations in the Belgian grid based on daily load patterns at 245 buses using a multivariate version of spectral clustering. Also, [19] analyzes the same dataset [16] from the perspective of structural building characteristics effects on season-adjusted consumption statistics. Recently the authors in [1] have proposed a statistically-meaningful customer segmentation technique that is based on *power demand distributions*, i.e., recurring shapes in consumption distribution profiles.

Extensive literature exists on HMMs [29] and on their many flavors and applications, ranging from DNA structure analysis to speech recognition to financial market modeling. Here we use HSMMs [39]. Typical computational issues include estimating the number of states [4] and dealing with missing data [37]. In the context of residential load data, HMMs have only seen a limited use to date, in particular for the *disaggregation* problem of recovering individual appliance signals from the aggregate load profile [20], [22]. Recently HMMs have been adopted with enthusiasm also in the targeted interventions and marketing literatures. For example, in [26] the authors estimate a model of the dynamics of alumni-university relationship using a panel of donation data. Over the last decade a growing interest is noticeable in incorporating external covariates into the HMM framework, which is most commonly done using generalized linear (e.g., logit) models for rows of the transition matrix and linear regression models for the state emission probabilities [36], [40]. Recently a growing body of literature developed estimation methodologies for such models defined over an entire population (in a panel fashion) [40].

Segmenting a collection of time series using HMMs is also a relatively new thread of research. In a clustering application, direct comparison of the parameter space of two models is ill defined, as the models may differ greatly in complexity [17]. Different techniques [17], [23], [34], [38] have been proposed to learn clustering solutions for HMM collections which typically involve computing a log-likelihood matrix L of every time series under each of the models, and then applying a standard clustering algorithm on L (or a suitable transformation thereof), such as `K-Means` or *spectral clustering* [27]. Other, fully-parametric approaches view the K-clustering problem as estimating a mixture of K HMMs on the entire dataset [6]. The analysis presented in this paper more closely resembles that in [17] and [38], as we use a combination of parametric and unsupervised learning.

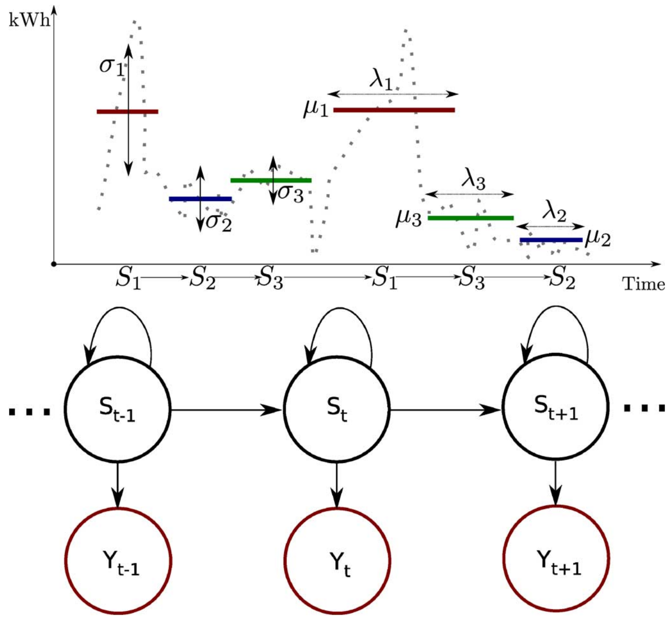

Fig. 3. Top: we model a user as a finite-state machine with states characterized by $\Lambda = (\mu, \sigma^2, \lambda)$; Bottom: the structure of a hidden Markov model. Observed values Y_t are generated by and underlying, unobserved Markov process Y_t. A hidden semi-Markov model is a generalization that allows the time spent in any state to obey a general distribution, including the geometric one for the HMM as a special case.

IV. Algorithmic Framework

A. Hidden Markov Models

We propose to model the time-dependent occupancy effects on consumption for an individual user as a trajectory through *occupancy states*. Each state is characterized by magnitude and duration spent in the state as described in Fig. 2. Moreover, the *dynamics* of transitioning between the states is modeled as a (discrete) stationary process. As we argue in Section V, a large degree of serial correlation is observable empirically in the consumption time-series. This suggests that modeling consumption as an HMM can transparently and parsimoniously capture both magnitudes and dynamics, while addressing the departures from the assumptions of the linear model observed in real data.

Definitions: A (discrete-time, homogeneous) *hidden Markov model* is a simple (stochastic) graphical model [29] with the following special structure (see Fig. 3, bottom panel):

1) A set of M (unobserved) states $S = \{S_1, \ldots, S_M\}$
2) An initial probability distribution over the states $\pi_i = P(S_1 = i)$
3) A state transition probability distribution $A_{ij} = P(S_t = i|S_{t-1} = j)$. This is the Markov assumption that the state of the system at time t only depends on its state at the previous time step $t-1$, but not on all past history.
4) An emission probability density $b(y) = p(Y_t = y|S_t)$.

In essence, a hidden Markov model can be viewed as a noisy observation of an underlying Markov chain. Note that while the chain $\{S_t\}$ obeys the Markov property, the process $\{Y_t\}$ does not. This allows us to model long-range serial correlations in a simple and flexible way.

The standard HMM has an implicit geometric distribution over the time spent in each state [40]. This construction may be too restrictive for modeling energy consumption, which depends on lifestyle and work schedules that span time scales longer than the 30-min unit of observation in our data that have characteristic time spans (e.g., 1–2 hour mornings preparing for going to work, 3- to 4-hour evenings at home relaxing at the end of the work day). Thus we wish to explicitly model the distribution $d(u_t|S_t = i)$ over the number of time steps u_t spent in each state. One class of models that internalize this behavior are the (discrete) HSMMs [39]. For this we used a simple Poisson counting process $u_t = k|S_t = i \sim \text{Poisson}(\lambda_i)$, i.e.,

$$d(u_t = k|S_t = i; \lambda_i) = \frac{\lambda_i^k e^{-\lambda_i}}{k!} \tag{2}$$

with rate parameters λ_i, $i = 1, \ldots, M$ to be estimated from the data. Our choice of the Poisson model is motivated by several observations. First, it is a standard, widely-used model for describing spread around an observed average duration of random event occurrences; such models have been used to model many aspects of human behavior, from communication activity (e-mail, text messages) to credit default and queuing behavior [3]. Second, it is a simple analytic model that simplifies computational inference.

We model the emission distributions as Gaussian, $b(y; \boldsymbol{\theta}) \sim \mathcal{N}(\mu_i, \sigma_i^2)$, i.e.,

$$p(Y_t = y|S_t = i; \boldsymbol{\theta}_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{(y-\mu_i)^2}{2\sigma_i^2}\right) \tag{3}$$

where $\boldsymbol{\theta}_i \equiv (\mu_i, \sigma_i^2)$ and $i = 1, \ldots, M$. We use Gaussian distributions because of 1) this choice offers direct correspondence with and least departure from the usual assumption of the linear model, for which many statistical results are known [13]; and 2) estimation via the expectation-maximization (EM) algorithm [14] is more efficient computationally because of the analytic maximization in the M-step (see below).

We denote $\boldsymbol{\Lambda} \equiv (\boldsymbol{\theta}, \boldsymbol{\lambda}) = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\lambda})$ the vector of model parameters that we need to estimate from the data (the bold face font indicates vector quantities). In this paper we estimate one such $\boldsymbol{\Lambda}_n$ for each individual household $n = 1, \ldots, N$.

Estimation: The (complete) likelihood for data $\{Y_t\}_{t=1}^T$ under such models may be written as

$$\mathcal{L}(\mathbf{y}) = \sum_{s_1,\ldots,s_T=1}^{M} P(\mathbf{Y} = \mathbf{y}, \mathbf{S} = \mathbf{s}), \text{ where}$$

$$P(\mathbf{Y} = \mathbf{y}, \mathbf{S} = \mathbf{s}) = \pi_{s_1^*} d_{s_1^*}(u_1) \prod_{r=2}^{R} p\left(s_r^*|s_r^*\right) d_{s_1^*}(u_r) \times \prod_{t=1}^{T} p(y_t|s_t) \tag{4}$$

where s_r^* is the rth visited state, $r = 1, \ldots, R$ and u_r the time spent in that state.

Fitting HMMs is referred to as the *learning problem*, i.e., computing maximum likelihood estimates (MLE) of the parameter vector $\boldsymbol{\Lambda}$:

$$\boldsymbol{\Lambda} = \arg\max \mathcal{L}(\mathbf{y}; \boldsymbol{\Lambda}).$$

This is usually performed via the Baum-Welch algorithm [29], which is a variant of the EM algorithm [14] for dealing with missing data (here the missing observations are the unobserved states $\{S_t\}$). The EM algorithm may be summarized as follows:

1) Start with a (reasonably chosen, see below) guess Λ^0.
2) *E-step*: compute expected value of complete-data log-likelihood $Q(\mathbf{\Lambda}|\mathbf{\Lambda}^k) = \mathbb{E}[\log \mathcal{L}(\mathbf{y}; \Lambda^k)]$.
3) *M-step*: choose $\mathbf{\Lambda}^{k+1} = \arg\max Q(\mathbf{\Lambda}|\mathbf{\Lambda}^k)$.
4) Repeat 2) and 3) until convergence.

Explicitly modeling durations u_r above complicates learning, but a number of efficient estimation methods have been developed for the task. Here we used the exposure and implementation in [28].

Note that the solution the EM algorithm converges to is typically a *local* maximum of the likelihood function. This is because $\mathcal{L}$ is non-convex in general. Careful initialization of model parameters or multiple re-estimations starting from different initialization points may ensure that a good such local optimum is found that is close to the global optimum. Following [23], here we first computed a `K-means` clustering on M (the number of HMM states) of the observation data $\mathbf{y}$; for each of those clusters we computed the mean and variance of the observations contained, which we then used as initial estimates for state means and variances.

Inference (computing the most likely sequence of states $\mathbf{s}$ that fits a given observation sequence $\mathbf{y}$) is referred to as the *decoding* problem in the HMM literature:

$$\arg\max_{\mathbf{s}} P(\mathbf{S}|\mathbf{Y}).$$

Efficient inference (the Viterbi algorithm [29]) that is based on dynamic programming is known for this problem. We use the implementation provided in the [28] in our calculations.

Choosing Model Size: The discussion above assumes that the number M of states is known *a priori*; this is not the case when dealing with real data. Many approaches have been proposed that address this problem in particular contexts [24]; here we employ a strategy outlined in [4]: for a given time series we fit an HSMM with a maximum allowed number of states $M_{\max}$ (here $M_{\max} = 10$) to the training set (about half of the observations); we then sequentially prune out the least probable state in the stationary distribution of the underlying Markov Chain, each time computing a measure of model fit (here the BIC criterion [14]) on a validation set (of similar size to the training set); we finally choose the optimum model size M^* the one with the highest BIC score.

B. HMM Clustering

As in [17], [23], we use *spectral clustering* to segment a collection of HMMs into classes of similar statistical properties. This is a graph-theoretic segmentation technique that relies on solving a minimum cut problem in a weighted graph [27]. For a K-clustering problem, the algorithm is as follows:

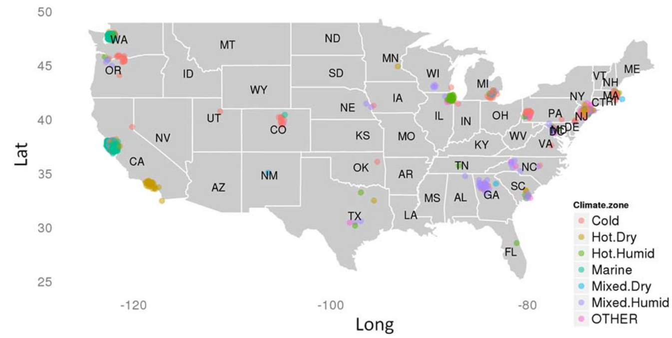

Fig. 4. Geographic distribution and climate zones of users in the sample.

1) compute a $N \times N$ symmetric matrix W, where

$$W_{ij} = \left| p\left(\{Y_t\}^j; \mathbf{\Lambda}^i\right) + p\left(\{Y_t\}^i; \mathbf{\Lambda}^j\right) - p\left(\{Y_t\}^i; \mathbf{\Lambda}^i\right) - p\left(\{Y_t\}^j; \mathbf{\Lambda}^j\right) \right| \quad (5)$$

for all pairs of sequences $\{Y_t\}$ and HMM fits $\mathbf{\Lambda}$;
2) compute the normalized Laplacian L of the graph defined by the adjacency matrix $\exp(-W_{ij}/2\sigma^2)$ (for σ^2 chosen as described in [27]);
3) compute the K principal eigenvectors corresponding to the largest eigenvalues of L and use them to form a new stacked $N \times K$ matrix P;
4) apply a standard clustering algorithm (here we used `K-Medoids`) on P, and output cluster membership.

V. Experimental Setup

A. Data Description

The primary dataset used in this paper was collected through an 8-month (March-October 2010) experiment [16]. The original goal of the study was to assess the impact of detailed consumption feedback on energy use. Data is available from about 1100 households of U.S.-based Google employees and contains 1) *power demand time series* of 10-min resolution for about 1100 households and 2) *socio-economic data* obtained via an online survey in which approximately 950 participants took part. In addition to this data we collected measurements on weather parameters (at 5- to 15-min resolution) at the locations (indicated by zipcode) of most users in the experiment using an online API.[1] The geographic distribution (and the corresponding weather patterns) experienced by the users varied substantially across the population (see Fig. 4). Most users however were located in Northern California.

Data was not free of inconsistencies (e.g., reading errors or incorrect timestamps); some time series contained very few data points (less than 200), such that they were left out of the final analysis. In addition, we only retained those households ($N = 952$) for which survey responses were available, and for which reliable fine-grained weather data could be collected. We

[1] http://www.weatherunderground.com

TABLE I
QUESTIONNAIRE SUMMARY (APPLIANCES AND DEMOGRAPHICS)

Question	Levels	Answers
Individuals.under.5	No/Yes	615/337
Individuals.36.to.54	No/Yes	546/406
Individuals.55.to.65	No/Yes	893/59
Individuals.over.65	No/Yes	927/25
Pets	No/Yes	573/379
Employed.full.time outside.home.	No/Yes	14/938
Employed.part.time outside.home.	No/Yes	835/117
Work.from.home	No/Yes	788/164
Unemployed.	No/Yes	669/283
Central.Heater	Other/Electricity	763/189
Spa.Hot.Tub.or.Pool.Heater	Other/Electricity	870/82
Water.Heating.Systems	Other/Electricity	814/138
Clothes.Dryer	Other/Electricity	366/586
Oven	Other/Electricity	435/517
Cooktop.Stovetop	Other/Electricity	615/337
Plasma.Televisions	No/Yes	670/282
Non.plasma.Televisions	No/Yes	180/772
Refrigerators	$< 2/ >= 2$	699/253
Computers	$< 2/ >= 2$	52/900
DVD.players	No/Yes	61/891
Gaming.Consoles	No/Yes	300/652
Dishwashers	No/Yes	53/899
Stand.alone.Freezers	No/Yes	821/131
Washing.Machines	No/Yes	91/861
Clothes.Dryers	No/Yes	94/858
Room.Units.or.Central.AC	No/Yes	339/613
Stand.alone.Unit.or Central.Heaters	No/Yes	70/882
Spas.Hot.Tubs.or.Pools	No/Yes	829/123

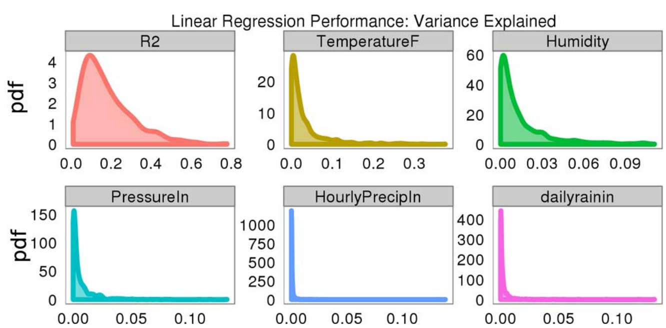

Fig. 5. Distribution of variance explained (R^2, expressed as fraction) by weather (overall and per-covariate) for all users in the sample.

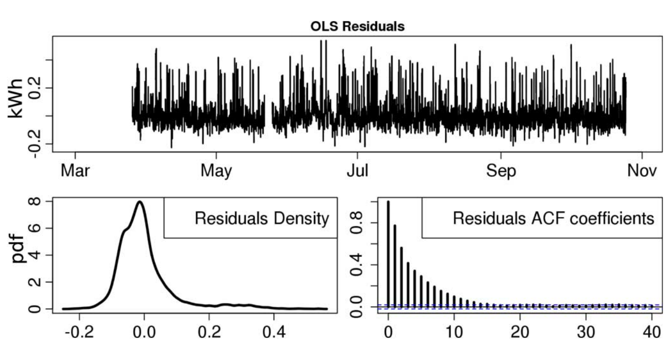

Fig. 6. Structure of the linear model error term for one example user. A time series of the residuals is shown in the upper panel. Note the non-Gaussian distribution density (lower-left panel) and the significant autocorrelation coefficients (lower-right panel, bars above the dotted blue line) for lags up to 16 (8 h).

snapped both the weather covariates and the consumption time series to the same time axis using nearest-neighbor interpolation to fill in small gaps (up to 2 h) and an EM algorithm (package `Amelia` in `R` with a multivariate normal target distribution) to imputate gaps for up to a day.

The online survey included information on many socio-demographic characteristics (we selected 89 covariates of interest, leaving out information such as political orientation and gender). A summary of some selected important attributes is given in Table I. For the purpose of this analysis, we were only interested in practical dichotomies on participant characteristics, e.g., whether a given household had a large number (>2) of fridges, or had any elderly people or infants.

B. Weather Effects

We estimate the (linear) weather effects in (1) as $f(W)$ by performing a regression analysis of the form

$$x_{nt} = \sum_{j=1}^{9} \beta_{nj} W_{nt} + \sum_{j=1}^{24} \delta_{nj} \mathbb{1}\{\mathrm{ToD}(t) = j\} + \sum_{j=1}^{7} \gamma_{nj} \mathbb{1}\{\mathrm{DoW}(t) = j\} + \nu_n \mathbb{1}\{\mathrm{Holiday}(t)\} + \epsilon_{nt} \quad (6)$$

where $n \in \{1, \ldots N\}$ refers to the user n as before, $t \in \{1, \ldots T\}$ represents time (on a 30-min resolution), W_{nt} are the weather covariates (temperature, pressure, humidity, and second-order interactions) for user n, the δ's, γ's and the ν's are dummy variables for time-of-day (in hours), day-of-week (Monday to Sunday) and federal holiday, respectively. We settled for specifying hourly and weekly fixed effects this way as opposed to defining a more flexible day-of-week to hour-of-day interaction terms because of the more stringent data requirements for estimating the latter specification ($24 \times 7 = 168$ coefficients as opposed to $24 + 7 = 31$ coefficients in the former case). The ϵ_{nt} terms are normally-distributed with zero mean (an ordinary least-squares regression). We used higher powers of weather covariates following [15], who argue for an up to fourth-order polynomial dependence with temperature of energy use based on a physical model of heat loss. The object of analysis in the next sections is the regression residuals $\hat{\epsilon}_{nt}$ which, according to Fig. 5 accounts for a significant portion of consumption: around 80% of variance in consumption for most users, and up to 35% of variance for even the users best-represented by linear models. Designing programs and evaluating their performance based on models with such poor predictive capabilities clearly miss important opportunities.

As Fig. 5 suggests, regression performance varies quite widely across the panel of users in explaining variance in consumption. Taken individually, weather covariates don't exhibit significant performance, with the best of the three being temperature (up to ~30% variance explained for some users) and the worst being humidity (green curve). However, the model (6) explains up to ~65% of the variance in consumption depending on the user. We retained the regression coefficients as classification features (Section VII).

C. Beyond the Linear Model: The Case for Occupancy Effects

Note that important assumptions on the idiosyncratic shock ϵ_{nt} in linear models such as (6) are independence and identical

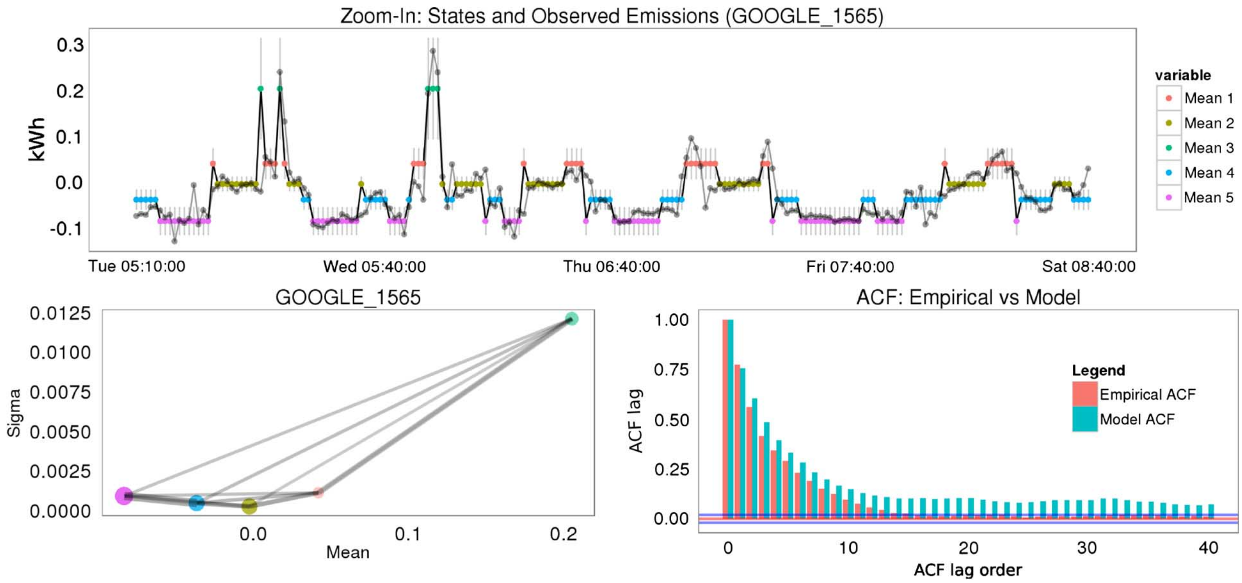

Fig. 7. HSMM model for one user. Top: example decoded sequence ($T = 200$ samples). States are shown in colors connected by a solid black line; original signal is shown in gray. Bottom-left: state space diagram as introduced in Fig. 2; Bottom-right: autocorrelation coefficients of HMM fit (blue bars) and original signal (red bars) up to 40 lags (20 h).

distribution (*i.i.d.*) according to a normal distribution (with constant variance) [13], i.e.,

$$\mathbb{E}[\epsilon_{nt}\epsilon_{nt'}] = 0 \quad \text{and}$$
$$\epsilon_{nt} \sim \mathcal{N}(0, \sigma^2). \tag{7}$$

These are very restrictive assumptions that do not necessarily hold in real consumption data, even with flexible specifications of the deterministic part in (6). For example, consider Fig. 6, where we present some properties of the regression residuals for one typical user. The residual density (lower-left panel) is clearly non-Gaussian, with apparent secondary peaks and a long tail, which suggests a mixture structure. Moreover, the autocorrelation coefficients for the first 15 lags (lower-right panel) are significantly greater then 0 (well above the 5% confidence level indicated by the horizontal dashed line).

We performed a similar analysis on the residuals for all users in our sample. First, we observe a high degree of heteroskedasticity (unequal variance σ^2 at different observations). In fact, of the $\sim$1000 users that we perform our analysis on, >94% display noticeable heteroskedasticity (as detected using a standard Breusch-Pagan test at a 5% level [13]). Second, we investigated the departure from normality of consumption for users in the sample, using an Anderson-Darling test of the null hypothesis $\mathcal{H}_0$ that the data was drawn from a normally-distributed population [13], and find that we are able to reject $\mathcal{H}_0$ at a 5% confidence level for 83% of the users in the sample. To test for serial correlation we used a Durbin-Watson test [13] (which tests $\mathcal{H}_0$ that the first-order autocorrelation lag $\rho = 0$), and find that we are able to reject $\mathcal{H}_0$ at a 5% confidence level for all (100%) the users in our sample. All these observations are clear indication that there is more to consumption that is not captured by the simple linear model that is pervasive in the energy consumption analysis literature.

VI. Occupancy States Analysis

A. Characteristics of Consumption

The analysis in this section proceeds on the occupancy component of consumption Y in our model (1). To each user's occupancy component Y we fit time-homogeneous HSMMs with Gaussian emission distributions and Poisson sojourn times, as outlined in Section VII and following the exposure in [28]. All estimations were performed in the open-source statistical language R. Model structure for an example user is illustrated in Fig. 7. The top panel presents the decoded sequence for about 5 days, with the means (colored dots) and standard deviations (vertical gray bars) of the inferred states plotted over the actual consumption (gray). Note that such a model is able to capture both the bursty events in consumption (the high-mean, high-variance state 3 depicted in green) and the relative flat consumption periods (state 5 depicted in magenta). The bottom-left panel is an illustration of the consumption characteristics of the occupancy states for the selected users following the schematic in Fig. 2, in which each state is a circle of radius λ in a (μ, σ^2)-space. Link widths are proportional to the transition probabilities between states and encode variability in consumption. The bottom-right panel compares the first 40 autocorrelation coefficients of the residuals $\hat{\epsilon_{it}}$ for the sample user with those of the HSMM fit. Such a model is (by construction) able to capture the long-range correlations in consumption to a significant degree at least for the first $\sim$10 lags (corresponding to 5–6 h).

In Fig. 8 we show the distribution of variance explained R^2 and model size M for the user population. Such modeling explains, on average, another $\sim$35% of the variance in the OLS residual, and as high as 80%. Most common model complexity is 5 states; note that the truncation at $M > 10$ was done for computational reasons (since the Baum-Welch has a runtime complexity quadratic in M, $\mathcal{O}(TM^2)$). There is wide variation on performance, which indicates that our simple modeling steps

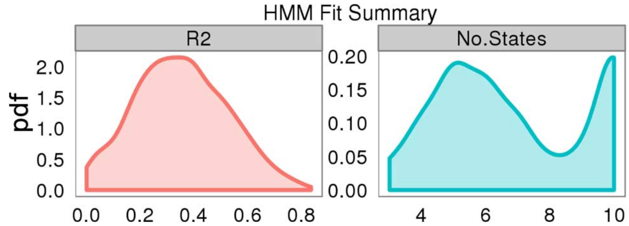

Fig. 8. HMM fit results: variance explained (as fraction, left panel) and model complexity M (right panel).

may not be sufficient for capturing the complex nature of consumption for certain users. We are currently developing a user model with superior predictive properties building upon the observations in this paper.

A general trend we notice in the case of magnitude characteristics (state mean μ and variance σ^2) is that the first four states tend to represent consumption levels below the baseline offered by the deterministic weather part in (1); models with 5 or more states represent complex dynamics including more pronounced bursty consumption. We also notice that typical sojourn times (rates λ of Poisson durations) range from 2–4 time steps (1–2 h) to upwards of 50 timesteps (a whole day). This indicates a large degree of heterogeneity in occupancy regimes for single users and across a population.

B. Occupancy Variability

To quantify empirically the randomness in the HSMM state sequence describing each individual user's consumption as estimated above, we turn to information-theoretic concepts. The (Shannon) entropy of a discrete (over M states), stationary, i.i.d. stochastic process with *known* p.m.f p is given by

$$S^{\rm shn} = -\sum_{j=1}^{M} p(j) \log_2 p(j). \qquad (8)$$

In our case $p(\cdot)$ is defined over *sequences* [7]. To approximately compute the entropy of arbitrary sequences for a finite, discrete-time process, we use the Lempel-Ziv estimator [35]

$$S^{\rm LZ} = \left(\frac{1}{T} \sum_i l_i \right)^{-1} \ln T \qquad (9)$$

with T the length of the sequence and l_i the length of the shortest substring that starts at position i and which does not appear at any position between 1 and $i-1$. In addition, we compute two additional benchmarks to characterize variability of the occupancy states under different models for $p(\cdot)$. If p is the uniform distribution, we have $S^{\rm rnd} = \log_2(M)$. If p is the limiting distribution $\hat{\pi}$ of the underlying Markov chain (the left-eigenvector of the transition matrix A, $\hat{\pi} = A\hat{\pi}$), we have $S^{\rm shn} = -\sum_{j=1}^{M} \hat{\pi}_j \log_2 \hat{\pi}_j$. In Fig. 9 (left panel) we present the distribution of the three benchmarks $S^{\rm rnd}$, $S^{\rm shn}$, and $S^{\rm LZ}$ over all users in the sample. Note that $S^{\rm LZ}$ peaks at $\sim$1.5 (corresponding to an average uncertainty of $2^{1.5} < 3$ states), whereas $S^{\rm shn}$ peaks at $\sim$2.4 (corresponding to an uncertainty $\sim$5 states). Moreover the distribution of $S^{\rm LZ}$ is much narrower than the others. These observations suggest that the temporal succession of states in the sequence indeed encodes information, allowing us to discriminate users by how flexible their consumption is.

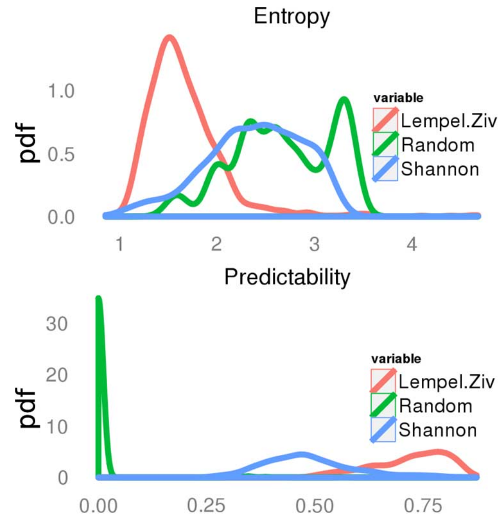

Fig. 9. Top: distribution of Random, Shannon, and Lempel-Ziv entropy benchmarks across the user population. Bottom: population distributions of predictability bounds $\Pi^{\max}$.

As an additional benchmark we calculated the *predictability* Π [35] for each user, defined as the probability that a suitable algorithm may correctly predict future states in a given sequence. For a sequence of length T and entropy S, we may calculate an upper bound $\Pi^{\max}$ on Π as the solution of

$$S = -\Pi^{\max} \log_2 \Pi^{\max} - (1-\Pi^{\max}) \log_2 (1-\Pi^{\max}) + (1-\Pi^{\max}) \log_2 (T-1). \qquad (10)$$

The distribution of $\Pi^{\max}$ estimates (random, Shannon, and Lempel-Ziv) is plotted in Fig. 9 (right panel). User predictability under our state-space model of consumption as estimated using the full entropy $S^{\rm LZ}$ peaks at 77%, as compared with the time-uncorrelated model $S^{\rm shn}$, which peaks at 47%. This again suggests that temporal patterns may be used to predict consumption, and that the user population is highly heterogenous in consumption predictability.

C. Clustering Analysis

We group users according to their temporal patterns as outlined in Section IV. To estimate the number K^* of clusters we used the Gap Statistic [14] technique and found $K^* = 8$. As mentioned in Section IV, we used a `K-medoids` algorithm as the last step in spectral clustering; this identifies "representative" users (cluster medoids) that can be used as prototypes in describing the entire population. The computed centers are illustrated in Fig. 10. Note that indeed the models are distinct in their properties: Cluster 1 is a complex 10-state model, whereas cluster 7 contains low-complexity models (3 states); cluster 5 contains medium-complexity models with small magnitude and variance states, out of which one is considerably "stickier" than the others. In addition, all models identified an occupancy state of high mean and variance, and relatively short duration. Anticipating the occurrence of such an occupancy regime is quite relevant in a Demand-Response context: an appropriate action (a so-called "Demand Response event" such as contacting the user to ask them to reduce or shift their consumption) may be taken for those individuals that are likely to exhibit a spike in consumption during peak.

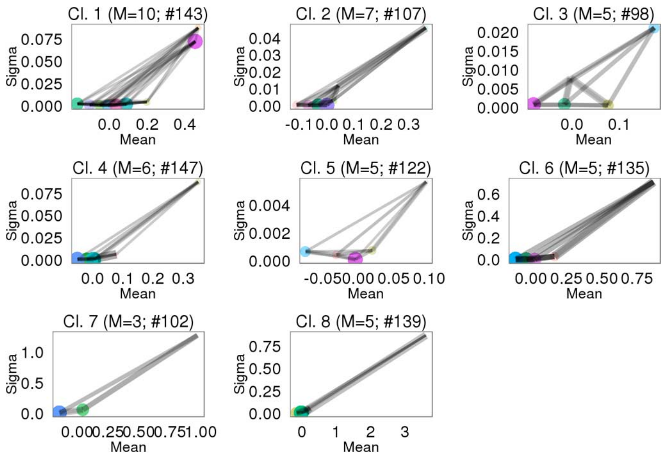

Fig. 10. Typical consumption models for the $N = 993$ users.

VII. Relating User and Consumption Characteristics

A. Inverse Problem

We set out to understand whether certain user characteristics may be inferred from knowledge of temporal patterns in their demand. Associate occupancy consumption characteristics with, e.g., whether the household has a certain appliance, or whether its inhabitants have certain lifestyles that are indicative of when and how they will consume can serve as a starting point to guide what kind of interventions aimed at consumption curtailment they might be more responsive to. As such, we learn (binary) classifiers to answer questions of the type "Does the household have a dishwasher?" or "Does the household contain an elderly person?" using as inputs consumption characteristics as described in Section VI. In essence, we explore the practical value of this modeling methodology to building a potential real-world system that is able to make statements about users' characteristics using their usage data. This is the opposite approach to the one that utilities currently take when designing their segmentation and targeting strategies. Typically, utilities would send out questionnaires of the type described in Section V to a sample of their users, and would use the answers (typical response rate is around 5%) for a "psychographic" segmentation [25], creating user groups such as "green enthusiasts" or "apathetic consumers". This approach is costly and scales poorly on the one hand, but also fails to take into account the actual energy usage. By comparison our proposed method allows to make (cheap) predictions on the lifestyle of the participants as to decide on the appropriate programs to offer to different users.

Here we are primarily interested in *prediction* performance without making any parametric assumptions, as opposed to building a microeconomically-sound model that readily lends itself to interpretation—which typically also has poor out-of-sample predictive performance. In fact, attempting generalizations of the type of typical interest in the related economics literature starting from this data would not be easily defensible: all users are relatively well-off, highly-educated Google employees, which suggests strong biases at the very least. Thus we are interested in a classification technique that may be rather opaque, but has good out-of-sample performance. Our preferred off-the-shelf classifier is the AdaBoost (the `ada` package in `R`). This is an ensemble learning technique that uses linear combinations of weak classifiers to improve performance (here we used classification trees). In essence, the algorithm computes optimal weights for a pre-determined number of individual classifiers, which are initialized with random assignments. A detailed description is given in [14]. For the purpose of this study we shall consider `AdaBoost` as a high-performance off-the-shelf classifier.

TABLE II
Classification Performance Metrics for the Top 10 Best Predictable Questionnaire Items: Relative Improvement Over Random Guessing (as Fraction)

Question	Precision	Recall	F_{measure}
Has electric clothes dryer?	0.50	0.10	0.19
Has washing machine?	0.52	0.10	0.19
Has central AC?	0.21	0.13	0.17
Has small children (under 5)?	0.06	0.18	0.11
Has unemployed occupants?	0.03	0.18	0.10
Has more than 1 fridge?	0.03	0.18	0.10
Has a plasma TV?	0.01	0.22	0.10
Has occupants aged 36-54?	0.05	0.13	0.09
Has pets?	0.03	0.15	0.08
Do occupants work from home?	0.00	0.16	0.07

B. Classification Performance Measures

We selected 33 items in the survey that relate to appliances and household occupancy characteristics (see Table I). We perform a 5-fold cross-validation analysis of classification, and report average results in Table II. To assess the out-of-sample classification performance we use the following concepts: the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Then we define the following performance metrics (following the discussion in [8]):

- *Precision*: $\mathrm{PREC} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$
- *Recall*: $\mathrm{RECL} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$
- *F-measure*: $\mathrm{FMES} = 2((\mathrm{PREC} \times \mathrm{RECL})/(\mathrm{PREC} + \mathrm{RECL}))$

All measures are between 0 and 1 and increase with classification performance. `FMES` is the geometric mean of precision and recall; we use this widely-accepted measure [8] as primary performance indicator.

The survey data is quite unbalanced; for example, only 91 users do not have washing machines, which is little over 10% of the population. Then a classifier that assigns labels at random (either has a washing machine or not) can appear to perform quite well. Consider a "Yes" response frequency p observed in the data, and a classification algorithm that assigns "Yes" with probability p *i.i.d.*. Then the True Positives and False Positives frequencies are $\mathrm{TP} = p^2$ and $\mathrm{FP} = p(1 - p)$. Other measures follow in the same manner as above. We call this algorithm a *random classifier*. Our task below is to assess whether the feature-based classifier can offer gains over the base case given by the random classifier.

C. Classification Results

In Table II we tabulate the performance metrics described above for a subset of questions that are typically associated with Demand-Response interventions: large appliances, occupancy information, heating/cooling infrastructure. All numbers are differences in percentages from performance obtained with

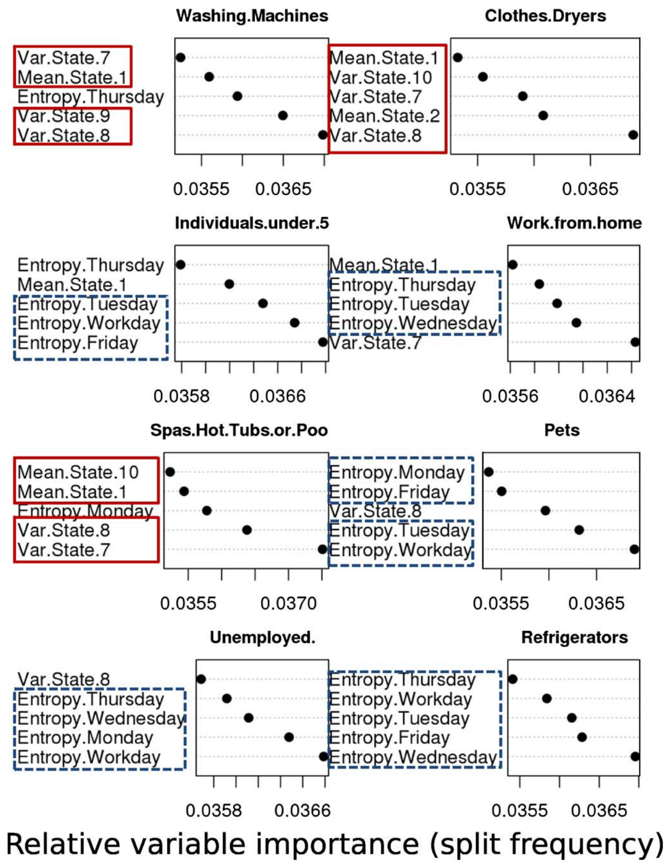

Fig. 11. Important variables for classification. Scores on the horizontal axis represent relative importance and are computed as in [14]. Magnitude-related consumption characteristics (magnitude, variance) are outlined with a red, solid line; variability-related consumption characteristics (entropy) are outlined with a blue, dashed line.

classification and performance obtained with the random classifier on imbalanced data. We observe that substantial gains in classification performance are achieved for some questions over random guessing: washing machines 19% improvement, clothes dryers 19% improvement, children under 5 11% (on the F-measure). Note the substantial improvement in precision (true positive rate) over random guessing for several questions, e.g., 50% for the electric clothes dryers or 52% for the presence of washing machines.

For each of the questions in Table II we also investigated the most relevant classification predictors. For ensemble learning such as the AdaBoost, relative importance is roughly the relative frequency with which the feature is selected at splits in the weak learners (decision trees) employed by the classifier [14]. The results are exemplified in Fig. 11 for 8 of the best-separated questions (only the first 5 most important features are shown). The states are ordered according to their means. The x-axis values represent the relative importance of a given covariate in the classifier.

A general trend observed was that appliance-related questions are best predicted by characteristics of states (roughly related to consumption magnitude), whereas lifestyle-related questions were best predicted by features more intimately related to temporally-varying consumption (entropy and predictibility). For example, randomness in consumption on workdays or Fridays can be a proxy for `Individuals under 5` (small children) in the household, whereas variance of high-mean states 8 and 9 predict the presence of washing machines. Interestingly, randomness in consumption on Thursdays is indicative of whether users owned washing machines, which may suggest that many users may do laundry on Thursdays. The presence of unemployed members of the household is also best predicted by randomness in consumption during work days—presumably unemployed people stay at home, and use the appliances, rather than go to work and follow a more regular schedule of consumption. Surprisingly cluster membership does not appear as a top predictor when using both features of the regression analysis step and HMM features, although it ranks higher than regression features. It does however appear as an important predictor for some questions when using HMM features alone (not shown here for space considerations). This suggests that similarity between users as determined with a segmentation algorithm may only capture second-order effects and its usefulness may be limited to summarization of data. Note also that sensitivity to weather and fixed-effects for time-of-day and day-of-week that we used in the initial weather normalization in the linear regression analysis (6) do not appear as top predictors, either (they come up with intermediate scores though and are not shown in Fig. 11). This is again surprising, but may reflect the fact that users in this sample are all professionals who do not base their use of appliances and their schedules on weather. However this requires further validation on larger and more representative datasets—a topic we are also pursuing currently. Given the limited number of people in the sample, and the issues with data quality described in Section V, generalizable conclusions may be drawn only with that further analysis.

VIII. Discussion: Data-Driven Program Targeting

HSMMs offer a compact characterization of users' consumption as noisy observations of a finite-state automaton. It allows us to address several important Demand-Response tasks [5], [18]

- identify users displaying a "random" consumption from those with more "regular" usage patterns;
- understand inter-temporal consumption dynamics as to anticipate negative-impact events such as high-magnitude spikes in consumption.

To the first point, "predictable" users should be treated in a fundamentally different way than irregular ones for the purpose of efficiency programs and control mechanisms for load-balancing demand on the grid: irregular users may benefit from enrollment in peak-pricing (to incentivize a more regular consumption), while relatively predictable users, who are thus less flexible in their consumption, may be targeted for rebates for efficient appliances (to reduce their consumption magnitude). Moreover, it has been argued that predictable loads are cheaper for the utility to service [1], [5].

In the case of demand-response management, it is of interest to understand timing of consumption at both the individual level and at the aggregate level. In particular, being able to specify a probability distribution over the occurrence of high-load events given the current state of consumption has practical value in determining, e.g., cost of service for an individual user [5]. For

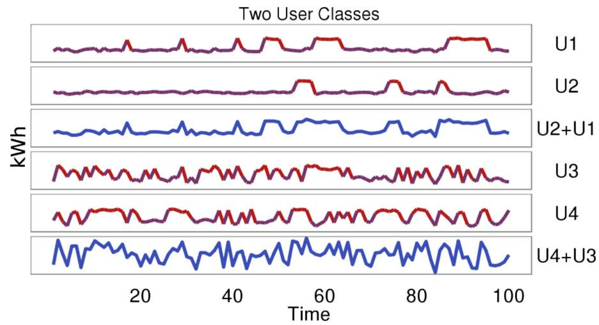

Fig. 12. Simulated consumption (HMM) time profiles: two classes of users with very different temporal consumption dynamics. Users $U1$ and $U2$ have a "regular" consumption, while users $U3$ and $U4$ have a less predictable usage pattern.

example, Pacific Gas and Electric (PG&E, a large California energy utility company) runs a differential pricing program (*Smart Rate*) for residential users that charges higher rates during the peak usage times in the summer.[2] A general consumption model over a population is not of practical use here: action is required at the individual level. Modeling each user as described above allows predictions of the type $P(Y_{t-1}|S_t)$, i.e., knowing the state of consumption at the current hour allows to define a probability distribution over consumption at the next hour. In the SmartRate context, a dynamic user model will allow the system operator (PG&E) to identify the customers that are likely to have high consumption during peak time, which in turn may be targeted for nudges such as text (SMS) alerts requesting that they reduce consumption in the next hour. A first-pass triage of the users may take place with a model-based segmentation such as the one illustrated in Section VI, which identifies prototypical users in terms of their model complexities and usage dynamics.

As an example, let us suppose that the system operator's objective is to reduce congestion and peak demand for the network, where the serviced population is composed of four users whose consumption is illustrated in Fig. 12. She may then note that users $U1$ and $U2$ have similarly predictable demands, which differ substantially from those of users $U3$ and $U4$ (which, in turn, are quite similar to each other). One reasonable intervention for the first group would be demand deferring (delaying or squeezing consumption over time), which would dis-align maxima and result in less pronounced aggregate peaks. Users $U1$ and $U2$ seem however to have relative little flexibility in the amount of consumption; thus they might also benefit from an efficiency program such as rebates for improved appliances that might not make their consumption more predictable, but at least would shave off some of the consumption. The utility thus needs to identify the appropriate users to whom it can make specific enough recommendations about what actions they may take that will be effective for demand curtailment and saving money.

But what specific actions can the utility recommend when for most users it only knows their consumption patterns? Being able to relate the consumption characteristics described in Section VI to user characteristics as done in Section VII would be useful in this context. Assessing whether the user is likely to consume in a certain way because of a pet in the household or a high number of computers can serve as a basis for tailored marketing actions.

[2]http://www.pge.com/en/myhome/saveenergymoney/energysavingprograms/smartrate/index.page

IX. Conclusion and Future Work

We have investigated the effectiveness of HMMs and model-based cluster analysis for smart meter time series data in producing meaningful features for classification. We show that one may correlate temporal consumption patterns with certain user characteristics. In particular, we calculated crude estimates of the amount of inherent randomness in users' consumption, and found that there is both a significant amount of predictability in users' state sequence and considerable variation in this quantity over the population, which may be used as first proxy for inferring certain lifestyle and appliance stock characteristics. This suggests that the dynamics of the time series as captured by HMM analysis can serve as valuable cues in guiding tailored interventions.

Methodologically, we plan to extend this work 1) with a more flexible consumption model that internalizes the dependency of occupancy states on weather and other exogeneous covariates and 2) with an explicit, dynamic model of user engagement with different energy programs. Clearly, one important challenge is to formulate segmentation algorithms and implementations that scale to very large user populations, and that may be estimated as closely on-line as possible. We are currently developing flexible population-level HMMs and estimation methodologies for distributed computing environments such as Hadoop/MapReduce.

Acknowledgment

A. Albert would like to thank the Precourt Center for Energy Efficiency at Stanford and its Director, Prof. J. Sweeney, for financial and intellectual support. R. Rajagopal would like to thank B. Smith at PG&E, Dr. A. Narayan at AutoGrid, and Profs. P. Varaiya and D. Callaway at UC Berkeley. The authors would like to thank C. Armel and ARPA-E for providing the dataset.

References

[1] A. Albert, R. Rajagopal, and R. Sevlian, "Power demand distributions: Segmenting consumers using smart meter data," in *BuildSys/SenSys, Seattle, WA*, Nov. 2011.
[2] C. Alzate, M. Espinoza, B. Moor, and J. A. Suykens, "Identifying customer profiles in power load time series using spectral clustering," in *Proc. 19th Int. Conf. Artificial Neural Networks: Part II, ICANN '09*, Berlin, Germany, 2009, pp. 315–324, Springer-Verlag.
[3] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, p. 207, 2005.
[4] M. Bicego, V. Murino, and M. A. T. Figueiredo, "A sequential pruning strategy for the selection of the number of states in hidden markov models," *Pattern Recognit. Lett.*, vol. 24, no. 9–10, pp. 1395–1407, Jun. 2003.
[5] E. Bitar, R. Rajagopal, P. Khargonekar, K. Poolla, and P. Varaiya, "Bringing wind energy to market," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1225–1235, Aug. 2012.
[6] J. Bulla, Application of Hidden Markov Models and Hidden Semi-Markov Models to Financial Time Series, 2006, tech. rep.
[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. New York, NY, USA: Wiley-Interscience, 2006.
[8] T. Fawcett, Roc Graphs: Notes and Practical Considerations for Researchers, 2004, tech. rep.

[9] V. Figueiredo, F. Rodrigues, Z. Vale, and J. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 596–602, May 2005.

[10] C. Flath, D. Nicolay, T. Conte, C. van Dinther, and L. Filipova-Neumann, "Cluster analysis of smart metering data—An implementation in practice," *Business Inf. Syst. Eng.*, vol. 4, no. 1, 2012.

[11] N. Gomes, D. Merugu, G. O'Brien, C. Mandayam, J. S. Yue, B. Atikoglu, A. Albert, N. Fukumoto, H. Liu, B. Prabhakar, and D. Wischik, "Steptacular: An incentive mechanism for promoting wellness," in *COMSNETS*, K. K. Ramakrishnan, R. Shorey, and D. F. Towsley, Eds., 2012, pp. 1–6, IEEE.

[12] W. F. Graf, Lagged Effects on Temperature on Residential Electricity Demand in Southern California, Department of Agricultural and Resource Economics, tech. rep.

[13] W. H. Greene, *Econometric Analysis*. Englewood Cliffs, NJ, USA: Prentice Hall, Aug. 2002.

[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, ser. Springer series in statistics. New York, NY, USA: Springer, 2009.

[15] A. Henley and J. Peirson, "Non-linearities in electricity demand and temperature: Parametric versus non-parametric methods," *Oxford Bull. Econ. Statist.*, vol. 59, no. 1, pp. 149–62, Feb. 1997.

[16] S. Houde, A. Todd, A. Sudarshan, J. Flora, and K. C. Armel, "Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence," *Energy Policy*, Jul. 2012.

[17] T. Jebara, Y. Song, and K. Thadani, "Spectral clustering and embedding with hidden Markov models," in *Proc. 18th Eur. Conf. Machine Learning, ECML '07*, Berlin, Germany, 2007, pp. 164–175, Springer-Verlag.

[18] R. H. Katz, D. E. Culler, and S. Sanders *et al.*, "An information-centric energy infrastructure: The Berkeley view," *Sustain. Comput.: Informat. Syst.*, vol. 1, pp. 7–22, 2011.

[19] A. Kavousian and R. Rajagopal, A Method to Disaggregate Structural and Behavioral Determinants of Residential Electricity Consumption, manuscript under revision, 2012.

[20] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *SDM*, 2011, pp. 747–758, SIAM/Omnipress.

[21] J. Z. Kolter and J. Ferreira, "A large-scale study on predicting and contextualizing building energy usage," in *AAAI*, W. Burgard and D. Roth, Eds. Menlo Park, CA, USA: AAAI Press, 2011.

[22] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," *J. Mach. Learn. Res.—Proc. Track*, vol. 22, pp. 1472–1482, 2012.

[23] C. Li and G. Biswas, "Applying the hidden Markov model methodology for unsupervised learning of temporal data," *Int. J. Knowledge-Based Intell. Eng. Syst.*, vol. 6, no. 3, pp. 152–160, 2002.

[24] I. MacDonald and W. Zucchini, *Hidden Markov and Other Models for Discrete—Valued Time Series: A Practical Introduction Using R*, ser. Monographs on Statistics and Applied Probability. New York, NY, USA: Taylor & Francis, 1997.

[25] S. J. Moss, "Market segmentation and energy efficiency program design," *CIEE Behavior and Energy Program*, 2008.

[26] O. Netzer, J. M. Lattin, and V. Srinivasan, "A hidden markov model of customer relationship dynamics," *Market. Sci.*, vol. 27, no. 2, pp. 185–204, 2008.

[27] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Information Processing Systems 14*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds. Cambridge, MA, USA: MIT Press, 2001.

[28] J. O'Connell and S. Höjsgaard, "Hidden semi Markov models for multiple observation sequences: The MHSMM package for R," *J. Statist. Softw.*, vol. 39, no. 4, pp. 1–22, 2011.

[29] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[30] T. Räsänen and M. Kolehmainen, "Feature-based clustering for electricity use time series data," in *Proc. 9th Int. Conf. Adaptive and Natural Computing Algorithms, ICANNGA'09*, 2009, pp. 401–412, Springer-Verlag.

[31] P. C. Reiss and M. W. White, Household Electricity Demand, Revisited, Working Paper 8687, National Bureau of Economic Research, 2001.

[32] P. C. Reiss and M. W. White, Demand and Pricing in Electricity Markets: Evidence From San Diego During California's Energy Crisis, Working Paper 9986, National Bureau of Economic Research, 2003.

[33] B. A. Smith, J. Wong, and R. Rajagopal, "A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting," *ACEEE Summer Study on Energy Efficiency in Buildings*, 2012.

[34] P. Smyth, "Clustering sequences with hidden markov models," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 1997, pp. 648–654.

[35] C. Song, Z. Qu, N. Blumm, and A.-L. Barabsi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[36] I. Visser, "depmix: An r-package for fitting mixture models on mixed multivariate data with Markov dependencies," R-Package Manual, 2007.

[37] H.-W. Yeh, W. Chan, and E. Symanski, "Intermittent missing observations in discrete-time hidden Markov models," *Commun. Statist.—Simulat. Computat.*, vol. 41, no. 2, pp. 167–181, 2012.

[38] J. Yin and Q. Yang, "Integrating hidden markov models and spectral analysis for sensory time series clustering," in *Proc. 5th IEEE Int. Conf. Data Mining*, Nov. 2005, 8 pp..

[39] S.-Z. Yu, "Hidden semi-Markov models," *Artif. Intell.*, vol. 174, no. 2, pp. 215–243, 2010.

[40] W. Zucchini and I. MacDonald, *Hidden Markov Models for Time Series: An Introduction Using R*, ser. Monographs on Statistics And Applied Probability. Boca Raton, FL, USA: CRC, 2009.

Adrian Albert received the B.S. and M.S. degrees in astrophysics, physics, and computational science from Jacobs University, Bremen, Germany, and the M.S. degree in management science and engineering from Stanford University, Stanford, CA, USA. He is currently pursuing the Ph.D. degree in electrical engineering at Stanford University. His dissertation focuses on modeling and characterizing variability in demand in large-scale infrastructure systems, with an emphasis on energy and the smart grid.

His research interests include applications of sensing, data mining, machine learning, stochastic control, and network analysis to improving the operations of natural resource systems and human-centric systems. Prior to his work at Stanford he has worked in business analytics research for Pixar Animation Studios and sensor data analysis and imaging for NASA and IBM.

Ram Rajagopal received the bachelor's degree in electrical engineering from the Federal University of Rio de Janeiro, Brazil, the master's degree in electrical and computer engineering from University of Texas, Austin, TX, USA, and the M.A. degree in statistics and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA.

He is an Assistant Professor of Civil and Environmental Engineering at Stanford University, Stanford, CA, USA, where he directs the Stanford Sustainable Systems Lab (S3L), focused on large scale monitoring, data analytics and stochastic control for infrastructure networks, in particular energy and transportation. His current research interests in power systems are in integration of renewables, smart distribution systems and demand-side data analytics. Prior to his current position he was a DSP Research Engineer at National Instruments and a Visiting Research Scientist at IBM Research. He holds more than 30 patents from his work, and has advised or founded various companies in the fields of sensor networks, power systems and data analytics.

Dr. Rajagopal is a recipient of the Powell Foundation Fellowship, Berkeley Regents Fellowship and the Makhoul Conjecture Challenge award.