A Scalable Stochastic Model for the Electricity Demand of Electric and Plug-In Hybrid Vehicles

Mahnoosh Alizadeh, Anna Scaglione, Jamie Davies, and Kenneth S. Kurani

***Abstract*—In this paper we propose a stochastic model, based on queueing theory, for electric vehicle (EV) and plug-in hybrid electric vehicle (PHEV) charging demand. Compared to previous studies, our model can provide 1) more accurate forecasts of the load using real-time sub-metering data, along with the level of uncertainty that accompanies these forecasts; 2) a mathematical description of load, along with the level of demand flexibility that accompanies this load, at the wholesale level. This can be useful when designing demand response and dynamic pricing schemes. Our numerical experiments tune the proposed statistics on real PHEV charging data and demonstrate that the forecasting method we propose is more accurate than standard load prediction techniques.**

***Index Terms*—Electric vehicles, load forecasting, load modeling, queueing theory, statistics.**

I. Introduction

DUE TO environmental and economic factors, the use of electric vehicles (EV) and plug-in hybrid EVs (PHEV) is expected to rise considerably in the near future [1]. Hence, in recent years, several researchers have examined the potential impacts of a large-scale integration of EVs on the power grid (for a comprehensive review see [2]). For example, in [3], PJM ran market model simulations to determine the effects of the addition of one million EVs to its load under 3 different load management scenarios. In [4], the electricity demand of the CAISO CAMX (California Independent System Operator for the California-Mexico Power Area) region, which accounts for 80% of the load of California, is examined under several EV adoption rates. In [5], [6], the effects of EVs on distribution network losses and voltage levels are examined.

Mapping a given level of EV penetration into a diurnal load pattern requires a *model* for customer charging behavior. The first step in providing such a model is to answer the following essential questions: 1) when do vehicles arrive where a charger is available? 2) how often do customers request battery charge when their vehicle is parked? 3) how much energy is required per each charge event? 4) how much flexibility accompanies each charge request?

Due to the scarcity of real-world data, and pressed by the need to assess the effects of EV load on the grid, most of the present literature hypothesizes simple models of aggregate arrival rates of EVs and their requests to charge, sometimes using internal combustion engine vehicles (ICEVs) travel patterns. Historical ICEV travel patterns are used in [7], [8], which propose a set of rules to map ICEV data into synthetic EV load traces that can then be used for Mote Carlo simulations. Attempts at mathematically modeling the stochastic process of battery charging vary in sophistication. For example, in [4], EV arrival times are modeled using a normal distribution with a mean of 6 P.M. and a standard deviation of 30 minutes. Both [9] and [10] develop a complete stochastic model for the load, by modeling arrivals of charge requests as a Poisson process and using queueing theory to derive the statistics of the load. However, neither of these models are based on real data. In fact, [9] simulates the arrival of vehicles as a homogeneous Poisson process and proposes to model the length of charge requests as an exponentially distributed. These, as well as conjectures in [10], are not supported neither by the real charging data, nor by the synthetic traces generated in [7], [8]. We showcase these discrepancies in our numerical results.

Our Contribution: In this paper, we answer the essential questions mentioned above, observing the characteristic features of real PHEV charging data [11], and provide a stochastic *mathematical model* for EV/PHEV aggregate load. Having access to such a model has key benefits that go beyond assessment of impacts of future EV load on the power grid via dynamic simulations. The model is useful for providing more accurate *short-term load forecasting*, especially when real-time sub-metering data is available. Crisper short-term forecasts of the volatile charging load of EVs in real-time can help the system operator in dispatching generation with the lowest possible costs in the electricity market and becoming less dependent on ancillary services. Furthermore, modeling the statistics of the load can facilitate planning and operation for a *demand response* (DR) aggregator that manages battery charging of vehicles. There is a growing literature on EV demand scheduling, e.g., [12]–[15], that relies on such statistics. Last but not least, thanks to its scalability, our model helps describe the flexibility of EV/PHEV electricity demand to the system operator (see [16] for a detailed explanation).

Note that our model presumes that we are looking at a large population of vehicles (theoretically infinite), with probability of a charging in the interval $[t, t+dt)$ that is in the order of $\lambda(t)dt$. The rate function $\lambda(t)$ captures aggregate customer behavioral characteristics, and can vary widely if one considers a small population. Hence, this model is not suitable to capture the PHEV/EV load at a feeder for a few houses, but can be used for substations and charging stations. Also, we leave for future work the inclusion of geographical information that would

Manuscript received June 29, 2012; revised December 05, 2012, March 25, 2013; accepted July 18, 2013. Date of publication September 11, 2013; date of current version February 14, 2014. This work was supported by DOE under CERTS. Parts of this work were presented at Allerton 2010. Paper no. TSG-00405-2012.

M. Alizadeh and A. Scaglione are with the Department of Electrical and Computer Engineering, University of California Davis, Davis, CA 95616 USA (e-mail: malizadeh@ucdavis.edu).

J. Davies and K. S. Kurani are with the Institute of Transportation Studies, University of California Davis, Davis, CA 95616 USA.

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/TSG.2013.2275988

allow us to understand the fraction of demand supported by a certain portion of the grid, and therefore ensure that physical power system constraints are met (e.g., transformer capacity limits). Either through the car global positioning system receiver, or through smart-phones applications, the vehicle location can be made available, and it will enrich the modeling and prediction engine. Location data would also allow us to estimate carbon dioxide emissions associated with the increase in electricity generation to charge the vehicles.

The paper is organized as follows. In Section II, we provide a queuing model to capture the load due to EV/PHEV charging that is not centrally controlled. We showcase how this model can be used for load forecasting in Sections II-A and II-C. Next, in Section III, we describe how the statistical information required to use our model can be estimated from sub-metering data. In Section IV-A, we introduce a scalable architecture for telemetry, monitoring, and management of charging requests. We argue that our model can fit arrival and charging data that are influenced by DR strategies, most notably pricing and direct load management (respectively in Sections II-B and IV-B). Next, we proceed to give preliminary numerical estimates of the model parameters using real-world PHEV home charging data in Section V. Lastly, in Section VI, we numerically assess the performance of our real-time forecasting algorithm.

II. Uncontrolled EV Load Modeling

Queuing theory is often used to mathematically analyze the effects of customers randomly arriving and being served by a system. The process of EVs being randomly plugged in and charged by the grid fits this paradigm. To model the charging load imposed by EVs on the grid, we assume that upon plugging in their vehicle, customers pick from a set of possible charging levels. The instantaneous charging power (rate) then remains constant throughout the charge. For example, standard charging rates for electric vehicles include level 1 (home) charging with a 1.1 kW rate, and 3.3 kWs or 6.6 kWs for level 2 (home and workplace) charging. To capture the temporal variations of demand due to EV charging that is not centrally controlled, we utilize a set of $M_t/GI_t/\infty$ service systems, each representing a different charging rate. For example, with the three mentioned charging rates, the aggregate demand can be modeled as the workload of three service systems. Scenarios that fall into this category include unmanaged EV charging load, and EV load under rate-based load shifting incentives such as dynamic pricing.

Queueing service models are traditionally represented using Kendall's shorthand $A/B/C$ notation, in which the first term (A) denotes the interarrival time distribution, which is a time-dependent Markovian (Poisson) distribution in our case, represented by the standard notation M_t. The second term (B) denotes the service time (charging duration) distribution, which is a general time-dependent distribution here, represented by GI_t. Note that the service times are assumed to be independent and identically distributed. Finally, the third term (C) is the number of servers. The number of servers in the model $M_t/GI_t/\infty$ is assumed to be infinite. This is because, when no central demand control technique is exercised, each EV is provided with energy immediately as it arrives in the system and *no queue is actually ever formed*. The number of customers receiving service from each queue, multiplied by the associated charging rate of that queue, provides us with the total power consumption of vehicles served under that queue.

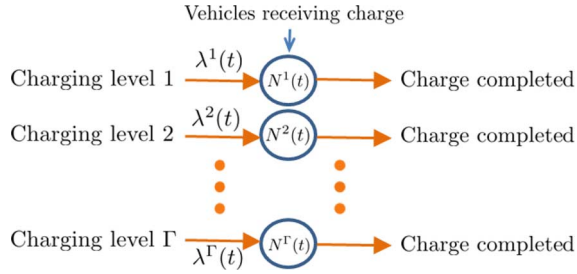

Fig. 1. Γ parallel $M_t/GI_t/\infty$ service systems for vehicle battery charge.

Our model is based on the following elements:

1) **Assignment to queues**: we assume that customers may use one of Γ distinct charging rates, which, as mentioned above, translates into being assigned to *one of* Γ *service systems*. We denote the charging power of the vehicles using charging rate γ as $\bar{R}^\gamma$, and assume that this power is constant throughout the charge cycle. Thus, if we denote the total number of vehicles receiving service from the γ-th queue at time t as $N^\gamma(t)$ (see Fig. 1), we can write the aggregate charging load at time t as

$$L(t) = \sum_{\gamma=1}^{\Gamma} \bar{R}^\gamma N^\gamma(t), \tag{1}$$

2) **Poisson arrivals**: we model the events of vehicles arriving in queues to request charge as a point process (a series of random arrival times t_j^a, with separate events indexed by j). Thus, for arrivals in each service queue, we consider a non-homogeneous Poisson process with a random arrival rate $\lambda^\gamma(t)$. We show in Section V-E why we think this is an appropriate model for charging events. In a Poisson process, the arrival of one customer carries no information about the arrival of others;
3) **Service times**: each plugged-in vehicle has an associated duration of charge S. We denote by $F_S^{t,\gamma}(s)$ the cumulative distribution function (CDF) of the duration of charge S required by each arriving vehicle using the γ-th charging rate at time t, considered to be independent and identically distributed for customers in the same queue.

Next, using assumptions 2 and 3, we find the statistics of $N^\gamma(t)$ and thus, that of the charging load in (1). As mentioned, these statistics could be beneficial for load forecasting and demand response scheduling.

A. Forecasting Load on the Previous Day or Earlier

Here we forecast the load by calculating its statistics, including its expected value, which we denote by $\mathbb{E}[L(t)]$. Note from (1) that we will have a full statistical description of $L(t)$ if we can learn the probability distribution of $N^\gamma(t)$. Since the arrival rate of customers in queues and their charge durations is non-homogeneous, the statistics of $N^\gamma(t)$ are time varying. We begin by assuming that the system started at $t = -\infty$. Then, we can use results in [17] to find the statistics of $N^\gamma(t)$. This assumption is usually made for initializing non-stationary models and is appropriate for forecasting EV charging load

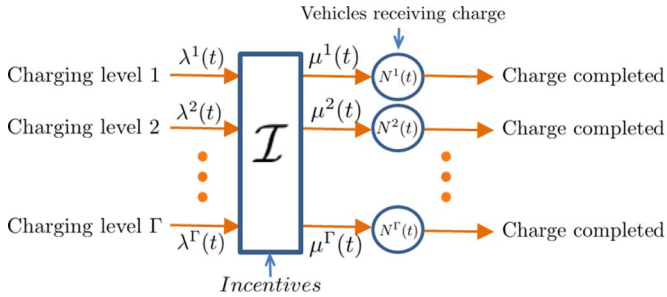

Fig. 2. Arrival rates are modified by setting load shifting rate incentives.

on the day-ahead, since EV charging cycles are limited to a maximum of 10–12 hours. The most interesting result for a $M_t/GI/\infty$ system, proven by [18], [19], is that $N^\gamma(t)$ is a *Poisson random variable* with mean $m^\gamma(t)$ given by:

$$m^\gamma(t) = \mathbb{E}[N^\gamma(t)] = \int_{-\infty}^{t} \left(1 - F_S^{t,\gamma}(t-u)\right) \lambda^\gamma(u) du, \quad (2)$$

which we can calculate if we are provided with estimates of $\lambda^\gamma(t)$ and $F_S^{t,\gamma}(s)$. The integral is basically summing up the expected number of arrivals at all times $u < t$ which are expected to still be receiving service at time t, i.e., their service time is expected to be longer than $t - u$.

We can now combine (1) and (2) to say that the aggregate load $L(t)$ of the vehicles on the next day is the sum of Γ scaled Poisson random variables, with time varying mean

$$\mathbb{E}[L(t)] = \sum_{\gamma=1}^{\Gamma} \bar{R}^\gamma m^\gamma(t). \quad (3)$$

B. The Effect of Demand Response Techniques: Rate-Based Load Shifting Incentives

It is widely believed that without proper management, the load due to electric vehicles can cause price spikes and decrease safety margins in the power grid [20]. In order to study the effects of rate-based charge incentives at specific times or locations on the aggregate load, we need to further expand our model to capture how these incentives can influence the arrival rates $\lambda^\gamma(t)$, i.e., how customers make a decision to plug in their EV in response to dynamic tariffs. Thus, we define a new mapping that models these influences:

$$\mathcal{I} : \{\lambda^\gamma(t), \gamma = 1, \ldots, \Gamma\} \mapsto \{\mu^\gamma(t), \gamma = 1, \ldots, \Gamma\},$$

where $\mu^\gamma(t)$ denotes the modified customer arrival rates due to the incentives (see Fig. 2). The mapping $\mathcal{I}$ can in general be nonlinear, have memory, and be dependent on various parameters such as location, grid conditions, time of day, etc. Customers may be incentivized to change their charging behavior through dynamic pricing or cheaper charging rates available at work or public charging stations.

After learning the mapping $\mathcal{I}$, potentially from empirical econometric studies, one can simply use (2) and (3) to map these modified arrival rates to aggregate load:

$$\mathbb{E}[L(t)] = \sum_{\gamma=1}^{\Gamma} \bar{R}^\gamma \int_{-\infty}^{t} \left(1 - F_S^{t,\gamma}(t-u)\right) \mu^\gamma(u) du. \quad (4)$$

Note that there exists another category of demand response techniques where vehicle charging is directly controlled by a central control unit. The effects of these techniques on our model and on the forecast will be later studied in Section IV-B.

Next, we use this same model to forecast the load accurately in real time.

C. Real-Time Load Forecasting

During real-time operations, the forecast unit can use a metering infrastructure to harness side information about the exact duration and rate of charge requested by vehicles once they are plugged in. Thus, to provide an accurate load forecast, the distribution of $N^\gamma(t)$ shall no longer be initialized at $t = -\infty$. Rather, the forecast unit can use the latest information available on $N^\gamma(t)$. Here we assume that the information on the queue index γ and the charge length S for each charge event is revealed through sub-metering when the car is plugged in, since the initial battery level allows to establish beforehand how much time is needed for a full charge. When a full charge is not possible, we assume that the amount of charge that can be delivered before the customers depart is explicitly revealed using sub-metering data.[1] This is important because there is no uncertainty about the service time of each vehicle. Hence, through telemetry, it is possible to determine *exactly* how many cars, among the ones that have previously started charging, will remain active at a future time t. Thus, the only uncertainty in load forecasting will be represented by the incoming new requests between the current time $t_\circ$ and future time t.

This approach is what we call the Observable Arrival Information *(OAI) predictor*. Specifically, let the present time be $t_\circ$, and $N^\gamma(t_\circ)$ be the customers present in the γ-th queue. Then the OAI predictor models the number of cars present in the system at time $t > t_\circ$ as:

$$N^\gamma(t) = N^\gamma_{\text{new}}(t) + \nu^\gamma(t), \quad (5)$$

where $N^\gamma_{\text{new}}(t)$ is the number of vehicles that arrived in the queue after $t_\circ$ and are actively charging at time t. $N^\gamma_{\text{new}}(t)$ is a random variable and the problem of determining its distribution is similar to the one we looked at in Section II-A. Thus, $N^\gamma_{\text{new}}(t)$ is a Poisson random variable with the mean given in 2, with $\lambda^\gamma(t)$ set to zero for $t \le t_\circ$. The term $\nu^\gamma(t)$, on the other hand, is deterministic and equal to $N^\gamma(t_\circ)$ minus the number of cars that departed (i.e., completed their charge) before time t. We need to calculate this term using sub-metering information. Let us denote the set of cars that arrived in the γ-th queue before time $t_\circ$ by $\mathcal{P}^\gamma_{t_\circ}$, and the arrival time and charge duration associated with the j-th charge request as (t^a_j, S_j). Ignoring the communication network delay, we can write $\nu^\gamma(t)$ as,

$$\nu^\gamma(t) = \sum_{j \in \mathcal{P}^\gamma_{t_\circ}} u\left(S_j - (t - t^a_j)\right), \quad t \ge t_\circ \quad (6)$$

[1]If the departure time of the customer is considered random, this technique will still work but yet another level of uncertainty should be modeled (jobs dropping out of the queue, before they are fully served).

with $u(.)$ as the unit step function. Given this, the real-time forecast of load due to battery charging is given by

$$\mathbb{E}[L(t)] = \sum_{\gamma=1}^{\Gamma} \bar{R}^{\gamma} \mathbb{E}[N^{\gamma}(t)] = \sum_{\gamma=1}^{\Gamma} \bar{R}^{\gamma} \left(\mathbb{E}\left[N^{\gamma}_{\text{new}}(t)\right] + \nu^{\gamma}(t) \right) \tag{7}$$

Notice that in order to use (3) and (7) to provide day-ahead and real-time load forecasts, the following information is required:

1) The PDF of charge lengths for each charging rate (queue), $F_S^{t,\gamma}(s)$;
2) Forecasts of future arrival rates of vehicles to request charge at different charging rates (queues), $\lambda^{\gamma}(t)$;
3) The triplet (γ, S_j, t_j^a) for each vehicle already plugged in, acquired from sub-metering. This information is specifically required to calculate $\nu^{\gamma}(t)$.

In the next section, we address how a load forecast unit can acquire the required statistics mentioned above, namely $F_S^{t,\gamma}(s)$ and $\lambda^{\gamma}(t)$, from real-world data. These estimates should be updated *adaptively* to account for temporal and geographical variations in customer behavior. Besides $F_S^{t,\gamma}(s)$ and $\lambda^{\gamma}(t)$, one can also look at the statistics of laxity (flexibility) of charge requests, which is not directly used by the forecast unit but is relevant for demand response purposes. After addressing these in Section, in Section IV-A, we propose scalable model to capture and aggregate the triplets (γ, S_j, t_j^a) with low communication and storage requirements.

III. Learning the Statistics

Given a large sample set of submetered data on vehicle charge requests, estimating $F_S^{t,\gamma}(s)$ is a straightforward task. We can either provide rescaled histograms, or find analytical distributions that fit the data well enough. In order to provide non-homogeneous distributions, we can assign different distributions to different hours of the day or week.

However, the real challenge lies in estimating future values of the arrival rates $\lambda^{\gamma}(t)$. The complexity is twofold: 1) current values of $\lambda^{\gamma}(t)$ are not observable. The information we can gather from the metering infrastructure provides the number of vehicles that actually plugged in and requested charge from a certain queue γ, which is *a specific realization* of the random Poisson arrival process with mean $\lambda^{\gamma}(t)$; 2) future values of $\lambda^{\gamma}(t)$ are random and correlated with its historical values. Fortunately, the same problem is faced in researching techniques to forecast future call volumes for call centers with stochastic inhomogeneous demand [21]–[24]. The most popular model for the call arrivals is an inhomogeneous (or piece-wise constant) Poisson process with a random arrival rate [23]. We show in Section V-E why we think this model is also appropriate for EV charging events.

Thus, next, given historical information about the number of vehicles plugged in at each charging rate, we wish to estimate the arrival rate $\lambda^{\gamma}(t)$ for a future time t. For brevity and ease of notation, in the rest of this section, we drop the superscript γ and estimate $\lambda(t)$ for only one charging rate.

A. Estimating and Forecating $\lambda(t)$ From Historical Data

We wish to estimate the inhomogeneous rate function $\lambda(t)$ that fully specifies the statistics of the EV Poisson arrival process. Since the customer behavior depends on many underlying factors, surely this rate is not best modeled as a deterministic function of the day of the week. It will have a random evolution that we need to predict on a daily basis, similar to classical load forecasting techniques. The assumption that this rate is random, on top of the random nature of Poisson arrivals, leads to a doubly stochastic model for the arrivals. Here, we approximate the arrival rate as being piecewise constant for intervals lasting (24 hours)/K (e.g., quarter or half hours) and we divide the day into K discrete epochs. Consequently, we denote the arrival rates on day j by a vector $\boldsymbol{\lambda}_j = [\lambda((j-1)K+\ell)]_{\ell=0,\ldots,K-1}$.

The goal of this section is to provide a method to forecast the vector $\boldsymbol{\lambda}_j$ for a future day. Successive values of the arrival rate are not independent and give valuable information for predicting future values of the series. One important tool to predict future values of such sequences is time series analysis [25]. However, since the rate profiles are unobservable, we need to build our forecasting model on the corresponding *count profiles*, which for day j are in our model given by the vector $\mathbf{c}_j = [c((j-1)K+\ell) \sim \text{Poiss}(\lambda((j-1)K+\ell))]_{\ell=0,\ldots,K-1}$. Note that our ultimate goal in observing the count profiles $\mathbf{c}_j$ is to estimate the rate vectors $\boldsymbol{\lambda}_j$ of our statistical model.

Since the counts $c(t)$ are realizations of Poisson random variables with different means $\lambda(t)$, their associated variances, which is identical to the mean values, is time-varying as well. In order to solve numerical issues that arise from this, we can use a simple variance-stabilizing transformation. For example, we can follow the suggestion in [23] and use a slightly modified version of the Anscombe square-root transform [26], which transforms Poisson distributed data like $c(t)$ to approximately normally distributed data with a constant variance of 1/4. The transformation is as follows:

$$c(t) \sim \text{Poiss}(\lambda(t)) \overset{\text{approx.}}{\longrightarrow} x(t) = \sqrt{c(t) + \frac{1}{4}} \sim \mathcal{N}\left(\theta(t), \frac{1}{4}\right), \tag{8}$$

where $\theta(t) = \sqrt{\lambda(t)}$. Following our previous notation, we also denote the vectors containing transformed counts $x(t)$ and their means $\theta(t)$ for day j as $\mathbf{x}_j$ and $\boldsymbol{\theta}_j$.

Now, there are two types of forecasts that have to be performed: 1) predict the arrival rates for the next day from historical data. This is referred to as inter-day forecasting; 2) dynamically update the forecasts during the day as new data becomes available in real-time. We call this intra-day forecast updating. Next, we address how we can provide such forecasts from historical and real-time data.

1) Interday Forecasting: Transformed count vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ representing different days are statistically dependent random variables. However, we consider them conditionally independent given the latent transformed rate vectors $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$. As we saw, $\boldsymbol{\theta}_j$'s are unobservable (latent) variables and need to be estimated. The first step in addressing this problem is to adopt a model that can describe the correlation between

successive values of $\boldsymbol{\theta}_j$'s. Time series models are a popular choice here. However, due to the large dimension of the vectors $\boldsymbol{\theta}_j$, multivariate time series models may not be the best way forward. Thus, to model inter-day dependencies among the $\boldsymbol{\theta}_j$'s, we adopt a model similar to what is proposed in [24] and we first reduce their dimension, using a low dimensional factor model to describe the correlation between the arrival rates on different days. Specifically, we assume:

$$\boldsymbol{\theta}_j = \sum_{i=1}^{\kappa} \alpha_j(i)\mathbf{u}_i, \tag{9}$$

with $\kappa < K$ and $\mathbf{u}_i'\mathbf{u}_j = 0$ for $i \neq i'$. To minimize the residual mean squared error associated with a given κ, $\mathbf{u}_i$'s can be chosen to be the principal components of the variables $\boldsymbol{\theta}_j$. The $\mathbf{u}_i$'s are assumed time-invariant in (9) and have a much slower dynamics than $\alpha_i(j)$. Alternatively, if we define the vector $\mathbf{a}_j = [\alpha_j(1), \ldots, \alpha_j(\kappa)]^T$ and the matrix $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_\kappa]$, (9) can be written as

$$\boldsymbol{\theta}_j = \mathbf{U}\mathbf{a}_j. \tag{10}$$

This approach maps the latent vector variable θ_j into κ scalar latent variables $\alpha_j(i)$. The correlation between successive values of $\boldsymbol{\theta}_j$ is modeled through the $\alpha_j(i)$. Due to the characteristics of singular value decompositions, a vector containing consecutive values of $\alpha_j(i)$'s for different days j would be orthogonal to that of $\alpha_j(m)$ for $m \neq i$. Thus, it is reasonable to assume that $\alpha_j(i)$ and $\alpha_j(m)$ are uncorrelated and the series $\alpha_j(i)$ can be modeled using separate univariate time series models. Denote by the index $t_j = 1, \ldots, 7$ the day of the week corresponding to day j. In our model we assume that $\alpha_j(i)$'s will have a periodic mean of $\bar{\alpha}_{t_j}(i)$, corresponding to the day of the week. Then, if we adopt an AR(1) model for each $\alpha_j(i)$, we can write

$$\alpha_j(i) - \bar{\alpha}_{t_j}(i) = \beta_i \left(\alpha_{j-1}(i) - \bar{\alpha}_{t_{j-1}}(i)\right) + w_j(i), \tag{11}$$

where $w_j(i) \sim \mathcal{N}(0, \sigma_i^2)$ is white noise and is uncorrelated for different i's. Higher degree models can be adopted if proven to be necessary by real-world data. In an AR(1) model, the transformed daily count vector $\mathbf{x}_j$ is conditionally independent of all other count vectors in the historical dataset given θ_j for day j and the previous day's transformed count vector $\mathbf{x}_{j-1}$. Rewriting (11) for all $i = 1, \ldots, \kappa$ in vector form gives

$$\mathbf{a}_j - \bar{\mathbf{a}}_{t_j} = \mathbf{D}\left(\mathbf{a}_{j-1} - \bar{\mathbf{a}}_{t_{j-1}}\right) + \mathbf{w}_j, \tag{12}$$

where $\bar{\mathbf{a}}_{t_j} = [\bar{\alpha}_{t_j}(1), \ldots, \bar{\alpha}_{t_j}(\kappa)]$, $\mathbf{D} = \mathrm{diag}[\beta_1, \ldots, \beta_\kappa]$, and $\mathbf{w}_j = [w_j(1), \ldots, w_j(\kappa)]$.

With (11), our model is complete. For given values of the parameters $(\sigma_i, \bar{\alpha}_{t_j}(i), \beta_i, \mathbf{u}_i)$, we can estimate the $\boldsymbol{\theta}_j$'s from the transformed count vectors $\mathbf{x}_j$ and use these estimates to forecast future arrival rates. By assigning normal diffuse priors to $\alpha_1(i), i = 1, \ldots, \kappa$, this estimation would be straightforward using classical Bayesian techniques such as maximum a posteriori (MAP) or minimum mean square error (MMSE) estimators [27]. In fact, with an AR(1) model, this can be easily done adaptively on a day-to-day basis, using standard Kalman filtering iterations [28]. Define $\mathbf{y}_j = \mathbf{x}_j - \mathbf{U}\bar{\mathbf{a}}_{t_j}$ as the transformed count vector with the average weekly trend removed. Then, (8), (10), and (12) can be combined to write the dynamics of $\mathbf{y}_j$ as

$$\mathbf{a}_j' = \mathbf{D}\mathbf{a}_{j-1}' + \mathbf{w}_j, \tag{13}$$

$$\mathbf{y}_j = \mathbf{U}\mathbf{a}_j' + \mathbf{v}_j, \tag{14}$$

where $\mathbf{y}_j$'s are the observations, $\mathbf{v}_j$ is zero mean Gaussian white noise $\sim \mathcal{N}(\mathbf{0}, (1/4)\mathbf{I})$ and is independent from $\mathbf{w}_j$, and $\mathbf{a}_j' = \mathbf{a}_j - \bar{\mathbf{a}}_{t_j}$. The dynamical model in (13) is in accordance with the framework of the Kalman filter.

In summary, in order to forecast arrival days for the next day (or later) on day j, one must perform the following steps:

1) Transform the observed count vector $\mathbf{c}_j$ to $\mathbf{x}_j$ via (8);
2) Remove the weekly trend $\mathbf{U}\bar{\mathbf{a}}_{t_j}$ from $\mathbf{x}_j$ to get $\mathbf{y}_j$;
3) Apply the Kalman filter update step to estimate $\mathbf{a}_j'$;
4) Apply the Kalman filter predict step the predict $\mathbf{a}_{j+1}'$;
5) Add the weekly component $\bar{\mathbf{a}}_{t_j}$ to forecast $\mathbf{a}_{j+1}$
6) Map this forecast of $\mathbf{a}_{j+1}$ into a forecast of the transformed rate $\boldsymbol{\theta}_{j+1}$ using (10);
7) Apply the inverse Anscombe transform to the forecast of $\boldsymbol{\theta}_{j+1}$ to get a forecast of the actual rate vector $\boldsymbol{\lambda}_{j+1}$

However, note that we are not provided with the model parameters $(\sigma_i, \bar{\alpha}_{t_j}(i), \beta_i, \mathbf{u}_i)$. We have to learn them from historical data. Due to the presence of latent variables $\boldsymbol{\lambda}_j$, and because of the non-linear nature of the model as a function of the unknown parameters, this is not a straightforward problem to solve analytically. Several approaches exist for calculating maximum likelihood (ML) estimates of the parameters (equivalent to a MAP estimate with a uniform prior). We initialize the model by assigning the first κ singular vectors of the historical transformed count matrix to the vectors $\mathbf{u}_i$. The rest of the parameters can then be estimated using iterative parameter estimation techniques with latent variables (an expectation maximization framework [29]). Note that the eigen-structure of the rate profiles can be better estimated as well once our estimates of the model parameters improve. Thus, the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_\kappa$ are also updated dynamically using recursive subspace tracking methods [30].

2) Intraday Updates: Now we have a forecast of the arrival rates for the next day (indexed by $j+1$). During real-time operations, the aggregator in charge of the vehicles will gradually observe the realized values of the random daily arrival counts, transformed to $\mathbf{y}_{j+1}$, through the metering infrastructure. At hour m of day $j+1$, the first m elements of $\mathbf{y}_{j+1}$ is available, which we denote by $\mathbf{y}_{j+1}^{(m)}$. As these entries become available, the load forecast unit can update its forecast of $\boldsymbol{\lambda}_{j+1}$. The update can be carried out in two steps:

1) *Direct estimation*: if we denote by $\mathbf{U}^{(m)}$ the first m rows of $\mathbf{U}$ and by $\mathbf{v}_{j+1}^{(m)}$ the first m elements of $\mathbf{v}_{j+1}$, then

$$\mathbf{y}_{j+1}^{(m)} = \mathbf{U}^{(m)}\mathbf{a}_{j+1}' + \mathbf{v}_{j+1}^{(m)} \tag{15}$$

can be used to directly estimate estimates of $\mathbf{a}_{j+1}'$ from $\mathbf{y}_{j+1}^{(m)}$ and thus, $\boldsymbol{\lambda}_{j+1}$ using steps (5), (6) and (7) mentioned above. Denote this estimate by $\hat{\boldsymbol{\lambda}}_{j+1}^{\mathrm{intra}}$.
2) *Penalized updating*: With low amounts of arrival count data, the direct estimation step is not accurate. Thus, we can use a weighted sum of $\hat{\boldsymbol{\lambda}}_{j+1}^{\mathrm{intra}}$ and the day ahead forecast as our estimate of $\boldsymbol{\lambda}_{j+1}$.

IV. Digital (Quantized) Submetering and Dispatch

A. Classification of Charge Requests

In Section II-C, we proposed a real-time forecasting technique in which forecast errors were reduced by incorporating submetering data. We divided the charging load at the future time t into two parts: load due to vehicles that arrived before the current time $t_\circ$ (denoted by set $\mathcal{P}^\gamma_{t_\circ}$), and load due to vehicles that will arrive between $t_\circ$ and t (denoted by set $\mathcal{P}^\gamma_t - \mathcal{P}^\gamma_{t_\circ}$). Thus, the load at a future time is given by

$$\begin{aligned} L(t) &= \sum_{\gamma=1}^{\Gamma} \bar{R}^\gamma N^\gamma(t) \\ &= \sum_{\gamma=1}^{\Gamma} \bar{R}^\gamma \left(N^\gamma_{\text{new}}(t) + \nu^\gamma(t)\right) \\ &= \sum_{\gamma=1}^{\Gamma} \bar{R}^\gamma \left(\sum_{j \in \mathcal{P}^\gamma_{t_\circ}} u(S_j - (t - t^a_j)) \right. \\ &\quad \left. + \sum_{j \in \mathcal{P}^\gamma_t - \mathcal{P}^\gamma_{t_\circ}} u\left(S_j - \left(t - t^a_j\right)\right) \right), \end{aligned} \tag{16}$$

where $\nu^\gamma(t)$ is deterministic and completely observable from submetering data. To track its value and improve the forecasts of $L(t)$ as described in Section II-C, we need the analog triplets (γ, S_j, t^a_j) for each charge request.

But high telemetry and storage costs and modeling complexity are two faces of the same coin. Thus, it is desirable to reduce the burden of communicating and storing the triplet (γ, S_j, t^a_j) for every single vehicle. Similar to the basic principles of all digital communication systems, in [31] we proposed to quantize the tuple (γ, S_j). This is carried out through a set of classifiers $\Psi^\gamma(S_j)$ that quantize the charging durations S of vehicles using the γ-th charging rate onto a set of Q^γ possible values $C^\gamma_q, q = 1, \ldots, Q^\gamma$. The effect of this clustering is twofold: 1) it obviously provides a digital representation of the charging load that can be communicated; 2) it separates the charge requests into $Q = \sum_\gamma Q^\gamma$ classes of service. We denote the class of all vehicles that use a charging rate γ and charge duration q by a tuple (γ, q). This *classification of demand* highly reduces out storage costs, as well as computational costs when using demand response techniques later in Section IV-B. We also choose to discretize time since forecast updates are not carried out continuously and, thus, we do not need to store the state of the queues at every time instant. Hence, from this point on, t is used to denote the index of discrete time epochs equally distanced by $\triangle$.

With this quantized charge description, at each epoch t, the new information that needs to be stored is the number of new arrivals in each class (γ, q) since the previous epoch, which we denote by $a^\gamma_q(t)$. Consequently, $\nu^\gamma(t)$ can now be written as,

$$\nu^\gamma(t) = \sum_{q=1}^{Q^\gamma} \sum_{k=t_\circ - C^\gamma_q}^{t_\circ} a^\gamma_q(k) \Pi\left(\frac{t-k}{C^\gamma_q}\right), \tag{17}$$

where $\Pi(\cdot)$ denotes a unit pulse function between [0,1]. Similarly, if we expand the summation over time beyond the current time $t_\circ$, we can model $N^\gamma(t) = N^\gamma_{\text{new}}(t) + \nu^\gamma(t)$ as,

$$N^\gamma(t) = \sum_{q=1}^{Q^\gamma} \sum_{k=t-C^\gamma_q}^{t} a^\gamma_q(k) \Pi\left(\frac{t-k}{C^\gamma_q}\right), \tag{18}$$

where the arrival counts $a^\gamma_q(t)$ beyond $t_\circ$ are not known yet but can be forecasted.

There are many benefits in using (18) over (16), the most notable of which is scalability. No matter how large the vehicle population is, we only need to store Q numbers per time index t to keep track of the sub-metered load. Also, when simulating EV load, we only need to generate Q Poisson random variables per time index t to generate random vehicle arrivals, which is extremely simple.

B. The Effect of Demand Response Techniques: Controlled Activations

The proposed load classification can also be used to characterize the *dispatchability* of the EV/PHEV load by an aggregator. An aggregator that wants to apply a demand management technique needs to model how its control signals will affect the load. In Section II-B, we saw that the effects of dynamic pricing techniques can be captured through modeling their effects on the aggregate arrival rates $\lambda^\gamma(t)$. Another scenario is that vehicle charge requests can be managed by a central unit that directly controls the exact times at which vehicles charge their batteries.

Under a direct demand management program, the number of requests that start being served by each queue at every time instant is *controlled* and customers that arrive in the grid to receive charge need to *wait* to receive an authentication from the control unit. This can be captured by denoting the number of vehicles that start charging at time t from the (γ, q)-th service class by $d^\gamma_q(t)$, while a number of vehicles are held in the queue (see Fig. 3). With this new definition, we can rewrite the aggregate charging load in (16), under a central demand control program, as

$$L(t) = \sum_{\gamma=1}^{\Gamma} \bar{R}^\gamma \sum_{k=t-C^\gamma_Q}^{t} \sum_{q=1}^{Q^\gamma} d^\gamma_q(k) \Pi\left(\frac{t-k}{C^\gamma_q}\right). \tag{19}$$

Consequently, studying the effects of different demand management techniques on the load would translate into understanding how the customer's initial arrival pattern $a_q(t)$ is affected by the control signal and yields $d_q(t)$. We denote the function that describes this relationship as $F(\cdot)$. The arguments of this function can be as simple as just $a_q(t)$, in the case of, for example, a heuristic threshold based demand management program holding back demand if it exceeds a certain value. At the other extreme, the control algorithm, described by $F(.)$, can have many input arguments and may not be an analytical function, but an optimization problem that should be solved numerically (see e.g., [31]).

The description in (19) can be generalized to also account for vehicle-to-grid (V2G) scenarios. To do so, we can assume that appliances waiting to receive service from each queue can choose between various service options. For example, a vehicle

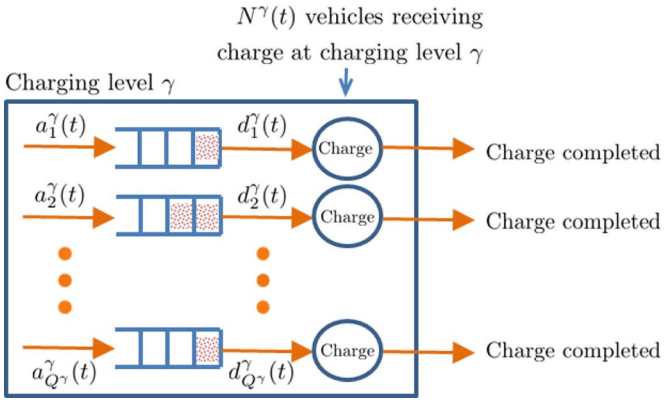

Fig. 3. Quantized sub-metering can help facilitate real-time forecasting and demand response. Each charging rate γ is mapped into Q^γ queues.

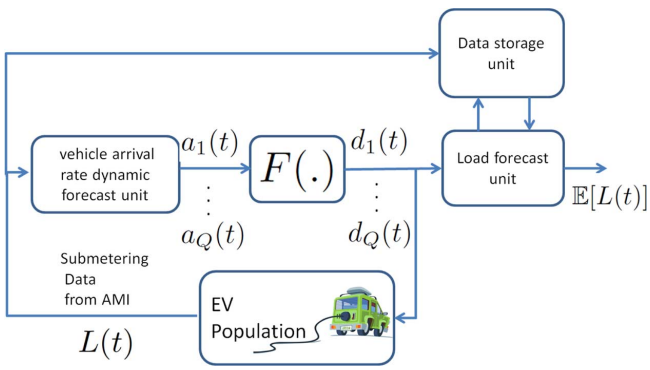

Fig. 4. Smart load forecasting with DR.

requiring charge for one hour can be charged and leave, or it can discharge the battery now and receive a longer charge afterwards. The details of how this can be achieved are beyond the scope of this paper and will be addressed in future work.

Note that, in the case where a DR aggregator is manipulating the access rate of vehicles to receive charge, the OAI predictor, which merely used estimates of the arrival rates $a_q(t)$ to predict the future load, now needs to forecast the values of $d_q(t)$ instead. There are two possible cases: 1) the load forecast unit has an analytical description of how the $d_q(t)$'s will be determined in response to an arrival rate of $a_q(t)$, due to $F(.)$; 2) the forecast unit needs to learn the behavior of each aggregator's demand control scheme from historical data available on the $a_q(t)$ and $d_q(t)$'s, i.e., estimate the function $F(.)$. The required interactions to forecast load in these cases is shown in Fig. 4. Again, this can be easily extended to include multiple charging rates or general charging pulses $g(\cdot)$.

V. Numerical Experiments—Part I: Tuning Model Parameters to Historical Data

Now that we have described how the system operator or an aggregator of electric vehicles can use sub-metering data to model the full statistics of charging demand for load forecasting or demand response purposes, we wish to numerically calibrate the parameters of our model to historical data. We provide these initial estimates with the caveat that they can be further improved when more data is available, considering that they might prove useful as a reference for research purposes. Due to the limitations of the real-world data we have access to, here we only look at level-1 home charging of PHEVs (Γ is equal to 1 since we only have one charging level). Hence, for brevity, we eliminate the subscript γ from our notation from this point on.

A. Data Set Description

We will compute the parameters of our model as follows. The probability density function (PDF) of charge durations and laxity (slack time) of PHEV charging is based on data from PHEVs driven and charged by households [11], provided by the UC Davis PH&EV center. These data confirm that the random number of vehicles being plugged in per unit of time, which we refer to as *vehicle arrival process*, has time-varying statistics. However, these statistics cannot be accurately learned from the PHEV database, due to its small size. Thus, we propose a new methodology to derive the parameters of the arrival model using ICEV data. More specifically, we infer a second distribution of charging events using the 2009 National Household Travel Survey (NHTS) [32], emulating the example of [7] in building synthetic traces.

The NHTS gathers information about daily travel patterns of different types of households. The available data include mode of transportation, duration, distance, and purpose of travel for ICEV owners. The specific parameters found for this portion of the model may be unrealistic for PHEVs and EVs in three ways. First, as [11] and [33] discuss, the probability that a PHEV or EV is charged upon its arrival at a charging location (taken to be home in this instance) 1) decreases with a longer driving range per battery charge; 2) increases with the distance the vehicle has been driven since its last charge; 3) increases with the time duration the vehicle is parked before it is driven again; and 4) increases as a function of how far the vehicle is expected to be driven subsequent to the charge. Because the entries in the NHTS travel data each cover only one day, the last point in particular cannot be incorporated into any charging probability distribution derived from [32], if charging is based on an expectation of the next day's travel. Second, we are not as sanguine as [7] that deviations between the observed travel in [32] and travel by PHEVs and EVs will decline in the future. Observations from [11] and [33] indicate that households change driver assignments to vehicles, trips, and activities, choose new locations, and buy different vehicles in response to the new driving and charging characteristics of PHEVs and EVs. That is, we view it as more likely that any future NHTS containing large numbers of PHEVs and EVs will be different from, not similar to, any data consisting primarily of ICEVs. Third, the amount of charged required by the vehicle is a function of several parameters such as driving/road conditions, driving habits, vehicle model, battery condition, dynamic prices, climate control settings, and user selected modes (charge depleting, blended, or charge sustaining). We cannot account of all of these parameters when mapping NHTS to PHEV charge requests. Although all of these issues will affect the statistics for PHEV charging events derived here, this does not limit our model from being re-parameterized when new larger data sets of PHEV or EV charging become available.

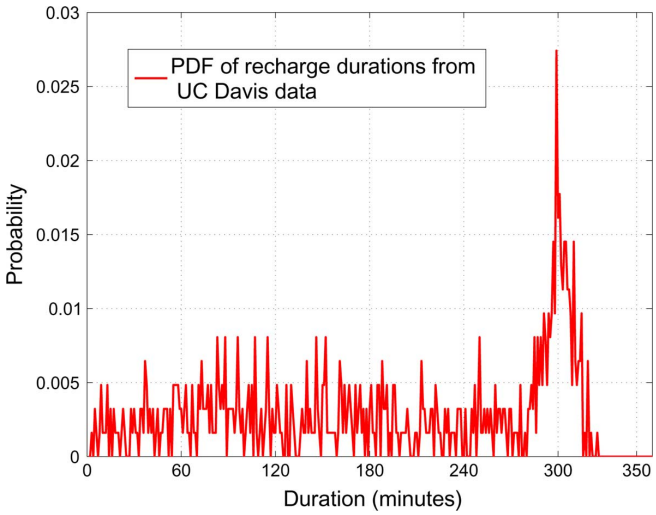

Fig. 5. The PDF of charging durations obtained from PH&EV center data. The noisiness is due to the rather small size of the dataset from UC.

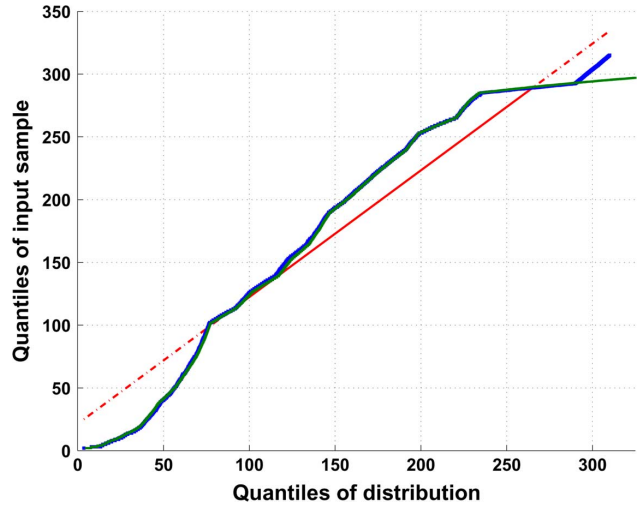

Fig. 6. q-q plot of PHEV charge estimates versus the traditional and the clipped lognormal distributions.

B. Statistics of Charge Duration S

Remember that in this work, the charging period of each vehicle is modeled as an independent identiacally distributed random variable S for vehicles plugged in at the same charging level and time. In this section we provide an appropriate PDF for S for the case of PHEV level-1 home charging. This answers the third essential question posed in the introduction: how much energy is required per each charge event?

Using level 1 charging, the full charging time of our studied PHEVs would be between four and five hours (four to six kWh of battery capacity). The actual power to charge typically starts at higher than the average rate and tapers off as the battery approaches a full charge. We ignore this change (as it has little impact on the duration of charging) and assume the PHEV charging power is equal to an average of $\bar{R} = 1.1$ kW, consistent with our real-world data.

Fig. 5 shows the probability distribution function of charge durations S_j derived from the PH&EV center data ($\mathbb{E}[S_j] = 189$ minutes). We tested numerous distribution functions that seem to fit the charge duration data. We found that a lognormal distribution $\ln \mathcal{N}(5.03, 0.78^2)$ best fits the data. Thus, we conclude that a suitable PDF to model the charge duration of PHEVs under level-1 home charging is

$$f_S^{\text{PHEV}}(s) \approx \frac{1}{0.78\sqrt{2\pi}s} \exp^{-\frac{(\ln s - 5.03)^2}{2*0.78^2}} \qquad (20)$$

where the unit of s is minutes. The expected value of this random variable is 207 minutes. The q-q plot in Fig. 6 compares the distribution to the samples. Another step to further improve this fit is to clip the above PDF at a maximum of 300–350 minutes. To see how much this improves the fit, we cut the PDF at 320 minutes and distributed the samples that were above this value uniformly in [280, 320] min (see Fig. 6).

Here, we cut off the distribution at a specific charge duration, which translates into a specific battery size. In the case of different battery sizes we expect that either a mixture of lognormal RVs or a single distribution with a tail that combines the cutoff effects of different battery capacities will fit the data.

One important issue that we wish to bring forward is that most of the home charging events in this dataset happen after 4 P.M. and continue throughout the evening and night hours. Thus, we do not have a large enough number of samples from charging events that happened during morning hours and the ones we have are shorter than the charges at night hours. Hence, the PDF in (20) is most appropriate to fit night charging events and the distribution $f_S^{\text{PHEV}}(s)$ may not be stationary and could be a function of time. However, when we studied the distribution of mileage driven before returning home from NHTS entries, and divided the night and day hours data, we observe very similar statistics. This suggests that any method converting travel miles into home charges will lead to a stationary distribution for the charge PDF, which at moment is neither confirmed nor denied by the real data. For the sake of illustrating our model, we proceed to use a stationary PDF of the charge requests for home charging, even though our model in Section II is compatible with a non-stationary charge request distribution.

C. Statistics of Laxity of Charge Requests

Next, we look at one parameter that may be interesting to DR aggregators in charge of managing EV charging load. We answer the fourth essential question mentioned in the introduction about the amount of flexibility accompanying each charge request. Fig. 7(a) shows the PDF of the laxity (slack time) of the charging requests gathered by UC Davis [11]. The PDF has two distinct peaks, one at 1–2 hours and the next at 8–10 hours. As seen in Fig. 7(b), we can confirm from the data that daytime requests contribute mostly to the first peak, while nighttime requests mostly belong to the second peak. Consequently, we fit two different probability density functions to represent the laxity offered by daytime and nighttime charge requests. We found that an exponential distribution with a mean of 1.089 best fits the daytime data, while the nighttime laxity is best represented by a lognormal distribution $\ln \mathcal{N}(2.25, 0.4^2)$. Due to a lack of enough samples, we refrain from finding a joint density

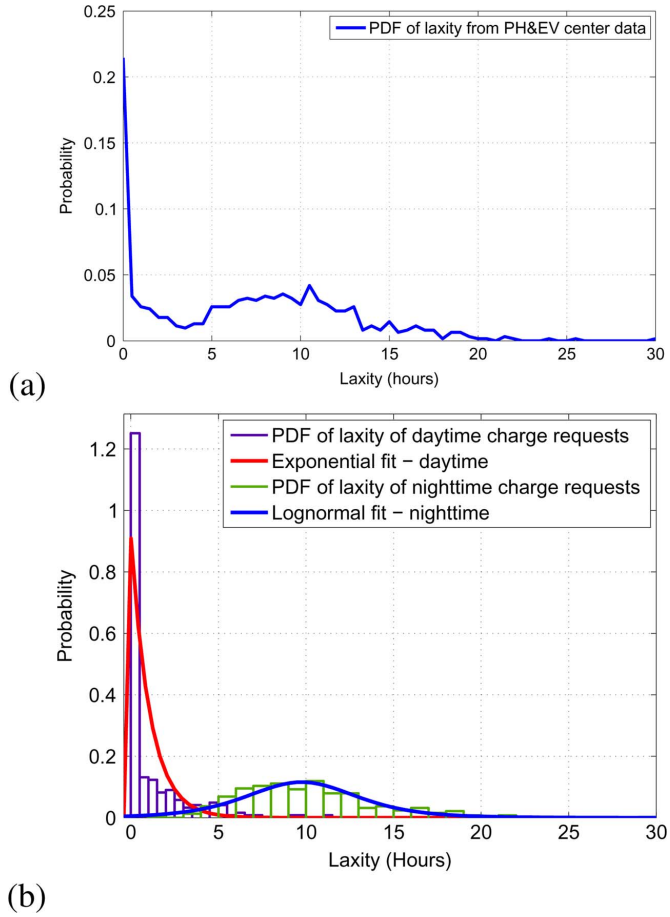

Fig. 7. (a) PDF of laxity of charge requests. (b) PDF of laxity separated for nighttime and daytime requests. Two distinct peak are observed.

for the charge laxity and length. However, we acknowledge that these two variables are most likely correlated.

D. Mapping NHTS Entries into Charge Durations

As mentioned, the size of the real-world data set available to us is rather small (620 samples). While this is a reasonable amount of data to derive the charge duration and laxity PDF, one cannot learn vehicle arrival statistics (the rate $\lambda(t)$) from this data set. One possible way to have more statistics on PHEV travel patterns, and thus arrivals, is to expand our sample set by using a methodology similar to [7], [8] and convert the NHTS mileage records traveled by ICEVs into PHEV arrival and charge durations [32]. Since NHTS entries only record one day worth of travel patterns per household for ICEVs, a common assumption made to convert *miles* to *kWhs* is that vehicles are charged at least once a day. If we adopt this assumption, we end up mapping at least one arrival of the customer at home per day into a battery charge request. To showcase why we think this is possibly inaccurate, we will introduce a simple miles to kWhs conversion next.

We assume that, on average, the energy required per each traveled mile is uniformly distributed as in Table I [34]. Also, we assume that the PHEVs are in a *charge depleting mode* until they run out of charge, when they switch to consuming gas. Therefore, if we denote the mileage traveled by vehicle i as M_i,

TABLE I
ENERGY PER TRAVELED MILE

Vehicle Type	kWh per mile
Sedans	0.18 - 0.3
Vans	0.3 - 0.4
SUVs	0.4- 0.5
Trucks	0.5 - 0.7

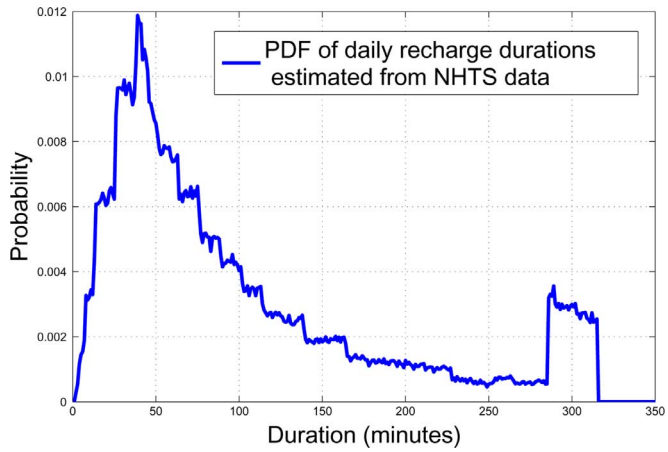

Fig. 8. The PDF of home charging durations obtained from NHTS is clearly not similar to that of the real-world data. We envision that this is because NHTS alone cannot account for the random nature of plug-in events.

the average energy required per mile by ϵ_i, the charge rate by R_i, the battery capacity by Υ_i, and the amount of time the vehicle is parked at home as ρ_i, we can write the mapping from NHTS entries into charge durations S_i as

$$S_i = \min\left\{\frac{\Upsilon_i}{R_i}, \frac{\epsilon_i}{R_i}M_i, \rho_i\right\}. \tag{21}$$

Fig. 8 displays the probability density function of the home charging durations S_i in minutes derived from the mileage entries based on (21). We can see that a large percentage of charge requests have a rather short duration, which is neither consistent with the real-world data (cf. Fig. 4) nor intuitive ($\mathbb{E}[S_i] = 102$ minutes). We envision that this problem will be eliminated if we account for the random nature of plug-in events. In fact, drivers are unlikely to charge at the end of every single day if they have only traveled a rather short distance and their battery is not depleted. The data from the UC Davis PH&EV center confirm that charging usually happens when the battery is nearly empty. Consequently, lacking sufficient arrival data in the PH&EV center dataset, we propose to account for this random plug-in behavior so as to transform NHTS entries to realistic PHEV arrival time data. The NHTS entries, unlike the PHEV data set, are abundant and expand our sample set sufficiently to denoise the estimate.

If customers do not plug in their vehicles every night, the total mileage driven by vehicle i in (21), M_i, will be the sum of daily traveled mileages over the days following the previous charge event. Thus, denoting the miles traveled by vehicle i on the j-th day since its last battery charge by $M_{i,j}$, we have

$$M_i = \sum_{j=1}^{K_i} M_{i,j}, \tag{22}$$

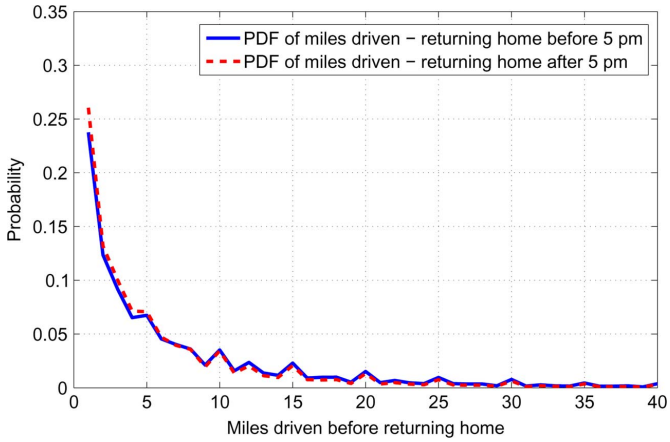

Fig. 9. Distribution of miles driven before returning home, derived from NHTS.

where K_i is the number of days since the customer has last charged the battery. Note that K_i is a random number. We assume a simple Bernoulli model to capture the customers decisions to charge their vehicles, with a few possible options on how to model the success probability. Picking different options would affect the statistics of K_i. Since the NHTS only gives us information on a single day of travel per vehicle i, i.e., one $M_{i,j}$ per household i, we propose two strategies to model K_i.

The first strategy is to assume that the plug-in probability is only a function of the last day's mileage, M_{i,K_i}. We found that by choosing

$$P_{\text{plug}}(M_{i,K_i}) = \begin{cases} 0.05, & M_{i,K_i} < 1.5 \\ 0.1, & 1.5 < M_{i,K_i} < 8 \\ 0.5, & 8 < M_{i,K_i} < 16 \\ 1, & 16 < M_{i,K_i} \end{cases}, \tag{23}$$

a two-sample Kolmogorov-Smirnov test does not reject the hypothesis that the 620 real-world samples and the 150 000 charging amounts derived from the NHTS database come from the same probability distribution (at a 5% significance level). However, note that this may not be the optimal way of capturing the customer's decision, which will most probably depend on M_i instead of $M_{i,j}$.

In the second strategy, instead, we assume that there is a constant probability of plugging in, no matter the miles, equal to P_{plug}. Hence, K_i will be a geometric random variable with a success probability of P_{plug}, i.e.,

$$P(K_i = n) = (1 - P_{\text{plug}})^{n-1} P_{\text{plug}} \tag{24}$$

Next, we need to estimate the parameter P_{plug} based on the real-world data. We can learn the PDF of the miles driven on a single day, $M_{i,j}$, from the NHTS data (see Fig. 9). If we assume that the mileage driven on consecutive days are independent random variables,[2] we can numerically calculate the PDF of M_i for different values of P_{plug}. Next, we record the log-likelihood that the 620 samples available from the PH&EV center data originated from each of these PDFs. The results indicate that our optimal estimate is $P^*_{\text{plug}} = 0.4$.

[2]It is likely this is not the case, and that one would have to account for serial correlation in travel across days for individual drivers, but we are avoiding the complexities associated with making these distinctions.

A further improvement would be to make the probability of plug-in a function of M_i, the number of miles driven since the last charge. This would mean that, if the battery is nearly empty, the customer will charge it with a higher probability. A simple model for this case would be

$$P_{\text{plug}}(M_i) = \begin{cases} P_1, & M_i < \text{Threshold} \\ P_2, & \text{else} \end{cases} \tag{25}$$

If we choose the threshold in (25) to be equal to 5 miles, we can estimate $P_1^* = 0.15$ and $P_2^* = 0.65$ by performing a likelihood ratio test on the real-data.

With this, we have answered the second essential question posed in the introduction: how often do customers request battery charge when their vehicle is parked? Here, we choose to proceed with the plug-in probability (23) to map the event of an ICEV arriving at home in the NHTS database to a PHEV plug-in event for our numerical experiments. Now we assume that we have an abundant amount of historical data on PHEV arrivals. Next, we show why we think a non-homogeneous Poisson process is an appropriate choice for modeling arrivals, as promised previously.

E. Uniform Conditional Test for a Poisson Process

In order to show that an inhomogeneous Poisson arrival model is a reasonable model for PHEVs, we construct a test of null hypothesis based on vehicle travel patterns from the NHTS data. The null hypothesis is that vehicle arrivals for charging follow a constant rate Poisson process in short intervals of time (the length of this interval is chosen to be 30 minutes in our case). To be consistent with our charging model, we assumed that vehicles request charge only when they arrive at home. The decision of customers to plug-in was assumed to be a Bernoulli random variable with a success probability given by (23). It is worth noting that, no matter the choice of plug-in probability, random selections of Poisson processes are still Poisson processes.

Before explaining how we carry out this test, we need to address an issue emerging from the NHTS trip start and end times. These times are almost always rounded to the nearest 5 minutes. Also, we observed that 1/2 hour intervals appear much more often than expected. Therefore, following [23], we added a random dithering to the data by adding normally distributed noise with a variance of 2 minutes to trip end and start times that were multiples of 5 and, with a variance of 15 to recorded times that were multiples of 30 minutes.

The idea behind the test of the Poisson hypothesis was explained first in [35]. Assume that there are a total of N arrival events in an interval $(0, t)$, occurring at times $t_1, \ldots, t_N$. Then, if we condition on N, the variables t_i/t are independent and uniformly distributed between $(0, 1)$. Thus, to test the conditional uniformity of the arrival times, standard tests of uniformity can be applied to this data (for example, a Chi-squared test of uniformity) [36]. We used a one-sample Kolmogorov-Smirnov test to compare the distributions of the normalized arrival times with the $U(0, 1)$ distribution. The test was repeated for different hours of the day and the null hypothesis was never rejected. Fig. 10 shows a q-q plot for arrivals between 13:30 to 14:00 on Mondays. Quantile-quantile (q-q) plots are used to check, in a

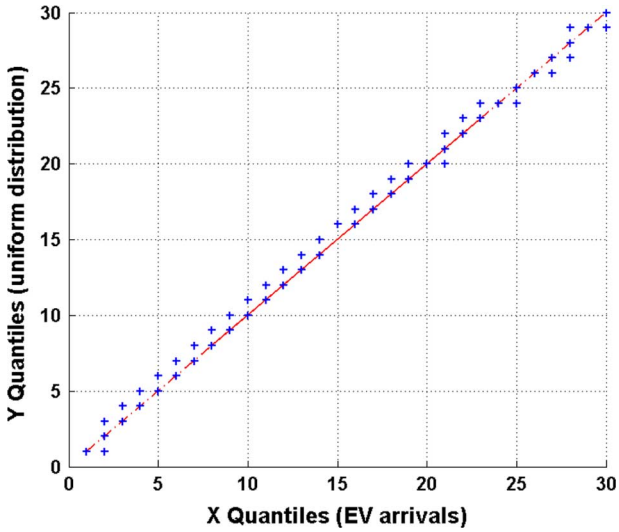

Fig. 10. q-q plot comparing conditionally normalized EV arrival times between 13:30–14:00 on a Monday to a uniform distribution.

non-parametric fashion, whether two sample set originated from the same probability distribution. If the two sets come from similar distributions, the points in the q-q plot will approximately lie on the line $y = x$. These tests have been run with standard Matlab functions. Obviously none of these tests can prove that the model is truthful, but they can at least not reject it.

F. Principal Components of Vehicle Arrival Counts

The NHTS gives us an ample amount of arrival count data for vehicle plug-ins, sorted by the day of the week. Unfortunately, since we do not know the exact date of each event, we cannot numerically demonstrate the principles described in Section III-A to calculate ML estimates of the coefficients of a time series model describing the $\alpha_j(i)$'s, e.g., the AR coefficient β_i and the noise variance σ_i^2 in (11). However, due the large size of the dataset, the principal components of the weekly arrival data give us a reliable estimate of the factors $\mathbf{u}_i, i = 1, \ldots, \kappa$ in (9). We perform an SVD on the transformed cumulative arrival count matrix $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_7]$, with each count vector $\mathbf{x}_t$ describing the arrival counts at different hours of a specific the week of the day t. It is interesting to mention that the six most important $\mathbf{u}_i$'s, shown in Fig. 11(a), account for 96% of the variance. Thus, the model in (9) reduces the size of the data (in our case from 96 to 6 elements per day) and make the next steps of predicting future arrival rates more tractable. One noteworthy observation is that $\mathbf{u}_1$ is very similar to the weekday travel patterns while $\mathbf{u}_2$ compensates for the different trends observed in weekend travel patterns. The next step is to estimate the periodic mean of the principal component coefficients, i.e., $\bar{\alpha}_t(1), \ldots, \bar{\alpha}_t(\kappa)$ for $t = 1, \ldots, 7$ and $\kappa = 6$. Fig. 11(b) displays this periodic mean for the first sixth principal component coefficients.

VI. Numerical Experiments Part II: Real-Time Load Forecasting

Our experiments in this section will showcase the advantages of sub-metering PHEVs separately through real-time smart meter data, just for forecasting purposes. We refer to the technique, presented in Sections II-C and IV-A, as the *OAI*

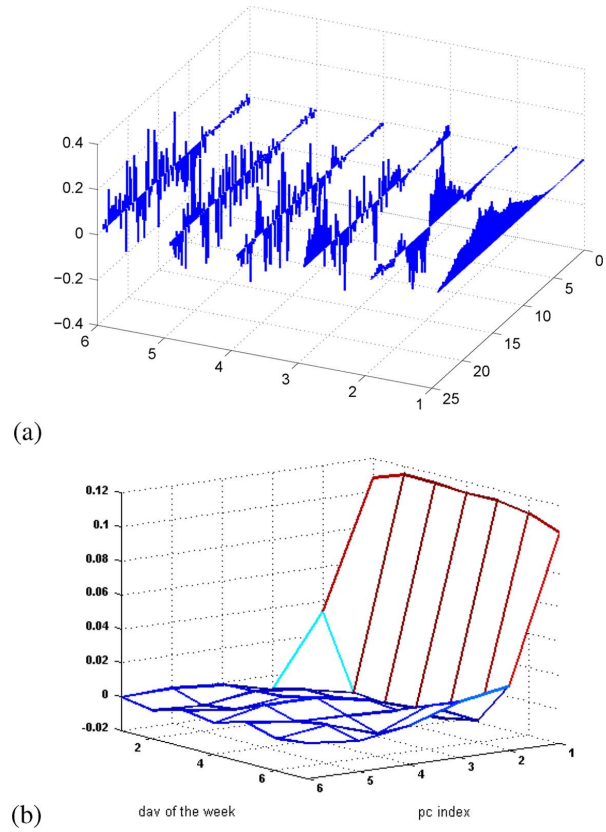

Fig. 11. (a) The first sixth principal components (index 1 on the x axis is the principal component with the highest variance). (b) Periodic mean of the six principal component coefficients (the x axis represents day of the week and the y axis is the principal component index).

predictor (prediction that uses Observable Arrival Information) and compare it to what we refer to as the *classical* prediction, which models the load as an ARMA process [37]. Univariate methods are frequently used for short-term load forecasting because exogenous variables such as the weather are assumed to change very smoothly over shorter time scales [38].

As explained in Section II-C, an OAI predictor needs to estimate the number of vehicles that will be receiving charge from each queue at a future time t. This load is proportional to the sum of a deterministic term $\nu(t)$ and a random term $N_{\text{new}}(t)$ with mean $m(t)$ in (2). To evaluate $m(t)$, we use two separate methods and compare their performances. In the first method, we forecast the arrival rate of the cars using the weekly component of the vehicle count vector and then evaluate the integral (2) numerically. We refer to this method as OAI1 predictor. In the second method, we use the results presented in [9], which we refer to as OAI2 predictor. This model assumes a homogeneous arrival process for the charging requests. Accordingly, we picked the average arrival rate we used for OAI1 as the arrival rate for OAI2, which was equal to 12 vehicles per every half hour.

The methods were tested on the simulated load for a substation serving a population of around 1000 PHEVs. The simulation covers 60 days, 30 of which are used for training the classical ARMA model and the forecast errors are averaged over the remaining 30 days. For fairness, we set the base load to be zero so that forecasting the base load does not contribute to the

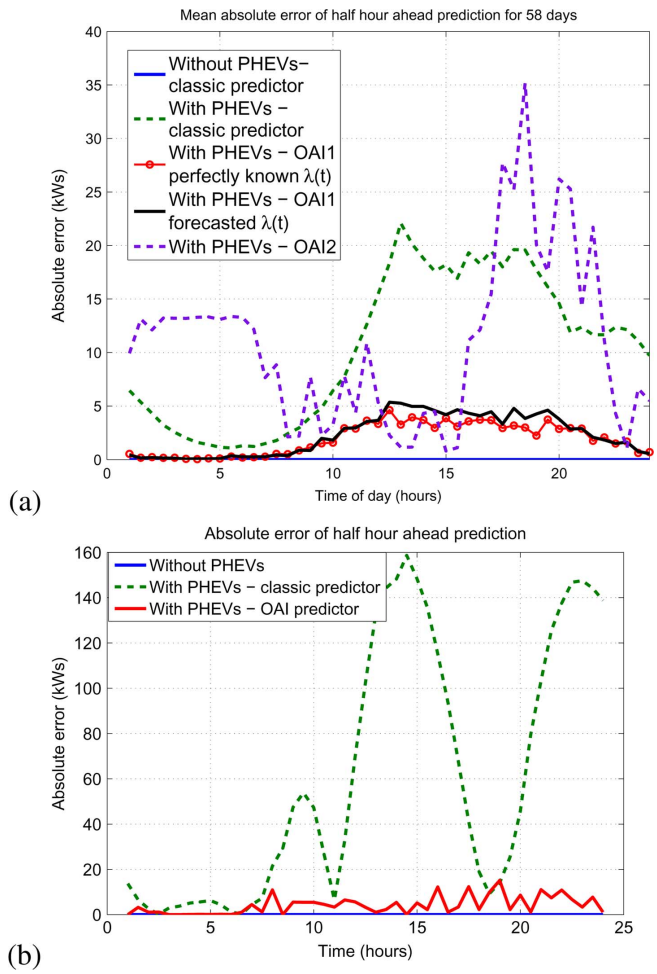

Fig. 12. Absolute error of half hour ahead prediction (a) arrivals based on NHTS data (b) arrivals based on UC Davis PH&EV center data.

error. We assume that the entire population of PHEVs uses level 1 charging (1.1 kW). We assume that: 1) the residents arrive to receive charge based on arrival rates derived from the NHTS data; 2) the charging time S of each car has the same distribution (20). Our results, displayed in Fig. 11(a), show the average absolute error (in kWs) of half-hour ahead prediction using the OAI1, OAI2, and the classical methods. In our simulations, the daily load due to battery charging of the 1000 PHEVs had a peak of around 250 kWs. One can clearly observe that the classical prediction technique has an absolute error that is up to 3–4 times worse than the OAI1 method. Also, the results clearly showcase that a non-homogeneous arrival model is necessary to capture EV charging load. Thus, a useful extension to the model proposed in [9] would be to incorporate non-homogeneous arrivals.

Furthermore, to showcase the benefits of individual energy requests being observable even if the statistics are incorrect, we simulate the performance of both predictors when the arrival events are taken from the PH&EV center data set, with all events assumed to happen on the same day. Due to the small size of this data, we refrain from training the ARMA model or predicting the arrival rates for the OAI predictor. Consequently, we used the same model developed on the NHTS data for the classic predictor and used a persistence predictor (i.e., $\hat{\lambda}(\ell+1) = \lambda(\ell)$) for the OAI predictor. This clearly degrades the performance of both predictors. However, as seen in Fig. 12, lowering the prediction accuracy of $\lambda(t)$, which is consistent with the scenario where the arrivals are more volatile and unpredictable, will hurt the classic predictor much more than the OAI predictor. This is simply due to the fact that, even if prediction of $\lambda(t)$ is inaccurate, the OAI predictor still has zero forecast error on the load due to previous charge requests.

VII. Conclusion

This paper provides a stochastic model useful to simulate and predict the electricity demand resulting from a general arrival and charging pattern for EVs/PHEVs. It also proposes a methodology to gather and fuse information online to perform future predictions that is useful both for the generation and demand side management of EVs. The model is validated using traces extracted from mapping NHTS data and via real PHEV measurements from the UC Davis PH&EV center. We would like to reiterate that the numerical values presented for the parameters of our model can and should be re-evaluated as new data (and in particular, multi-day charing data) become available, charging infrastructure develops, and business plans for demand side management/dynamic pricing techniques emerge.

References

[1] J. Axsen *et al.*, "Plug-in hybrid vehicle ghg impacts in california: Integrating consumer-informed recharge profiles with an electricity-dispatch model," *Energy Policy*, vol. 39, no. 3, pp. 1617–1629, 2011.

[2] R. Green, L. Wang, and M. Alam, "The impact of plug-in hybrid electric vehicles on distribution networks: A review and outlook," *Renewable Sustain. Energy Rev.*, vol. 15, no. 1, pp. 544–553, 2011.

[3] S. Schneider, R. Bearman, H. McDermott, X. Xu, S. Benner, and K. Huber, "An assessment of the price impacts of electric vehicles on the PJM market," A Joint Study by PJM and Better Place,, May 2011.

[4] N. DeForest *et al.*, "Impact of widespread electric vehicle adoption on the electrical utility business threats and opportunities," Center for Entrepreneurship and Technology (CET) Technical Brief,, Aug. 2009.

[5] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 371–380, Feb. 2010.

[6] G. Putrus, P. Suwanapingkarl, D. Johnston, E. Bentley, and M. Narayana, "Impact of electric vehicles on power distribution networks," in *Proc. IEEE VPPC*, Sep. 2009, pp. 827–831.

[7] D. Wu, D. Aliprantis, and K. Gkritza, "Electric energy and power consumption by light-duty plug-in electric vehicles," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 738–746, May 2011.

[8] J. Taylor, A. Maitra, M. Alexander, D. Brooks, and M. Duvall, "Evaluation of the impact of plug-in electric vehicle loading on distribution system operations," in *Proc. IEEE PES GM*, Jul. 2009, pp. 1–6.

[9] G. Li and X.-P. Zhang, "Modeling of plug-in hybrid electric vehicle charging demand in probabilistic power flow calculations," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 492–499, Mar. 2012.

[10] M. Alizadeh, A. Scaglione, and Z. Wang, "On the impact of smartgrid metering infrastructure on load forecasting," in *Proc. Allerton Conf.*, 2010.

[11] K. Kurani *et al.*, "Learning from consumers: Plug-in hybrid electric vehicle (PHEV) demonstration and consumer education, outreach, market research program, volumes I and II," Institute of Transportation Studies, UC Davis, Res. Rep. UCD-ITS-RR-10-21, 2010.

[12] Y. Xu and F. Pan, "Scheduling for charging plug-in hybrid electric vehicles," in *Proc. IEEE 51st Annu. Conf. Decision Control (CDC)*, Dec. 2012, pp. 2495–2501.

[13] S. Chen, Y. Ji, and L. Tong, "Deadline scheduling for large scale charging of electric vehicles with renewable energy," in *Proc. IEEE 7th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2012, pp. 13–16.

[14] K. Mets, T. Verschueren, W. Haerick, C. Develder, and F. De Turck, "Optimizing smart energy control strategies for plug-in hybrid electric vehicle charging," in *Proc. IEEE/IFIP Netw. Oper. Manage. Symp. Workshops (NOMS Wksps),* , Apr. 2010, pp. 293–299.

[15] S. Han, S. Han, and K. Sezaki, "Development of an optimal vehicle-to-grid aggregator for frequency regulation," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 65–72, Jun. 2010.
[16] M. Alizadeh, T.-H. Chang, and A. Scaglione, "On modeling and marketing the demand flexibility of deferrable loads at the wholesale level," in *Proc. HICCS 2013*, to be published.
[17] S. G. Eick, W. A. Massey, and W. Whitt, "The physics of the Mt/G/infty Queue," *Oper. Res.*, vol. 41, no. 4, pp. 731–742, 1993.
[18] C. Palm, *Intensity Variations in Telephone Traffic*. Amsterdam, The Netherlands: North Holland, 1988, book edition of Ph.D. thesis of 1943.
[19] A. Khinchin, *Mathematical Methods in the Theory of Queueing*. London, U.K.: Griffin, 1955.
[20] S. W. Hadley and A. A. Tsvetkova, "Potential impacts of plug-in hybrid electric vehicles on regional power generation," *Electricity J.* vol. 22, no. 10, pp. 56–68, 2009 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S104061900900267X
[21] J. Weinberg, L. D. Brown, and J. R. Stroud, "Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. to appear," *J. Amer. Stat. Assoc.*, 2007.
[22] D. Matteson, M. McLean, D. Woodard, and S. Henderson, "Forecasting emergency medical service call arrival rates," *Ann. Appl. Stat.*, vol. 5, no. 2B, pp. 1379–1406, 2011.
[23] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao, Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective Tech. Rep., Novemer.
[24] H. Shen and J. Z. Huang, "Interday forecasting and intraday updating of call center arrivals," *Manuf. Service Oper. Manage.*, vol. 10, pp. 391–410, Jul. 2008.
[25] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis, Forecasting and Control*, ser. Wiley Series in Probability and Statistics, 4th ed. Hoboken, NJ, USA: Wiley, 2008.
[26] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, 1948.
[27] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[28] G. Welch and G. Bishop, "An introduction to the Kalman filter," [Online]. Available: http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf 1995
[29] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Roy. Stat. Soc. B*, vol. 39, no. 1, pp. 1–38, 1977.
[30] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.
[31] M. Alizadeh, A. Scaglione, and R. Thomas, "From packet to power switching: Digital direct load scheduling," *IEEE J. Sel. Areas Commun. (Special Issue on Smart Grid Communications)*, vol. 30, no. 6, pp. 1027–1036, Jul. 2012.
[32] U.S. Dept. Transportation, Federal Highway Administration, "2009 national household travel survey" [Online]. Available: http://nhts.ornl.gov
[33] T. Turrentine, D. Garas, A. Lentz, and J. Woodjack, "The UC Davis MINI E consumer study," Inst. Transportation Studies, Univ. California, Davis, Res. Rep. UCD-ITS-RR-11-05, 2011.
[34] M. Kintner-Meyer, K. Schneider, and R. Pratt, "Impacts assessment of plug-in hybrid vehicles on electric utilities and regional U.S. power grids, part 1: Technical analysis,", 2007.
[35] G. A. Barnard, "Time intervals between accidents: A note on Maguire, Pearson and Wynn's paper," *Biometrika*, vol. 40, pp. 212–213, 1953.
[36] P. A. W. Lewis, "Some results on tests for Poisson processes," *Biometrika*, vol. 52, no. 1/2, pp. 67–77, 1965.
[37] G. Gross and F. Galiana, "Short-term load forecasting," *Proc. IEEE*, vol. 75, no. 12, pp. 1558–1573, Dec. 1987.
[38] J. W. Taylor, "Triple seasonal methods for short-term electricity demand forecasting," *Eur. J. Oper. Res.*, vol. 204, no. 1, pp. 139–152, 2010.

Mahnoosh Alizadeh (S'08) is a Ph.D. candidate at the University of California Davis, CA, USA. She received the B.Sc. degree in electrical engineering from Sharif University of Technology, Iran, in 2009. Her research interests mainly lie in the area of real-time control and optimization for cyber-physical systems, with a special focus on designing scalable and economic demand side management frameworks for the smart grid.

Anna Scaglione (S'97–M'99–SM'09–F'11) is currently Professor in Electrical and Computer Engineering at the University of California Davis, CA, USA. She joined UC Davis in 2008, after leaving Cornell University, Ithaca, NY, USA, where she started as Assistant Professor in 2001 and became Associate Professor in 2006; prior to joining Cornell she was Assistant Professor in the year 2000–2001, at the University of New Mexico.

Dr. Scaglione is the Editor in Chief of the IEEE SIGNAL PROCESSING LETTERS, and served as Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2002 to 2005, and from 2008 to 2011 in the Editorial Board of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, where she was Area Editor in 2010–2011. She has been in the Signal Processing for Communication Committee from 2004 to 2009 and is in the steering committee for the conference Smartgridcomm since 2010. She was general chair of the workshop SPAWC 2005. Dr. Scaglione is the first author of the paper that received the 2000 IEEE Signal Processing Transactions Best Paper Award; she has also received the NSF Career Award in 2002, and is a co-recipient of the Ellersick Best Paper Award (MILCOM 2005), and the 2013 IEEE Donald G. Fink Award. Her expertise is in the broad area of signal processing for communication systems and networks. Her current research focuses on signal processing algorithms for networks and for sensors systems, with specific focus on smart grid, demand side management, and reliable energy delivery.

Jamie Davies received the B.S. degree in environmental science and policy in 2008, and the M.S. degree in transportation in 2011, from the University of California at Davis, CA, USA. From 2008 to 2011 he was a Graduate Student Researcher at the Institute for Transportation Studies at UC Davis. Since 2011, he has been Consumer Research Analyst at the Plug-in and Electric Vehicle Research Center at UC Davis. His research emphasis is on the design and practice of quantitative and qualitative research to collect, measure and evaluate plug-in electric vehicle impacts, policy goals and vehicle design potential based on consumer's real world travel and charging behavior. His broader research goals also include understanding factors which affect PEV market and testing the effectiveness of market development strategies on the sale of plug-in electric vehicles to organizations and consumers.

Kenneth S. Kurani received the Ph.D. degree in civil and environmental engineering from the University of California, Davis, CA, USA, in 1992.

He is presently an Associate Researcher at the Institute of Transportation Studies at the University of California, Davis. He is the co-author of nearly 100 publications including scholarly articles, book chapters, and technical reports. His research focuses on the intersection of lifestyle, automobility, energy, and the environment. He is especially interested in enriching the behavioral approaches to consumers responses to new transportation technologies such as new propulsion systems and fuels for automobiles.

Dr. Kurani was a Federal Highway Administration Graduate Research Fellow in 1985 and a Chevron Fellow in 1991. He is a friend of the committee for the Energy, Alternative Transportation Fuels, Travel Behavior and Values, and User Information Systems Committees of the Transportation Research Board.