This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

Reinforcement Learning-Based Plug-in Electric Vehicle Charging with Forecasted Price

Adriana Chiş, *Student Member, IEEE,* Jarmo Lundén, *Member, IEEE,* and Visa Koivunen, *Fellow, IEEE*

***Abstract*—This paper proposes a novel demand response method that aims at reducing the long term cost of charging the battery of an individual plug-in electric vehicle (PEV). The problem is cast as a daily decision making problem for choosing the amount of energy to be charged in the PEV battery within a day. We model the problem as a Markov decision process with unknown transition probabilities. A batch reinforcement learning algorithm is proposed for learning an optimum cost reducing charging policy from a batch of transition samples and making cost reducing charging decisions in new situations. In order to capture the day-to-day differences of electricity charging costs the method makes use of actual electricity prices for the current day and predicted electricity prices for the following day. A Bayesian neural network is employed for predicting the electricity prices. For constructing the reinforcement learning training data set we use historical prices. A linear programming based method is developed for creating a data-set of optimal charging decisions. Different charging scenarios are simulated for each day of the historical time frame using the set of past electricity prices. Simulation results using real world pricing data demonstrate cost savings of 10%-50% for the PEV owner when using the proposed charging method.**

***Index Terms*—Cost reduction, demand response, plug-in electric vehicle, price prediction, reinforcement learning, smart charging.**

I. Introduction

THE transportation system is one of the biggest contributor to the diminishing oil resources and environmental problems such as increasing level of pollution. In order to address these problems, plug-in electric vehicles (PEVs) will soon become an important component of the transportation system. PEVs imply zero fossil fuel consumption and very low carbon emissions. The rapidly advancing PEV technologies are continuously providing larger PEV battery capacities which allow these vehicles to be driven over longer distances with one charge, without the need of daily charging. Uncontrolled PEV charging on the other hand, can lead to a significant increase of the residential electricity consumption cost for the PEV owner. In addition to that, PEV charging outside of home might lead to further increase in costs. In order to support their price-conscious customers, many utility companies offer time varying prices which give their customers the possibility of controlling their electricity usage. Smart PEV charging programs may take advantage of the electricity price variation and the flexibility of the load scheduling in order to shift the electricity usage to times when the prices of electricity and loads in the power grids are low. Moreover, these programs, also called demand response programs, do not require much effort from the user.

The authors are with the Department of Signal Processing and Acoustics, Aalto University, FI-02150 Espoo, Finland (e-mail: adriana.chis@aalto.fi, jarmo.lunden@aalto.fi, visa.koivunen@aalto.fi)

A variety of methods for reducing the costs of charging of a large number of electric vehicles have been proposed in the literature [1]–[6]. A decentralized price based technique for PEVs' charging coordination is presented in [1]. A distributed optimization model for integration of PEVs and incorporation of renewable energy within the power grid is proposed in [2]. In [3] a centralized approach is presented for scheduling the charging of a large number of PEVs. The approach considers jointly the aggregators' and customers' revenues.

Methods that employ a Markov decision process (MDP) formulation for optimization of electric vehicles' charging of have been investigated in [7]–[11]. In [7] a finite horizon MDP problem formulation is proposed to optimize the charging of an individual plug-in hybrid electric vehicle that can provide ancillary services to the grid. Day-ahead electricity prices are assumed to be known and hourly decisions are taken to minimize the 24-hour cost of charging. A finite horizon stochastic MDP is proposed in [8] to optimize the charging of a PEV with grid-to-vehicle frequency regulation possibility. A method for controlling the battery usage of a PEV that can act as an energy storage for the power grid is proposed in [9]. The stochastic driving patterns of the PEV are modeled through an inhomogeneous Markov model. The operational behavior of a charging station for plug-in hybrid electric vehicles is modeled as a MDP in [10]. The authors in [11] use trip chain and MDP to simulate the daily stochastic driving, parking and charging events of moving PEV loads among power system buses.

Reinforcement-learning (RL) techniques have been used for obtaining solutions to different demand response problems. A batch RL method was proposed in [12] to learn the day-ahead hourly charging behavior of a fleet of PEVs. Here the EV fleet aggregator participates to a day-ahead market bidding and wants to minimize its energy purchase risk. Online RL techniques were proposed in [13], [14] to perform demand response to daily schedule the residential loads.

In this paper, we propose a novel method that schedules the home charging of an individual PEV such that the long term electricity costs for the PEV owner are reduced. The method exploits the day-to-day fluctuations of electricity prices and the large PEV battery capacities common nowadays in order to make daily decisions upon the amount of energy to be charged. It is assumed that the PEV owner buys electricity from the utility company at a rate determined by the actual hourly market exchange price. The PEV owner's household is equipped with a smart device that coordinates the charging of the PEV's battery. The smart device receives the day-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

ahead electricity prices from the utility company to which it is connected through a communication system. The device can record historical electricity pricing and daily PEV consumption information. In this work the PEV's owner's driving patterns are assumed to be known. Using the available information the smart device computes an optimized charging strategy for reducing the electricity consumption cost, while respecting the constraints imposed by the driving patterns. Our main contributions in this paper are described below:

- We formulate the problem as a Markov decision process with unknown transition probabilities. A *fitted Q-iteration* batch reinforcement learning algorithm is proposed to learn an optimum cost reducing charging policy from a data-set of historical transition samples. The learned policy is then exploited to make charging decisions in new situations such that the long term cost of charging is reduced.
- The transition samples data-set used for training the *fitted Q-iteration* algorithm is constructed by exploring the day-to-day fluctuations of known electricity prices over multiple days in the past. Different charging scenarios are simulated for each day of the historical time frame. A linear programming (LP) optimization is performed to determine the optimum charging action in each scenario. The optimization uses the known plug-in and plug-out states and consumption patterns of the PEV for each day of the week.
- The method makes use of known day-ahead prices for the current day and predicted prices for the following day. A Bayesian neural network is employed to predict the electricity prices. A suitable set of input features and prior distributions for the network parameters are chosen in order to obtain accurate predictions of the electricity prices.
- A linear program is used to optimally schedule the PEV battery charging during the current day, after the charging decision is taken.

The existing demand response methods for individual PEV charging [7]–[9] optimize the charging over a 24-hour time frame. The goal of the proposed method is to extend the 24-hour optimization time frame and hourly decision making scenarios. In this paper, we make use of daily effective electricity prices and next-day predicted electricity prices. We explore the daily intertemporal fluctuations of electricity prices in order to make daily charging decisions that reduce the cost of charging over a long time horizon. For this purpose we propose a novel method that consists of using the MDP formulation combined with price prediction and a *fitted Q-iteration* based batch reinforcement learning method. The *fitted Q-iteration* uses a kernel-averaging regression function to determine the cost reducing charging solution of the problem. The PEV charging problem addressed in this work and its MDP formulation are different from the those considered in [10], [11]. We propose an MDP model based on daily time steps and a continuous state space. The control action is defined by a discretized value representing an amount of energy to be charged in a battery during each day such that the cost of charging over an infinite time horizon is minimized and daily driving constraints may be satisfied. An optimal scheduling of the battery charging within a day is found using linear programming.

Our preliminary ideas and results presented in [15], [16] are extended as follows:

- We propose a Bayesian neural network model to forecast the electricity prices employed in the PEV charging technique;
- We extend the state space of the MDP model by introducing new pricing variables that capture more accurately the fluctuations of electricity prices from one day to another;
- A heuristic technique was used in the preliminary work for determining the reward information involved in the reinforcement learning method. In this work we propose to use LP optimization both to construct the batch of transition samples and to schedule the charging during the current day;
- We provide thorough simulation results that demonstrate the performance of the proposed method.

The performances of the price prediction and the proposed PEV charging methods are tested using true electricity pricing data [17]. Our PEV charging method is compared in simulation with the daily optimum PEV charging method (daily deterministic charging method in which the exact consumption and electricity prices are known) and with the conventional unoptimized charging method. Simulation results show that the cost of electricity for charging the PEV can be reduced by 10% to 50% when using the method proposed in this paper.

The rest of the paper is organized as follows. Section II includes the system model that formulates the charging scenario as a MDP. The batch reinforcement learning method and the LP optimization method are presented in Section III. The Bayesian network for price prediction is introduced in section IV. The numerical results that show the performance of the price prediction method and the PEV charging scheme are presented in section V. Section VI gives the conclusions.

II. Markov Decision Process Model

The goal of the proposed PEV charging strategy is to daily choose the amount of energy to be charged such that the long term cost of charging is reduced. Since the actual market electricity prices are known only for a day ahead, this problem is formulated as an MDP with unknown transition probabilities. An MDP is a mathematical formulation for modeling decision making in uncertain situations [18]. An MDP is defined by a sequence of discrete time steps *n=1,…,N*, a set of possible states $\mathcal{X}$ that describe the environment and a set of possible actions $\mathcal{U}$ for each of the states. An agent takes actions according to a policy μ and a state transition function $\rho : \mathcal{X} \times \mathcal{U} \times \mathcal{X} \rightarrow [0, 1]$ defines the state transition probabilities. A real valued reward function r: $\mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R}$ gives rewards for each action taken.

A. System model

The PEV charging problem is described as an MDP defined on daily time steps. We denote by *n* the index of the current day for which the charging strategy needs to take an action. At the beginning of day *n* the utility company gives the set of true hourly market electricity prices for 24 hours: $\mathbf{p}_n = \{p_n^{(t)}\}_{t=1}^{24}$. The set of hourly prices for the following day *n*+1 are predicted using a Bayesian price forecasting method. This set of prices is denoted by: $\hat{\mathbf{p}}_{n+1} = \{\hat{p}_{n+1}^{(t)}\}_{t=1}^{24}$. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

battery capacity is limited by a maximum value $C_{\max}$. The battery charger has an hourly maximum charging rate which we denote by c_{rate}. The knowledge of the PEV's owner's driving patterns is a prerequisite to the proposed PEV charging method. We assume that the smart charging device can record the daily plug-in and plug-out times of the vehicle as well as the energy consumed during each trip. This data can be used to create statistical models of the driver's daily charging patterns. Methods for construction of such models are proposed in [19], [20]. In this paper we assume that the daily consumption as well as plug-in and plug-out times of the PEV are already known. We denote by ϵ_n the consumption parameter indicating the estimated amount of energy that the vehicle is expected to consume during the current day n. Let *t* be the index of an hour within a day. The sets $\mathbf{h}_n = \{t|\ PEV\ is\ at\ home\}_{t\in\{1,\dots,24\}}$ and $\mathbf{h}_{n+1} = \{t|\ PEV\ is\ at\ home\}_{t\in\{1,\dots,24\}}$ show the hours of the day during which the PEV is at home during the current day *n* and during the next day *n*+1. Finally, we denote by $\zeta_n(\epsilon_n)$ and $\hat{\zeta}_{n+1}(\epsilon_n)$ the optimum costs of charging the amount of energy ϵ_n during the current day *n* and during the next day *n*+1. The optimum cost is computed using the LP optimization method described by the equations (5)-(8) in section III.B. The optimization is performed over one day time horizon using the actual market electricity prices $\mathbf{p}_n$ and the predicted electricity prices $\hat{\mathbf{p}}_{n+1}$.

B. MDP state space

The state space of the proposed MDP formulation allows for continuous variables. Let $\mathbf{x}_n$ be the state of the MDP corresponding to the current day *n* for which a charging decision must be taken. This state is represented by a vector of four variables: $\mathbf{x}_n = [d_n\ b_n^{(0)}\ p_{n_{\min}}\ \Delta_n]$. The variable $d_n \in \{1,\dots,7\}$ indicates the current day of the week. The variable $b_n^{(0)}$ shows the state of charge of the PEV battery at the beginning of the current day, before the charging schedule is computed: $0 \le b_n^{(0)} \le C_{\max}$. The minimum hourly electricity price of the current day is indicated by $p_{n_{\min}} = \min_{t\in\mathbf{h}_n} p_n^{(t)}$. For recording the daily fluctuation of the charging cost we introduce a fourth state variable component $\Delta_n = \zeta_n(\epsilon_n) - \hat{\zeta}_{n+1}(\epsilon_n)$. This variable acts as a cost fluctuation indicator. The variable shows the difference in the cost of charging the ϵ_n amount of energy during the current day *n* and the cost for charging the same amount of energy ϵ_n during the next day *n*+1.

C. MDP action space

The action of the proposed PEV charging strategy is defined by the amount of energy that is charged in the battery during the current day *n*. The action should reflect the day-to-day fluctuations of electricity prices as follows: the algorithm can choose to charge a large amount of energy if the electricity prices of the current day are low compared to the historical electricity patterns and the predicted prices for the following day. Moreover, it can even choose not to charge at all if the electricity prices of the current day are high and the state of charge of the battery fulfills the consumption constraint of the day. The action must always satisfy the consumption constraints.

In case of the proposed MDP model the action must be chosen from a finite set of possible actions. Thus, the action space of the MDP defining the PEV battery charging problem is discrete. The battery capacity is discretized into *L* equally sized levels. An action is denoted by u_n and is defined by an operation that controls the charging of the battery. The control operation chooses a number of battery levels to be charged during the current day: $u_n \in \mathcal{U} = \{0, ..., L+1\}$.

III. Batch Mode Reinforcement Learning

A. Principle of batch reinforcement learning

Let $\mu(n, \mathbf{x}_n) : \mathcal{X} \to \mathcal{U}$ be a policy that determines an action for every day *n* of the PEV charging time horizon, when the system is in the state $\mathbf{x}_n$. Choosing the optimum action for any given MDP state implies finding the optimal control policy μ_{opt} that generally requires knowledge of the state to state transition probability distribution. Since the transition probability distribution is unknown in our PEV charging setting, we propose a batch reinforcement learning method to learn the optimal control policy from a given set of transition samples. The goal of batch reinforcement learning is to find the policy that maximizes the rewards received by the agent when taking actions in the current environment. The overall expected discounted reward over a temporal horizon of *N* days, conditioned on an initial state $\mathbf{x}_0$, when the system is controlled by an arbitrary policy μ_N is:

$$J_{\mu_N}(\mathbf{x}_0) = \mathbb{E}\left\{\sum_{n=1}^{N} \gamma^n r_n | \mathbf{x}_0\right\}, \tag{1}$$

where γ $(0 \le \gamma \le 1)$ is the discount factor. Define $Q_{\text{opt}}(\mathbf{x}_n, u)$ as the optimal action-value function associated with the overall reward received by taking an action *u* when the system is in state $\mathbf{x}_n$: $Q_{\text{opt}}(\mathbf{x}_n, u) = \max_\mu \mathbb{E}\left\{\sum_{n=1}^{N} \gamma^n r_n | \mathbf{x}_0\right\}$. Then, the optimum action in each state is the one that maximizes the action-value function, also called the *Q*-function :

$$u_{n_{\text{opt}}} = \arg\max_{u\in\mathcal{U}} Q_{\text{opt}}(\mathbf{x}_n, u). \tag{2}$$

The optimum action-value function is difficult to determine when the state transition probability distribution is not known. We propose a reinforcement learning method based on a *fitted Q-iteration* algorithm that uses a kernel averaging regression function in order to approximate the action-value function: $\hat{Q}(\mathbf{x}_n, u)$. Batch reinforcement learning derives the optimal control policy from a set of four tuples $(\mathbf{x}_l, u_l, \mathbf{y}_l, r_l)$ sampled from the environment. The learning implies observing each state $\mathbf{x}_l$, the action u_l taken when the system is in state $\mathbf{x}_l$ and its associated reward r_l. By $\mathbf{y}_l$ we denote the next state that the system reaches by taking action u_l when the system was in state $\mathbf{x}_l$.

Fitted Q-iteration algorithm is an offline model free learning method in which the transition dynamics of the MDP problem are learned from an available batch of transition samples:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

$$\mathcal{F} = \{(\mathbf{x}_l, u_l, \mathbf{y}_l, r_l) | l = 1, ..., |\mathcal{F}|\}. \quad (3)$$

In this section by *l* we denote the index of the sample from the set $\mathcal{F}$, while by *n* we denote the daily MDP step. $|\mathcal{F}|$ denotes the cardinality of this set.

In the next subsection we explain how we sample from the environment in order to construct the batch of transition samples $\mathcal{F}$ needed in the PEV charging problem.

B. Sampling from environment using linear programming

The proposed PEV charging strategy needs to choose an action that determines the amount of energy to be charged during the current day *n*. The action should reduce the charging cost for the PEV owner by predicting the fluctuation of electricity prices of the current day in comparison to the possible price fluctuations during the following days. This operation cannot be performed due to the lack of knowledge of electricity prices for multiple days ahead. Instead, we evaluate different possible charging experiences using the available set of past prices and determine the best action.

We consider a time interval of λ days in the past for which the true hourly electricity prices are known. Denote by $\imath = n - \lambda, \ldots, n-1$ the day index within the given interval of λ days. For each day $\imath$ within the interval λ we determine an optimum amount of energy to be charged such that the consumption cost is minimized according to the price fluctuation over δ days ahead. The optimal amount is determined by employing a linear programming (LP) optimization method. The LP optimization is repeatedly performed over each finite time interval of δ days, $\delta < \lambda$, starting with each day $\imath$. This time interval is behaving as a sliding window within the time interval of λ days.

The optimization is performed by taking into account the known hourly electricity prices over the time interval of δ days as well as the current known plug-in, plug-out and consumption patterns of the PEV. $\epsilon_j, j = \imath, \ldots, \imath + \delta$, is the estimated daily consumption of the car for each day of the time frame of δ days. Then, let $\boldsymbol{\varepsilon}_j = \{\varepsilon_j^{(t)}\}_{t \in \mathbf{h}_j}$ be a consumption indicator vector such that:

$$\varepsilon_j^{(t)} = \begin{cases} \epsilon_j, & \text{if } t \in \mathbf{h}_j \text{ is the hour when the vehicle is plugged} \\ & \text{-in after the trip for } j = \imath, \ldots, \imath + \delta; \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Let $b^{(0)}$ be the initial state of charge of the battery at the beginning of day $\imath$. $b^{(0)}$ is a continuous value within the state space of the battery $[0, C_{\max}]$. Then, let $\eta^{(t)}$ be a continuous variable describing the amount of energy to be charged during hour *t* and let $b^{(t)}$ be the state of charge of the battery at the end of hour *t*. The objective of the optimization is to minimize the charging cost over the time frame of δ days ahead subject to the hourly charging rate constraint, battery capacity constraint and battery dynamics constraint. More formally the optimization problem may be stated as follows:

$$\min \sum_{j=\imath}^{\imath+\delta} \sum_{t \in \mathbf{h}_j} p_j^{(t)} \eta_j^{(t)}, \quad (5)$$

with constraint over the hourly charging rate:

$$0 \leq \eta_j^{(t)} \leq c_{\text{rate}}, \quad (6)$$

and battery capacity constraint:

$$0 \leq b_j^{(t)} \leq C_{\max}, \quad (7)$$

as well as battery dynamics constraint:

$$b_j^{(t)} = b_j^{(t-1)} + \eta_j^{(t)} - \varepsilon_j^{(t)}, \quad (8)$$

for all $t \in \mathbf{h}_j$, $j = \imath, \ldots, \imath + \delta$, where $b_j^{(0)} = b_{j-1}^{(24)}$ for $j = \imath + 1, \ldots, \imath + \delta$.

It can be observed that the objective function and the constrains of the optimization possess linear relationships between the variables of the problem. This makes the optimization problem to be of LP type.

In order to construct the training set $\mathcal{F}$ we compute the price variables $p_{n_{\min}}$ and Δ_n corresponding to the days $\imath = n - \lambda, \ldots, n-1$. For each day $\imath$ we solve the LP optimization problem in (5)-(8) and simulate different possible initial battery charge states $b^{(0)}$ by randomly taking several values sampled from the battery state space. Given the initial state of charge of the battery, the optimum charging amount for day $\imath$ is the cumulated hourly amounts of energy to be charged in the battery during the hours of the day $\imath$: $u^* = \sum_{t \in \mathbf{h}_\imath} \eta^{(t)}$. The *fitted Q-iteration* based reinforcement learning method proposed to solve the PEV charging method determines discrete charging actions. Consequently, the optimum continuous action (amount to charge) determined by the LP optimization must be quantized to a discrete action from the set of possible actions $\mathcal{U}$ defined for the PEV charging problem. In order to make sure that the consumption constraints are always fulfilled, the optimum discrete action is chosen to be the nearest action from $\mathcal{U}$, larger than the determined continuous optimum action: $u_{\text{opt}} = \lceil u^* \rceil \in \mathcal{U}$. The minimized cost of charging the optimum amount during day $\imath$: $\zeta(u^*) = \sum_{t \in \mathbf{h}_\imath} p_\imath^{(t)} \eta_\imath^{(t)}$. The reward for taking the optimal action u_{opt} is then: $r_{\text{opt}} = \zeta(u^*)$. For any other action the corresponding reward is the minimum price plus a penalty value $r = \zeta(u^*) + \zeta_{\text{penalty}}$.

C. Fitted Q-iteration algorithm with kernel-based approximation of the action-value iteration

The PEV's owner's driving patterns may be different during each day of the week. Due to this fact, the action-value function is approximated separately for each day of the week: $\hat{Q}_d(\mathbf{x}_n, u)$. In order to approximate the action-value function $\hat{Q}_d(\mathbf{x}_n, u)$, the given set of transition samples $\mathcal{F}$ is divided into equally sized subsets according to each day of the week *d* and according to each possible action *u*: $S_{d,u} = \{(\mathbf{x}_l, r_l, \mathbf{y}_l) | l = 1, ..., m\}$. Each subset $S_{d,u}$ represents the collection of *m* state transition samples in which action *u* was taken. The *fitted-Q iteration* algorithm for approximating the action-value function $\hat{Q}_d(\mathbf{x}_n, u)$ is described in Table I. We propose a kernel averaging regression operator to fit the $\hat{Q}$ action-value function to the data. The kernel averaging operator is defined as:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

Table I
FITTED Q-ITERATION ALGORITHM WITH KERNEL APPROXIMATION OF THE ACTION-VALUE

1:	Input: Set of samples $S_{d,u} = \{(\mathbf{x}_l, r_l, \mathbf{y}_l) \vert l = 1, ..., m\}$, discount factor γ;
2:	Initialize: Q function approximate $\hat{Q}_{d,0}(\mathbf{y}'_l, u) = 0$;
3:	**repeat** at every iteration $\tau = 1, 2, \ldots$
4:	**for** $\mathbf{y}'_l \in S_{d,u}, l = 1, \ldots, m$ **do**
5:	$\hat{Q}_d^{(\tau+1)}(\mathbf{y}'_l, u) = \sum_{(\mathbf{x}_l, \mathbf{r}_l) \in S_{d,u}} \kappa(\mathbf{x}_l, \mathbf{y}'_l)[r_l + \gamma \max_u \hat{Q}_d^{(\tau)}(\mathbf{y}'_l, u)]$;
6:	**end for**
7:	**until** $\| \sum_{l=1}^{m} \sum_{u \in \mathcal{U}} (\hat{Q}_d^{(\tau+1)}(\mathbf{y}'_l, u) - \hat{Q}_d^{(\tau)}(\mathbf{y}'_l, u)) \| \simeq 0$;
8:	Output: $\hat{Q}_d(\mathbf{y}', u)$;

$$\kappa(\mathbf{x}_l, \mathbf{x}_n) = \frac{\phi\left(\frac{\|\mathbf{x}_l - \mathbf{x}_n\|}{\beta}\right)}{\sum_{\mathbf{x}_k \in \mathcal{F}} \phi\left(\frac{\|\mathbf{x}_k - \mathbf{x}_n\|}{\beta}\right)}, \tag{9}$$

where β is a *bandwidth* parameter that controls the smoothness of the kernel function and $\| \cdot \|$ can be any suitable norm. $\kappa(\mathbf{x}_l, \mathbf{x}_n)$ acts as a weighting kernel function that depends on the training set $\mathcal{F}$, the state $\mathbf{x}_n$ and a kernel function ϕ.

A kernel-based approximation of the value iteration was introduced in [21]. Given an initial approximation value of the action-value function $\hat{Q}_d(\mathbf{x}_n, u)$, each iteration τ of the method consists of solving the exact Bellman-equation [22].

The learning stage is completed when the difference in values of $\hat{Q}_d(\mathbf{y}'_l, u)$ between two consecutive iterations becomes sufficiently small. The convergence of the method was proven in [21]. When the learning stage is over, the value of $\hat{Q}_d(\mathbf{x}_{\text{new}}, u)$ in the new PEV charging setting for the current day of the week *d* described by a state $\mathbf{x}_n = \mathbf{x}_{\text{new}}$ can be obtained from:

$$\hat{Q}_d(\mathbf{x}_{\text{new}}, u) = \sum_{(\mathbf{x}_l, \mathbf{r}_l) \in S_{d,u}} \kappa(\mathbf{x}_l, \mathbf{x}_{\text{new}})[r_l + \gamma \max \hat{Q}_d(\mathbf{y}_l, u)]. \tag{10}$$

The optimal action to be taken for the new state $\mathbf{x}_{\text{new}}$ is the one that satisfies (2). This optimal action represents the amount of energy to be charged in the battery during the current day *n* represented by the state $\mathbf{x}_n = \mathbf{x}_{\text{new}}$, as described in section II.B of this paper. The scheduling of the PEV battery charging during the hours of the current day *n* will be done optimally based on the LP method described in (5-8) using the known electricity prices $\mathbf{p}_n$ and consumption constraint ϵ_n.

IV. PRICE PREDICTION USING BAYESIAN NETWORKS

Electricity price forecasting plays an important role in the development and operation of smart electric power grids. Reduced electricity costs will give incentives for the electricity residential users to participate in demand response programs. Accurate price forecasts will determine also the activity of utility companies and electricity suppliers in the electricity market. In recent years a wide variety of methods and algorithms have been developed for electricity price forecasting. The autoregressive integrated moving average (ARIMA) [23] is a time series approach for predicting electricity prices. A neural-network based prediction method [24] was employed for modeling the nonlinearities of electricity prices. Recently, probabilistic interval forecasting methods [25] were studied in order to capture the heteroskedasticity of electricity price forecasting errors.

In this work a multi-layered perceptron (MLP) neural network with a single hidden layer is employed to forecast the electricity prices. Artificial neural networks such as the MLP possess advantages that allow them to efficiently perform nonlinear statistical modeling. Some of the key advantages include: less rigorous statistical training demand, capability of capturing nonlinear relationships between the predictor and output variables and capability of identifying interactions between the predictor variables. A Bayesian approach is proposed to learn the network parameters. The Bayesian learning provides a natural way to model the nonlinear structure of electricity prices due to its capability to cope with the model complexity. Given a prior distribution of the unknown parameters and a set of observations the goal of this approach is to provide reliable estimation of the parameter values. Moreover, the Bayesian neural network together with the hierarchical model presented allows for the method to incorporate more data, thus obtaining accurate predictions of the electricity prices.

A Bayesian approach for predicting 24-hour electricity prices was previously introduced in [26]. The authors demonstrate through simulations that the presented Bayesian method produces good results. However, the method proposed in [26] is lacking in many aspects. Only a vague, incomplete description of the prior assumptions is provided. In order to predict electricity prices for 24 hours the method requires that the electricity load is also known in advance for the entire 24-hour time frame. The method requires re-training of the network for each 24-hour price prediction. The Bayesian method proposed in this work for forecasting the electricity prices brings significant gains. In the following paragraphs we give an explicit description of the prior assumptions and parameters of these assumptions required by the Bayesian approach. We provide a different noise model suitable in case of heteroskedastic regression problems such as electricity price forecasting. A set of input features based only on known past pricing and load data is proposed for deriving the Bayesian inference for prediction. The method estimates posterior prediction distributions that can be used to predict electricity prices with good accuracy for multiple days, without the need of retraining the network. The Bayesian model for electricity price prediction used in this work may be described as follows.

Let $\mathcal{D} = \{(\mathbf{v}_i, z_i)\}_{i=1}^{S_\mathcal{D}}$ be the set of training data. Here $\mathbf{v}_i$ for $i = 1, \ldots, S_\mathcal{D}$ are the input vectors composed by a set of *M* features containing relevant information for deriving the Bayesian inference for prediction and z_i for $i = 1, \ldots, S_\mathcal{D}$ are the target samples. $S_\mathcal{D}$ represents the total number of samples in the training set $\mathcal{D}$. Denote by $\boldsymbol{\theta}$ the set of all network parameters: $\boldsymbol{\theta}$=$\{\mathbf{b}_1, \mathbf{w}_1, \mathbf{b}_2, \mathbf{w}_2\}$, where $\mathbf{w}_k, \mathbf{b}_k$, for $k = 1, 2$, represent the vectors of weights and biases of the output and hidden layers of the neural network. The principle of Bayesian learning for MLP networks is to construct the poste-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

rior distributions for the network parameters $\boldsymbol{\theta}$ given the data samples $\mathcal{D}$ and an explicit definition of the prior probability distributions of the parameters. In order to describe these prior probabilities a hierarchical prior structure is constructed. A complete description of the functionality of hierarchical prior structure is described in [27]. In this structure the parameters of the prior probability distributions for the weights and biases of the neural network are called hyperparameters. The hyperparameters have also their own prior distributions with fixed parameters. The prior distribution of the networks' weights and biases is the Gaussian distribution:

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \sigma_{w_k}^2 \mathbf{I}), \quad \mathbf{b}_k \sim \mathcal{N}(\mathbf{0}, \sigma_{b_k}^2 \mathbf{I}), \text{ for } k = 1, 2, \quad (11)$$

where σ_w^2, σ_b^2 are the variance hyperparameters for the distributions of weights and biases, respectively. $\mathbf{I}$ is an identity matrix whose size is equal to the number of neuron units in each layer. Let $\sigma^2 = \sigma_w^2 = \sigma_b^2$ be a common notation for the variance hyperparameters. The hyperparameters σ^2 are described using a conjugate inverse-gamma prior distribution:

$$\sigma^2 \sim \text{inv-gamma}(\alpha^2, \nu_\sigma) \propto (\alpha^2)^{-(\frac{\nu_\sigma}{2}+1)} e^{-\frac{\nu_\sigma \alpha^2}{2\sigma^2}}, \quad (12)$$

where α is the scale parameter and ν_σ represents the number of degrees of freedom. A conjugate type of prior is designated in the case when the posterior distribution is desired to belong to the same family as the prior probability distribution. The inverse-gamma distribution is used as a noninformative prior distribution. Noninformative priors indicate that limited or no information is known in advance about the value of the parameter.

The posterior distribution of the MLP network's parameters $\boldsymbol{\theta}$ and hyperparameters σ given the data $\mathcal{D}$ is derived according to Bayes' rule as:

$$P(\boldsymbol{\theta}, \sigma|\mathcal{D}) = \frac{P(\mathcal{D}|\boldsymbol{\theta}) P(\boldsymbol{\theta}|\sigma) P(\sigma)}{P(\mathcal{D})}, \quad (13)$$

where $P(\mathcal{D}|\boldsymbol{\theta})$ represents the likelihood of the network's parameters, $P(\boldsymbol{\theta}|\sigma)$ and $P(\sigma)$ represent the prior distribution of the network's parameters and hyperparameters, respectively. $P(\mathcal{D})$ is the normalization term. The posterior predictive distribution of the output z_{new} for a new unseen input $\mathbf{v}_{\text{new}}$ given the training data set $\mathcal{D}$ is obtained by integrating the predictions of the network with respect to the posterior distribution of the network's parameters:

$$P(z_{\text{new}}|\mathbf{v}_{\text{new}}, \mathcal{D}) = \int P(z_{\text{new}}|\mathbf{v}_{\text{new}}, \boldsymbol{\theta}, \sigma) P(\boldsymbol{\theta}, \sigma|\mathcal{D}) d\boldsymbol{\theta} d\sigma. \quad (14)$$

Due to the functional form of the model, the integral in (14) does not have a closed form solution. One method for approximating the integration in neural networks is Markov Chain Monte Carlo (MCMC) sampling. In this paper we use the MCMC framework proposed in [27]. It uses a Markov chain to draw samples from the joint distribution of all network's parameters and hyperparameters. In the used framework the hybrid Monte Carlo (HMC) algorithm [28] is used for sampling the network's parameters and Gibbs sampling is used to derive samples for the hyperparameters.

In order to accurately model the heteroskedasticity of the electricity price forecasting error a probability model with additive error must be considered:

$$z = f(\mathbf{v}, \boldsymbol{\theta}) + e, \quad (15)$$

where $f(\cdot)$ is the output function of the MLP network:

$$f(\mathbf{v}, \boldsymbol{\theta}) = \mathbf{b}_2 + \mathbf{w}_2 \tanh(\mathbf{b}_1 + \mathbf{w}_1 \mathbf{v}). \quad (16)$$

In heteroskedastic regression each input-output sample $(\mathbf{v}, z) \in \mathcal{D}$ may have a specific noise variance σ_e, but all noise variances are defined by a single prior distribution model. In our electricity price forecasting problem the noise is assumed to obey Student's *t*-distribution due to the nonstationary nature of the electricity price time series, which often has sharp peaks and valleys:

$$e \sim t_\nu(0, \sigma_e^2). \quad (17)$$

The prior for the noise hyperparameter σ_e^2 is again the inverse-gamma distribution (12). The number of degrees of freedom ν of Student's *t* distribution is selected from a set of discrete values of ν by fitting the model to the data. The integration over the degrees of freedom is done through Gibbs sampling.

V. Performance Evaluation

In this section we first show the performance of the Bayesian neural network for forecasting the price of electricity. We then evaluate the performance of the proposed PEV charging method using simulations. In our examples we use real world pricing data.

A. Price Prediction using Bayesian Networks

The proposed PEV charging method requires accurately predicted electricity prices. A number of *M*=9 features containing relevant information for deriving the predictive Bayesian inference were chosen to compose the input vectors $\mathbf{v} = \{v_1, v_2, \ldots, v_9\}$ that are fed into the MLP network. These features are:

- v_1: a value from {1, …,7} indicating the day of the week for which the electricity price is being predicted;
- v_2: a flag {0,1} indicating whether the day of the week is a working day or weekend;
- v_3: a value from {1, …,24} showing the corresponding hour of the day;
- v_4, v_5: the hourly local marginal electricity price from one day before and from the same day one week before;
- v_6, v_7: the hourly system load from one day before and from the same day previous week;
- v_8, v_9: the average hourly local marginal electricity price and the average hourly system load from the day before.

The configuration of the MLP network contains 9 neurons (same as number of input features *M*) in the input layer, 10 neurons in the hidden layer and 1 neuron unit in the output layer. We choose the number of degrees of freedom in the description of the inverse-gamma prior (12) for all the hyper parameters σ as ν_σ=0.5. The scale parameter α instead depends on the distribution. In case of the Gaussian

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

prior for the network's parameters we choose the scale factors as: $\alpha_{w_1} = \alpha_{b_1} = 0.05$, $\alpha_{w_2} = \alpha_{b_2} = 0.05/U_2^{1/\nu_\sigma}$. Here the quantity U_2 represents the number of units in the hidden layer which in our model was chosen as 10. In case of Student's *t*-distribution for the noise model we choose the scale factor as $\alpha_e = 0.1$. The choice of number of degrees of freedom in case of the Student's *t*-distribution for the noise model is done by fitting the noise model to the data. In this sense Gibbs sampling is performed for a set of discrete values $\nu \in \{2, 2.3, 2.6, 3, 3.5, 4, 4.5, 5, 6, \dots 50\}$. The training of the MLP network was performed for 3000 iterations. Bayesian inference for prediction was performed using the set of samples composed by every 20th sample from the last 2000 iterations. Each iteration implied a trajectory length of 100. The training of the Bayesian MLP was carried out in Matlab using the MCMC toolbox [29]. For more information about the parameters of the model consult the documentation of the software and [27].

In order to evaluate the performance of the Bayesian network for price prediction the collection of input vectors, for learning and testing purposes, was taken from the ISO New England database [17]. Hourly data samples corresponding to 672 days starting from 24th of August 2009 until 26 of June 2011 were first taken for training purposes. In total a number of $S_{\mathcal{D}}$=16182 input vectors were fed into the input of the Bayesian network for learning. Due to the large training dataset the training stage of the MLP network takes several hours in Matlab on a conventional desktop computer. The training can be done offline, hence the complexity of the method is not a critical issue. The major advantage of the method is that electricity price can be forecasted even for 168 days using the posterior predictive distribution estimated after one training round.

In order to implement and evaluate the performance of the proposed PEV charging method electricity prices were forecasted for a total of 608 days from July 2011 until February 2013. This implied four training rounds of the MLP network. Numerical results showing the forecasting accuracy are presented in Table II. The performance of the Bayesian forecasting method is studied using the mean absolute forecasting error (MAE) and the mean absolute percentage forecasting error (MAPE). The performance of the Bayesian network (BN) for electricity price forecasting is compared against the forecasting performance of the ARIMA [30] method. The ARIMA method was used to forecast the day-ahead prices by fitting the actual prices from previous week. We also simulated the performance of the Bayesian model proposed in [26], for which we use the acronym VJ-BN. Note however, that some numerical values for the simulation were not specified in [26]. Consequently, the simulation for obtaining the VJ-BN price forecasts was performed using the same numerical values as in our proposed method. Table II shows the quantitative error criteria for forecasting the electricity prices of October 2012. The first 7 columns of the table show the daily forecasting accuracy for the first 7 days of October 2012. Then, the remaining column shows the forecasting error for the whole month. It can be observed that the developed Bayesian forecasting method improves the MAE performance by 1.16 when compared to the ARIMA method and by 0.42 when compared to the VJ-BN method. Also the MAPE values show an improvement of 3.34% and 0.94% respectively, when comparing our model to the ARIMA and VJ-BN methods. Fig. 1 shows the sequence of true electricity prices and the corresponding predicted electricity prices obtained by the BN method for first eleven days of October 2012.

Table II
OBTAINED MAE AND MAPE IN FORECASTING

		1st	2nd	3rd	4th	5th	6th	7th	Oct
MAE	**BN**	**2.95**	**3.65**	**4.54**	**1.74**	**1.75**	**1.51**	**2.4**	**3.27**
($)	ARIMA	3.04	9.44	8.5	3.28	3.17	3.03	2.36	4.43
	VJ-BN	2.07	3.46	5.95	2.62	2.44	3.44	3.23	3.69
MAPE	**BN**	**11.6**	**11.6**	**11.1**	**5**	**5.31**	**4.77**	**7.88**	**9.21**
(%)	ARIMA	13.49	29	20.56	8.71	9.92	9.89	7.35	12.55
	VJ-BN	8.97	10.99	14.53	7.58	7.47	10.59	10.11	10.15

B. The PEV Battery Charging

In order to evaluate the performance of the proposed PEV battery charging method we assume a PEV model having battery capacity $C_{\max}$=24kWh. The distance range of the PEV battery is considered to be 160km, while the battery charging rate is c_{rate}=6.6 kW. For simulating the daily consumption of the car the assumed driving distances in kilometers for each day of the week from Monday to Sunday are $\{40, 25, 35, 27, 25, 12, 30\}$. The hour intervals when the car is assumed to be away from home from Monday to Sunday are $\{8-16, 8-16, 8-18, 8-17, 8-15, 10-18, 17-19\}$. The battery consumption is taken to be linear in time and it is calculated based on the assumed daily driving distance and battery distance range. Uniform random variation between 0 and $\pm$ 20 % of the calculated consumption is added to this consumption value in order to take into account any random fluctuation of true consumption.

The price component $p_{n_{\min}}$ from the set of variables defining the state $\mathbf{x}_n$ is limited within the interval [60$, 250$] in terms of $ per MWh of energy. These values are the approximated extreme values of the histogram of $p_{n_{\min}}$ variables determined from the training data set. The 24 kWh battery capacity is quantified into *L*=24 levels, each level representing an interval of 1kWh.

The amount of energy to be charged in the battery during a day is computed using the proposed batch RL method by taking into consideration the current charge level in the battery, the amount of energy that it is expected to be consumed during the day and the period of time when the car is assumed to be at home during the day. In practice the smart device that controls the charging may be featured with a control option through which the car owner can choose himself the amount of energy to charge for the following day. Thus, the car user can override the charging decision of the smart device. The scheduling of the charging starts every night at 12 AM and is performed using the LP formulation in (5)-(8) such that the charging of the chosen amount happens during those hours of the day when the electricity price is minimum.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

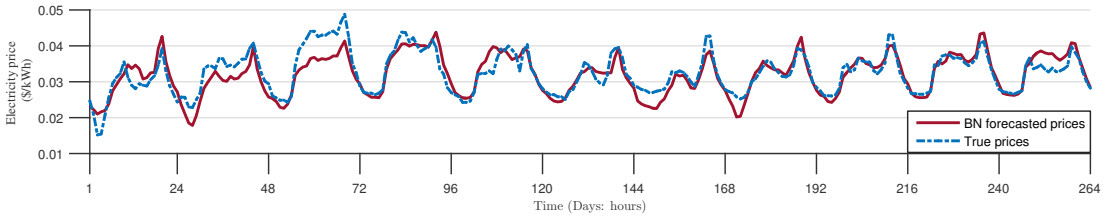

Fig. 1. A sequence of true hourly electricity prices and the corresponding sequence of forecasted electricity prices.

The performance of the PEV charging method was assessed by using actual pricing data taken from the ISO New England. The predicted prices are the ones obtained using the Bayesian Neural Network method. A set of 292 daily true and predicted prices was selected starting from the 14th of July 2011 until the 31st of April 2012. A subset of 182 daily prices was used in the learning stage of the developed *fitted Q-iteration* batch RL method proposed to solve the PEV charging problem. The remaining 110 daily prices were used for testing the cost reducing performance of the proposed method.

The four-tuple sample set $\mathcal{F}$ used in the learning stage of the *fitted Q-iteration* batch RL method was constructed according to the method presented in section III.B. This set was constructed using the employed data-set of λ=182 days of true historical electricity prices. The sampling starts with the first day of the time frame λ. For each day of the interval the corresponding day of the week was observed and the pricing components Δ_n and $p_{n_{\min}}$ were calculated. Multiple values representing the initial states of charge of the battery $b^{(0)}$ were randomly sampled from the battery state space [0, 24 kWh] and associated to the pricing variables. Several actions were randomly chosen with equal probability for each value in the set of possible actions. The reward was determined according to the LP optimization which was performed over a time window of δ=14 days. The penalty value was selected to be $\zeta_{\text{penalty}} = 0.2|u - u_{\text{opt}}|\zeta(u^*)$. The objective of our PEV charging strategy is minimizing the cost of charging. However, the proposed *fitted Q-iteration* based reinforcement learning algorithm computes a charging policy by maximizing the received rewards. The reward values in the proposed method contain a pricing component. In order to run the reinforcement learning algorithm such that the cost of charging is minimised, the reward values are all inversed such that: $r = -r + r_{\max} + r_{\min}$. Here $r_{\max}$ and $r_{\min}$ are the maximum and minimum reward values in the batch data-set $\mathcal{F}$. The next state is calculated using the price information of the next day from the time frame λ. The construction of the four-tuple sample set $\mathcal{F}$ was performed in Matlab and the process lasted for 1 hour 8 minutes and 43 seconds on a conventional desktop computer. The process can be performed offline using the provided set of data samples. For solving the LP optimization problem we employed standard interior point algorithms [31] provided by the CVX package for convex optimization [32].

The fitted Q-iteration algorithm was employed separately for each day of the week. For this purpose, the initial set of generated samples $\mathcal{F}$ was further divided into subsets based on the day of the week and action: $S_{d,u}$. In order to have equal number of samples for each action in each subset $S_{d,u}$, m=3120 samples were chosen randomly from $\mathcal{F}$ for each action, according to a uniform distribution. The overall training time of the algorithm was 22 minutes and 30 seconds.

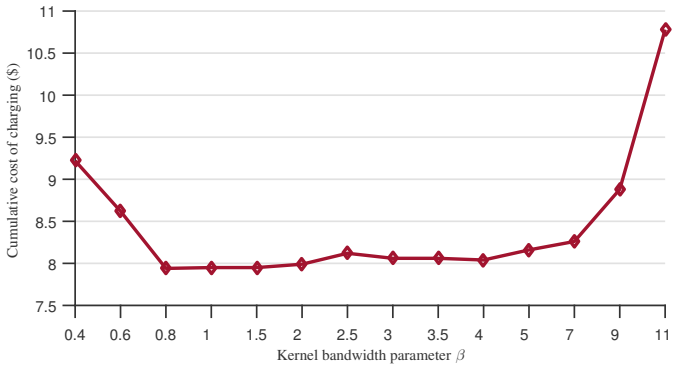

Fig. 2. Variation of the cumulative cost of charging as a function of the bandwidth parameter of the kernel function employed in the *fitted-Q iteration* algorithm. The most suitable value is β=0.8.

The weighting kernel function $\kappa(\mathbf{x}_l, \mathbf{x}_n)$, which depends on the current state of our PEV charging problem, is computed using (9). The state variables involved in the calculation of the kernel function were the initial state of charge of the battery $b_n^{(0)}$ and the pricing variables Δ_n and $p_{n_{\min}}$. Different kernel regression functions such as Laplace, Uniform and Gaussian were tested using different bandwidth values. For our PEV charging problem we found the Gaussian kernel function to give the best performance.

In the case of the Gaussian kernel, the bandwidth parameter actually represents the standard deviation of the Gaussian function. In order to choose the value of the bandwidth parameter we use a set of data different from the one used for the actual validation of the PEV charging method performance. Pricing data corresponding to the time period starting from 1st of May 2012 until 6th of February 2013 was used. This set was also divided in two subsets containing 182 daily prices for the learning stage and 100 daily prices for testing the performance of the PEV charging methods for different values of the bandwidth parameter.

Fig. 2 shows the variation of the cumulative cost of charging the PEV battery as a function of the bandwidth parameter β of the Gaussian kernel function employed in the *fitted-Q iteration* algorithm. In this plot we show the cumulative cost of charging the battery over 100 days. The most suitable bandwidth value is the one for which the cost is minimum. It can be observed that the result of the proposed PEV charging

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

Table III
SIMULATION PARAMETERS

Parameter	Value or expression
Discount factor γ	0.9
Kernel function $\phi(\frac{\|\mathbf{x}_l-\mathbf{x}_n\|}{\beta})$	$\frac{1}{\sqrt{2\pi}\beta} e^{\frac{-\|\mathbf{x}_l-\mathbf{x}_n\|^2}{2\cdot\beta^2}}$
Bandwidth parameter β	0.8

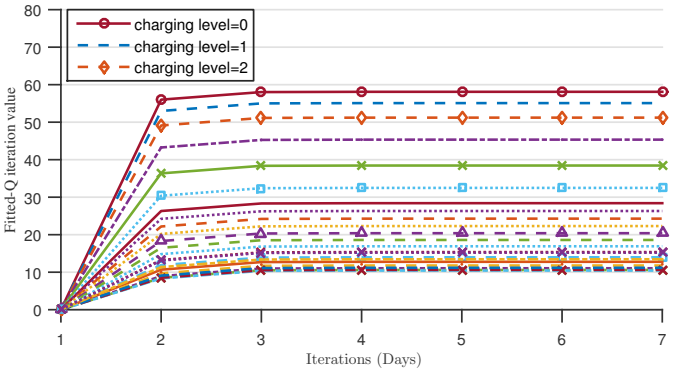

Fig. 3. Convergence to a steady-state of the approximate value iteration for the fitted Q-iteration algorithm during the learning process. For this particular case the best solution would be to charge zero levels.

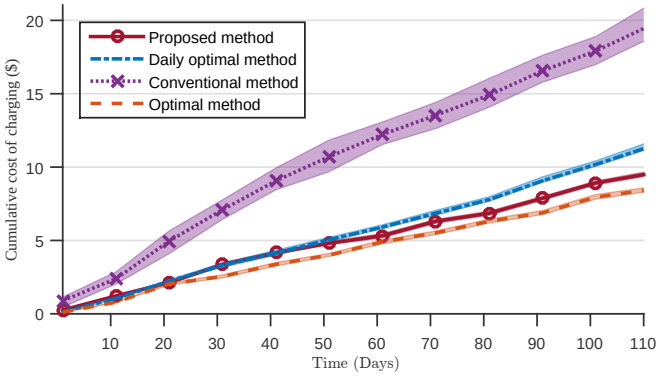

Fig. 4. Comparison of charging costs between the proposed novel charging strategy and other three charging strategies. The average cost values and the variance intervals of the methods are presented. The proposed novel method reduces the costs of charging by roughly 50% when compared with the conventional charging method and by roughly 10% when compared with a daily optimal charging strategy. To reach the global optimal solution the cost should further be reduced by 8%.

method is not very sensitive to the choice of this parameter value. The method shows a very slight increase in cost for β values within 0.8 and 5. For the rest of the simulations in this work we used a bandwidth parameter value β=0.8.

Table III presents the values of all parameters of the fitted Q-iteration algorithm used for solving the proposed PEV charging problem. These values were found experimentally for the PEV charging problem at hand. L_1 norm was used for the calculation of the kernel function.

The speed of convergence may also be used as a criterion for evaluating the performance of the proposed fitted Q-iteration algorithm. In order to ensure convergence, the algorithm was run for multiple iterations. Fig. 3 shows the evolution of the approximate value iterations of the fitted-Q function during the learning process for a particular state over 7 iterations. It can be seen that the state action values converge very quickly, typically after 5 iterations, to steady state levels after the initial learning phase. The best action is the one that maximizes the value of the Q-function at the end of the learning stage. The legend in Fig. 3 explains the curves of the Q-function values corresponding to charging zero, one and two levers. These actions are the ones that maximizing the the value of the Q-function for this particular state. The best action for this particular state would be to charge zero levels, which means not to charge at all during that day. The values of the Q-function corresponding to the rest of the possible actions are plotted as well.

The proposed PEV charging strategy was tested over time intervals of 14 days. After each 14-day interval, the training data set was updated with the newest pricing values and the learning stage was executed again. The performance of our method was tested for a total number of 110 days. Fig. 4 shows the cumulative costs of charging the PEV over 110 days. The proposed novel PEV charging method is compared with other PEV charging policies. The initial state of charge of the battery at the beginning of the test was 0 for all the methods. Three other charging strategies were tested for comparison. The conventional method simulates the charging behavior of non-price conscious drivers that charge the battery of the car when the battery is almost empty without considering the cost. In this case the charge amount is a random value between the amount of energy needed for the next drive and charge to the full capacity of the battery. The daily optimal charging method is a smart deterministic charging method in which the exact daily consumption is considered to be known at the beginning of the day and the battery is charged with that amount of energy such that the electricity cost is minimized. Finally, the optimal charging method is a fully deterministic method for which it is assumed that the price of electricity and the exact daily consumption amounts are known in advance for the whole considered time frame of 110 days. The battery is scheduled to be charged such that the overall cost of energy for the entire time frame is minimized. This method is the global optimal method which is not feasible in practice due to the lack of knowledge of the electricity prices for such a long time horizon in advance. The daily consumption values were generated randomly using the method explained earlier in this section and 50 different car usage realizations were performed. The plots in Fig. 4 show the average cost values and the variance intervals for each employed method. These variance intervals are given by the random component of the consumption. It can be observed that roughly a 50% decrease in charging costs is obtained when the proposed method is used instead of the conventional charging policy. Even if the battery would be charged daily at minimum price of the day with the exact knowledge on the amount of consumed energy, the car owner would still save money when using the proposed approach. A reduction of about 10% in the cost of charging was observed in our simulations. This happens because the proposed learning algorithm is able to avoid charging on certain days when the price of electricity is higher. In order

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2016.2603536, IEEE Transactions on Vehicular Technology

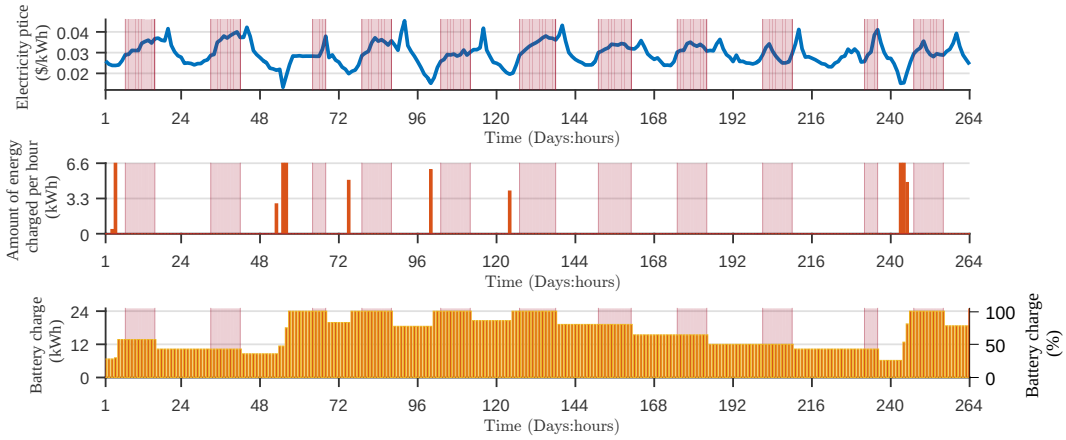

Fig. 5. An example of charging pattern of the proposed method over 11 consecutive days showing: the hourly price of energy (a), the amount of energy chosen to be charged during each hour of these days (b) and the charge level of the battery(c). The highlighted regions in each sub-figure indicate the hours of the day when the car is not at home. The method chooses to charge a higher amount of energy during the days when the price is lower and not to charge at all when the prices are higher.

to reach the global optimum solution the charging cost would still have to be reduced by roughly 8%. However, the proposed charging strategy reduces the charging costs and gets the result significantly closer to the globally optimal solution.

Fig. 5 shows the charging behavior of the proposed PEV charging method that avoids charging on days when the prices are higher and charges more on days when the prices are lower. The highlighted regions in each sub-figure show the periods of time when the car is not at at home. The sequence of hourly electricity prices per kWh corresponding to 11 consecutive days and the amount of energy that the proposed PEV charging policy chose to charge during every hour for these days are presented in Fig. 5(a) and Fig. 5(b). The maximum hourly charging rate is 6.6 kW. Fig. 5(c) shows the amount of energy existing in the battery during each hour of the time frame. The left side scale of Fig. 5(c) shows the battery's charge amount in kWh, while the right side of the scale shows the charge amount in percentage of the total capacity $C_{\max}$. It can observed that the charging policy chooses to charge a certain level of energy during those days in which the price of electricity is lower than in the rest of the days. The policy chooses not to charge at all during the days when the price of electricity is higher. When the price of electricity decreases again the policy chooses to charge a larger amount of energy.

VI. Conclusion

A novel method for scheduling the PEV battery charging was proposed in this paper. The method was formulated as a daily decision making problem that chooses the amount of energy to be charged in the PEV battery within a day. The method exploits the knowledge of true day ahead prices and predicted prices for the second day ahead. The goal of the proposed PEV charging method is to decrease the cost of charging to a consumer over a long term time horizon. A method using Bayesian neural networks has been proposed for forecasting the electricity prices employed in the PEV charging method. The PEV charging problem has been formulated as a MDP with four state variables. The problem was solved using a batch reinforcement learning method that employs the *fitted Q-iteration* algorithm with a kernel based approximation of the value iteration. In order to avoid the battery depletion the user's driving patterns are taken into account. A LP based method is developed for making optimal charging decisions when constructing the batch of transition samples employed by the reinforcement learning algorithm. Simulations indicate that by using the proposed Bayesian method the predicted electricity prices show an improvement of 3.34% and 0.94% in the prediction accuracy in comparison to the ARIMA and VJ-BN methods. The performance of the proposed PEV charging scheme was studied using real world pricing data [17] and predicted prices obtained using the Bayesian method. Simulations show that the method reduces the charging cost by 10% - 50%, depending on the scenario.

References

[1] X. Xi and R. Sioshansi, "Using price-based signals to control plug-in electric vehicle fleet charging," *IEEE Trans. on Smart Grid*, vol. 5, no. 3, pp. 1451–1464, May 2014.

[2] Z. Tan, P. Yang, and A. Nehorai, "An optimal and distributed demand response strategy with electric vehicles in the smart grid," *IEEE Trans. on Smart Grid*, vol. 5, no. 2, pp. 861–869, March 2014.

[3] C. Jin, J. Tang, and P. Ghosh, "Optimizing electric vehicle charging: A customer's perspective," *IEEE Trans. on Vehicular Technology*, vol. 62, no. 7, pp. 2919–2927, Sept 2013.

[4] Y. He, B. Vankatesh, and L. Guan, "Optimal scheduling for charging and discharging of electric vehicles," *IEEE Trans. on Smart Grid*, vol. 3, no. 3, pp. 1095–1105, September 2012.

[5] M. Sarker, M. Ortega-Vazquez, and D. Kirschen, "Optimal coordination and scheduling of demand response via monetary incentives," *IEEE Trans. on Smart Grid*, vol. 6, no. 3, pp. 1341–1352, May 2015.

[6] S. Yoon, Y. Choi, S. Bahk, and J. Park, "Stackelberg game based demand response for at-home electric vehicle charging," *IEEE Trans. on Vehicular Technology -(to appear)*.

[7] N. Rotering and M. Ilic, "Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets," *IEEE Trans. on Power Systems*, vol. 26, no. 3, pp. 1021–1029, August 2011.

[8] J. Donadee and M. Ilic, "Stochastic optimization of grid to vehicle frequency regulation capacity bids," *IEEE Trans. on Smart Grid*, vol. 5, no. 2, pp. 1061–1069, March 2014.

[9] E. B. Iversen, J. M. Morales, and H. Madsen, "Optimal charging of an electric vehicle using a Markov decision process," *Applied Energy*, vol. 123, pp. 1–12, 2014.

[10] M. M. Karbasioun, I. Lambadaris, G. Shaikhet, and E. Kranakis, "Optimal charging strategies for electrical vehicles under real time pricing," in *Proc. of the IEEE International Conference on Smart Grid Communications (SmartGridComm)*, November 2014.

[11] D. Tang and P. Wang, "Probabilistic modeling of nodal charging demand based on spatial-temporal dynamics of moving electric vehicles," *IEEE Trans. on Smart Grid*, vol. 7, no. 2, pp. 627–636, March 2016.

[12] S. Vandael, B. Claessens, D. Ernst, and T. H. G. Deconinck, "Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market," *IEEE Trans. on Smart Grid*, vol. 6, no. 4, pp. 1795–1805, July 2015.

[13] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, October 2010.

[14] W. Zheng, D. O'Neill, and H. Maei, "Optimal demand response using device-based reinforcement learning," *IEEE Trans. on Smart Grid*, vol. 6, no. 5, pp. 2312–2324, Sept 2015.

[15] A. Chiş, J. Lundén, and V. Koivunen, "Scheduling of plug-in electric vehicle battery charging with price prediction," in *Proc.4th IEEE Power and Energy Society Conference on Innovative Smart Grid Technologies Europe*, October 2013.

[16] ——, "Optimization of plug-in electric vehicle charging with forecasted price," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015.

[17] [Online]. Available: www.iso-ne.com

[18] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994, 2005.

[19] P. Grahn, J. Munkhammar, J. Widén, K. Alvehag, and L. Söder, "PHEV home-charging model based on residential activity patterns," *IEEE Trans. on Power Systems*, vol. 28, no. 3, pp. 2507–2515, August 2013.

[20] J. Donadee, M. Ilic, and O. Karabasoglu, "Optimal autonomous charging of electric vehicles with stochastic driver behavior," in *Proc. of 2014 IEEE Vehicle Power and Propulsion Conference (VPPC)*, October 2014.

[21] D. Ormoneit and S. Sen, "Kernel-based reinforcement learning," *Machine learning*, vol. 49, no. 2-3, pp. 161–178, 2002.

[22] R. E. Bellman, *Dynamic Programming*. Princeton University Press, 1957.

[23] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. on Power Systems*, vol. 18, no. 3, pp. 1014–1020, August 2003.

[24] N. Amjady, "Day-ahead price forecasting of electricity markets by a new fuzzy neural network," *IEEE Trans. on Power Systems*, vol. 21, no. 2, pp. 887–896, May 2006.

[25] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong, and K. P. Wong, "A hybrid approach for probabilistic forecasting of electricity price," *IEEE Trans. on Smart Grid*, vol. 5, no. 1, pp. 463–470, January 2014.

[26] V. Vahidinasab and J. Jadid, "Bayesian neural network model to predict day-ahead electricity prices," *European transactions on electrical power*, vol. 20, pp. 231–246, 2010.

[27] J. Lampinen and A. Vehtari, "Bayesian approach for neural networks-review and case studies," *Neural Networks*, vol. 14, no. 3, pp. 257–274, 2001.

[28] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Physics Letters*, vol. 195, no. 2, pp. 216–222, September 1987.

[29] [Online]. Available: www.becs.aalto.fi/en/research/bayes/mcmcstuff

[30] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis Forecasting and Control, Third edition*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[31] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming.*, vol. B, no. 95, pp. 189–217, 2003.

[32] M. Grant and S. Boyd. CVX: Matlab Software for Disciplined Convex Programming, Version 2.0 Beta, Sept 2013.

Adriana Chiş (S'13) received the B.S and M.S degrees in electronic engineering and telecommunications from the Technical University of Cluj-Napoca, Romania, in 2009 and 2011, respectively. She is currently working towards the Ph.D degree with the Department of Signal Processing and Acoustics, School of Electrical Engineering, Aalto University, Finland.

Her research interests include machine learning, statistical signal processing, game theory, optimization, and their applications in smart grids.

Jarmo Lundén (S'05, M'10) received the M.Sc. (Tech.) degree with distinction in communications engineering and the D.Sc. (Tech.) degree with distinction in signal processing for communications from the Helsinki University of Technology, Espoo, Finland, in 2005 and 2009, respectively.

From 2010 to 2011, he was a Visiting Postdoctoral Research Associate at Princeton University, Princeton, NJ, USA. He is currently a Staff Scientist at Aalto University, Espoo, Finland. His research interests include statistical signal processing and machine learning, and their applications in cognitive radio, radar, and smart grids.

Visa Koivunen (F'11) received his D.Sc. (EE) degree with honors from the University of Oulu, Dept. of Electrical Engineering. He received the primus doctor award among the doctoral graduates in years 1989-1994. He is a member of Eta Kappa Nu. From 1992 to 1995 he was a visiting researcher at the University of Pennsylvania, Philadelphia, USA. Since 1999 he has been a full Professor of Signal Processing at Aalto University (formerly known as Helsinki Univ of Technology), Finland. He received the Academy professor position (distinguished professor nominated by the Academy of Finland). Years 2003-2006 he was also adjunct full professor at the University of Pennsylvania, Philadelphia, USA. He has also been a part-time Visiting Fellow at Nokia Research Center (2006-2012). He has spent multiple mini-sabbaticals and two full sabbaticals at Princeton University.

Dr. Koivunen's research interests include statistical, communications, sensor array and multichannel signal processing. He has published more than 370 papers in international scientific conferences and journals and holds 6 patents. He co-authored the papers receiving the best paper award in IEEE PIMRC 2005, EUSIPCO'2006,EUCAP (European Conference on Antennas and Propagation) 2006 and COCORA 2012. He has been awarded the IEEE Signal Processing Society best paper award for the year 2007 (with J. Eriksson). He served as an associate editor for IEEE Signal Processing Letters and IEEE TR on Signal Processing. He has served as co-editor for two IEEE JSTSP special issues. He was a member of editorial board for IEEE Signal Processing Magazine. He has been a member of the IEEE Signal Processing Society technical committees SPCOM-TC and SAM-TC. He was the general chair of the IEEE SPAWC 2007 conference in Helsinki, Finland June 2007 and Technical Program Chair for the IEEE SPAWC 2015. He was awarded the 2015 EURASIP (European Association for Signal Processing) Technical Achievement Award for fundamental contributions to statistical signal processing and its applications in wireless communications, radar and related fields. He is a member of the IEEE Fourier Award committee and IEEE SPS Distinguished Lecturer for 2015-2016.