Adaptive Home Heating under Weather and Price Uncertainty using GPs and MDPs

Mike Shann and Sven Seuken
Department of Informatics
University of Zurich
{shannm, seuken}@ifi.uzh.ch

ABSTRACT

We consider the problem of adaptive home heating in the smart grid, assuming that real-time electricity prices are being exposed to end-users with the goal of realizing demand-side management. To lower the burden on the end-users, our goal is the design of a *smart thermostat* that automatically heats the home, optimally trading off the user's comfort and cost. This is a challenging problem due to two sources of uncertainty: future weather conditions and future electricity prices. Our main technical contribution is a general technique that uses predictive distributions obtained from Gaussian Process (GP) regressions to compute the state transition probabilities of an MDP, such that the solution to the resulting MDP constitutes a sequentially optimal policy. We apply this general approach to the home-heating problem, where we use the predictive distributions of the GPs for the day-ahead external temperatures and electricity prices. The solution to the home-heating MDP constitutes an optimal heating policy that maximizes the user's utility given the probability information gathered by the Gaussian process model. Via simulations we show that our MDP-based approach outperforms various benchmarks, especially for cost-sensitive users.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Plan execution, formation, and generation*

Keywords

Smart Grid; Home Heating; Real-time Prices; MDPs; GPs

1. INTRODUCTION

The electricity grid is undergoing big changes as many countries are now moving from fossil fuel burning power stations to renewable energies (solar, wind, tidal). This creates a number of challenges because energy from renewable sources is very volatile, energy is inherently difficult to store, and the classic model in energy markets is one where supply follows demand. Until now, end-users have generally faced fixed energy prices and were not aware of changes in supply and demand of energy. But with more and more renewable energy sources, this inelastic demand side becomes an increasingly severe problem [2]. For this reason, governments around the world are investing billions in the development of the next generation of the electricity grid, the so-called *smart grid*.

One important part of the smart grid vision is to create a paradigm shift that enables *demand-side management*. This means that in times when energy is scarce (and expensive), the demand for energy should adjust and go down, and when energy is plentiful (and cheap), the demand for energy should go up. One way to achieve demand-side management is by exposing real-time prices to end-users. While currently the biggest potential for demand-side management still lies in the industrial sector, this will change very soon, with more and more people driving electric vehicles and heating their homes with electricity instead of oil or gas.

1.1 Home Heating in the Smart Grid

The energy used for heating homes is a major part of many countries' energy consumption and consequently also accounts for a substantial part of their CO2 emissions. In the US, approximately 40% of household energy is used for heating [15], and in the UK it even accounts for 66% of household energy usage [7]. Thus, if the international community wants to meet its goal of reducing CO2 emissions as stated in the Kyoto protocol [14], reducing the energy used for heating must be part of the agenda. Of course, individual home owners also have an interest in this, given that home heating accounts for the majority of their household energy costs. Thus, the optimization of energy usage for home heating is an important lever to reduce CO2 emissions, to enable demand-side management, and to reduce individual home owners' energy costs.

There are two main avenues for improving the energy efficiency of homes. One is better insulations, which reduces the leakage of heat to the outside. However, this is often very expensive or not even worth it, especially for old buildings. The other avenue, which we consider here, is the optimization of the home heating control process. We assume that the heating device is a heat pump that works with electricity (an assumption that will be true for many households once renewable energy makes up the majority of the energy mix). Heat pumps offer the advantage of higher energy efficiency and lower CO2 emissions compared to conventional forms of heating, in particular when renewable energy sources are used to produce the electricity. When heat pumps are powered by electricity, home heating is obviously directly connected to the electricity market. Thus, for demand-side management to be effective, the home heating controller must be responsive to price changes, which adds a new complication to this problem.

1.2 Coping with Weather & Price Uncertainty

Our goal is to design a *smart thermostat* that has a model of the user's preferences and automatically adjusts the temperature as the environmental conditions affecting the heating (such as external temperature and electricity prices) change. To optimize the heating

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*
Copyright © 2014, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

strategy, the smart thermostat must "plan ahead," e.g., if electricity prices are about to rise, then the current cheap electricity should be used to heat up the house, such that it is already warm during times of high energy costs. Therefore, the smart thermostat first needs to predict the future development of the electricity price and the external temperature, and then use this information to compute a heating strategy that is optimal for the user.

Our work is motivated by earlier research on home heating by Rogers et al. [10], who developed an adaptive heating algorithm that first predicts future external temperatures using Gaussian Processes (GPs), and then computes a heating plan using mixed-integer programming. However, their algorithm implicitly assumes that the weather predictions are correct. Consequently, the home heating policy they compute might be sub-optimal because it does not account for the uncertainty inherent to weather forecasts. This issue is exacerbated when prices are dynamic and therefore not known perfectly in advance. In our approach, we also use GPs to predict future outside temperatures, as well as future electricity prices. However, we use Markov Decision Processes (MDPs) to explicitly account for the uncertainty of these predictions.

1.3 Overview of Contributions

Our main technical contribution in this paper is a general technique that uses the probabilistic predictions obtained from Gaussian Process regressions to define the state transitions for an MDP, such that a solution to the resulting MDP constitutes a sequentially optimal policy for the problem. This approach can be applied to any problem that requires a stochastic policy that is contingent on the future values of certain state variables.

We illustrate this general technique by applying it to the problem of computing a sequentially optimal home heating policy. Using the predictive distributions of the GPs for the day-ahead external temperature and electricity prices, the solution to the MDP constitutes a heating policy that maximizes the user's total utility. Our MDP formulation is very general, and can easily be extended to incorporate other sources of uncertainty (e.g., home occupancy).

We use simulations based on real-world weather data to compare our MDP-based algorithm against multiple benchmarks from the literature (including MIPs and MPCs). We demonstrate that our approach achieves the same or higher performance, and is particularly effective for cost-sensitive users.

2. RELATED WORK

Rogers et al. [11] provide an introduction to the smart grid from a multi-agent systems perspective, and Ramchurn et al. [8] describe the opportunities for AI research in this field. Vytelingum et al. [16] study autonomous agents for micro-storage in the smart grid that automatically react to price changes. However, their approach does not explicitly account for the uncertainty in the domain.

In our own prior work on home heating, we have studied how to automatically learn the user's preferences (trading off comfort with costs) with minimal interactions [13]. In this paper, we assume that the thermostat already has a good model of the user's preferences, and focus on computing a sequentially optimal heating policy. However, our MDP-based approach naturally lends itself towards incorporating preference elicitation techniques as presented in our prior work, which is subject to our ongoing research.

Various researchers have studied the problem of energy efficient heating control. A notable approach involves predicting the occupancy of the building with the goal of reducing the inside temperature when the building is unoccupied. For example, Scott et al. [12] use motion sensing, and find patterns in user behavior to heat adaptively. Occupancy prediction is complementary to weather and price prediction, but our MDP-based approach can easily be extended to also include an occupancy prediction component.

In the control community, the state-of-the-art method for home heating is *model predictive control (MPC)*. For example, *Opti-Control* is a project aiming at energy efficient heating of office buildings [5]. They consider weather forecasts and occupancy predictions, and use MPCs to compute a heating policy. A similar approach is followed by Yu et al. [4]. MDPs and MPCs share some commonalities, but there are also important differences. In Section 6.2, we provide a detailed comparison of the two methods.

3. THE MODEL

We consider the problem of computing a sequentially optimal home heating policy that reacts to changing environmental conditions.[1] We discretize every day into T intervals, each consisting of $\Delta t = 24h/T$ (a typical interval length is 10 minutes). Each time step, we consider three environmental variables: the internal temperature T_t^{int}, the external temperature T_t^{ext}, and the price of electricity p_t. The heat pump is controlled via a decision variable h_t. Depending on how well the heat pump can be controlled, this variable is either binary, $h_t \in \{0, 1\}$, corresponding to the heating being turned off or on; or the variable is continuous, $h_t \in [0, 1]$, corresponding to the pump operating at a certain level between zero and maximum power. To compute the optimal heating policy, we need the following four components:

1. A thermal model of the house,
2. a model of the user's preferences,
3. a prediction of future environmental conditions, and
4. an optimization method that, given the thermal model and the predictive information, computes an optimal heating plan according to some criterion of optimality.

We now explain each of these components in detail.

3.1 Thermal Model of the House

To model the thermal properties of the house, we adopt an approach that is widely used in the home heating literature [4, 10]. In this model, the internal temperature of the home, T_t^{int}, is affected by two antagonistic effects. On the one hand, the heat pump delivers heat at a certain rate that is the product of the electrical power of the pump, r_h, times its thermal efficiency, called *coefficient of performance (COP)*. Mathematically, the heat delivered by the pump is $r_h \cdot COP$, measured in Watt (W). On the other hand, heat leaks from inside the house to the environment at a rate that is proportional to the difference between the internal and external temperatures. The heat loss per time unit depends on the insulation of the house, which is quantified by the leakage rate λ (in W/K). Given this, the instantaneous gain (or loss) of energy in the home at time step t is computed as

$$Q_t = h_t r_h \cdot COP - \lambda \cdot (T_t^{int} - T_t^{ext}) + \epsilon_t, \qquad (1)$$

where ϵ_t is a random variable denoting fluctuations in the heat flow due to random effects not accounted for in the model (e.g., opening doors or windows). Note that Equation (1) is stochastic due to the random effect ϵ_t. However, the thermal properties of the home (i.e. the variables r_h, COP, and λ) can be learned, as demonstrated in [10], assuming that ϵ is independently distributed. Therefore it is sufficient to consider a deterministic version of Equation (1).

[1] Note that all models and techniques presented in this paper can also be applied in a straightforward way to compute an optimal cooling strategy (i.e., to control an air conditioner). However, we restrict ourselves to heating in this paper to simplify the exposition.

The internal temperature at a new time step is then computed as the sum of the previous internal temperature and the heat delivered to (or lost from) the home:

$$T_{t+1}^{int} = T_t^{int} + \frac{Q_t}{c_{air} \cdot m_{air}} \Delta t, \tag{2}$$

where we let c_{air} (unit: J/kg K) and m_{air} (unit: kg) denote the heat capacity and the mass of the air inside the building, respectively.

3.2 The User's Utility Function

Inherent to the home heating problem is the need for the user to trade off *comfort* (i.e., coziness) with the *costs* of heating. Therefore, the optimization has to take both aspects into account. In contrast to most of the prior work in the home heating domain, we follow a decision-theoretic approach and formalize this trade-off with the help of a *utility function*. In this paper, we use the following class of utility functions:

$$u(T_t^{int}, p_t) = \Big(\underbrace{a - b(T_t^{int} - T^*)^2}_{\text{value}} - \underbrace{c(p_t)}_{\text{cost}}\Big)\Delta t, \tag{3}$$

where T^* is the user's most preferred temperature, and p_t the price of electricity. The term $a - b(T^{int} - T^*)^2$ is the value function. The parameter a is the user's willingness to pay for his most preferred temperature (per unit of time), and $b(T^{int} - T^*)^2$ is a quadratic loss function, quantifying the amount of discomfort experienced (per unit of time) due to temperatures deviating from T^*.[2] The parameter b measures the user's sensitivity to temperature deviations.

The cost function $c(p)$ quantifies how much it costs to let the heater run per unit of time. It is given by:

$$c(p_t) = h_t r_h p_t, \tag{4}$$

and is determined by the state of the heater, h_t, the heater's electricity consumption, r_h, and the electricity price, p_t (in Cents/kWh).

4. TEMPERATURE & PRICE PREDICTION

We use GPs to predict future external temperatures as well as electricity prices because GPs are a powerful and flexible framework and have been successfully used to predict external temperatures [10, 6] as well as electricity prices [3]. Our approach is adapted from [10] and [6]. Due to space constraints, we only give a brief overview of GPs. For a more detailed treatment see [9].

4.1 The Prediction Task

Consider a time series $\mathbf{S} = (S(t_1), \dots S(t_N))$, e.g., for the external temperature. We use the vector notation $\mathbf{t} = (t_1, \dots, t_N)$ for past time steps, and $\hat{\mathbf{t}} = (\hat{t}_1, \dots, \hat{t}_T)$ for future time steps for which we want to make predictions. We assume that our training data $\mathbf{y} = (y_{t_1}, \dots, y_{t_N})$ is distorted by additive i.i.d. Gaussian noise: $y_{t_i} = S(t_i) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. Given historical data $\mathbf{y}$, we want to make a (probabilistic) prediction of our time series for the next T time steps: $\widehat{\mathbf{S}} = (\widehat{S}(\hat{t}_1), \dots \widehat{S}(\hat{t}_T))$.

4.2 Gaussian Process Predictions

A Gaussian process approximates the time series $\widehat{\mathbf{S}}$ via a multivariate normal distribution. It is specified by its mean $m(t_i)$ and covariance function $k(t_i, t_j)$. The prior distribution for $\widehat{\mathbf{S}}$ is given by

$$Pr(\widehat{\mathbf{S}}) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}(\hat{\mathbf{t}}, \hat{\mathbf{t}})\right), \tag{5}$$

where $\mathbf{K}(\hat{\mathbf{t}}, \hat{\mathbf{t}})$ is the covariance matrix of the input points, i.e. $\mathbf{K}_{i,j} = k(\hat{t}_i, \hat{t}_j)$. The posterior distribution after having learned data points $\mathcal{D} = \{(\mathbf{t}, \mathbf{y})\}$ is computed as

$$Pr(\widehat{\mathbf{S}}|\mathcal{D}) \sim \mathcal{N}(\mathbf{m}_{post}, \mathbf{K}_{post}), \text{ where}$$
$$\mathbf{m}_{post} = \mathbf{K}(\hat{\mathbf{t}}, \mathbf{t})\left(\mathbf{K}(\mathbf{t}, \mathbf{t}) + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{y}, \text{ and}$$
$$\mathbf{K}_{post} = \mathbf{K}(\hat{\mathbf{t}}, \hat{\mathbf{t}}) - \mathbf{K}(\hat{\mathbf{t}}, \mathbf{t})\left(\mathbf{K}(\mathbf{t}, \mathbf{t}) + \sigma_n^2 I\right)\mathbf{K}(\mathbf{t}, \hat{\mathbf{t}}).$$

[2]Note that we adopted this approach towards modeling discomfort from Rogers et. al [10].

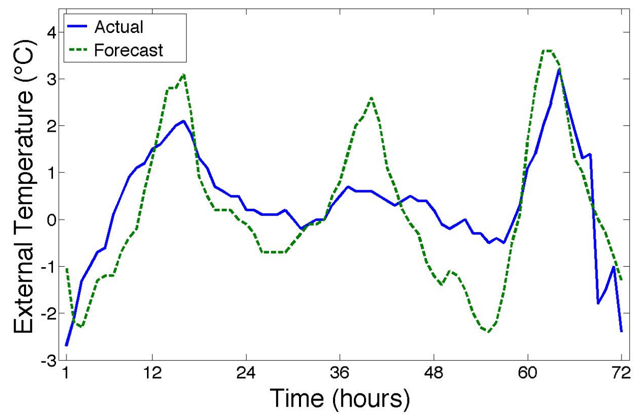


Figure 1: Three days of historical weather data from Zurich.

4.3 External Temperature Prediction

The main idea for the prediction of the external temperature is to train a GP using historical temperature measurements from the actual house as well as weather forecast data from a nearby meteorological service. Obviously, the local weather and the forecast should be highly correlated. In our data, provided by the Swiss national meteorological service MeteoSwiss, the correlation between forecasts and actual temperatures is approximately 0.9. Figure 1 shows a small sample of historical temperature data from Zurich. The green line is the temperature forecast for Zurich from the meteorological service, and the blue line is the actual temperature that was measured in one specific location. As we can see, the two time series are highly (but not perfectly) correlated.

Formally, we consider two temperature time series, one for the local measurements, which we denote as $T^L(t)$, and one for the forecasts, denoted $T^F(t)$. For both, we have historical data (i.e., one data point for every hour), but additionally we have a forecast for the next 24 hours. To use the GPs, we have to specify a model (via the covariance function of the GP) that captures the features of the external temperature sufficiently well. The four features that we model are: (i) daily rise and fall, (ii) rise and fall over longer periods of time (i.e. several days), (iii) erratic movements, and (iv) the correlation between the two time series. The covariance function k_{temp} for two data points (l, t) and (l', t') ($l \in \{Local, Forecast\}$ is the label of the series) is then given by

$$\begin{aligned} k_{temp}((l,t),(l',t')) = k_1(l,l')\,(k_2(t,t') + k_3(t,t')) \\ + k_4(t,t') + k_5(t,t'). \end{aligned} \tag{6}$$

Here, k_1 is a function that measures the cross-correlation between the time series, which is equal to one if the data points are from the same time series, and otherwise equal to θ_1:

$$k_1(l, l') = \begin{cases} 1 & \text{if } l = l', \\ \theta_1 & \text{otherwise.} \end{cases} \tag{7}$$

The covariance function k_2 encodes the daily rise and fall in temperatures. This is modeled using a periodic covariance function with period one day. However, the actual periodicity of weather is only approximately, but not exactly one day. To account for this, we multiply the periodic function with a squared exponential covariance function to allow for more complicated patterns:

$$k_2(t, t') = \theta_2^2 \exp\left(-\frac{(t-t')^2}{2\theta_3^2} - \frac{2\sin^2(\pi(t-t'))}{\theta_4^2}\right). \tag{8}$$

The rise and fall of the temperature over longer periods of time is modeled via a squared exponential covariance function:

$$k_4(t, t') = \theta_5^2 \exp\left(-\frac{(t-t')^2}{2\theta_6^2}\right). \tag{9}$$

The fourth covariance function in Equation (6), k_4, models erratic movements that are uncorrelated between the two time series (e.g. due to specific conditions at the measurement site). For this, we use a Matern class kernel as it is able reproduce such fluctuating temperature movements:

$$k_4(t, t') = \delta_{l,l'}\theta_7^2\left(1 + \frac{\sqrt{3}(t-t')}{\theta_8}\right)\exp\left(-\frac{\sqrt{3}(t-t')}{\theta_9}\right). \tag{10}$$

To account for measurement noise, which we assume to be i.i.d additive Gaussian, we use the following noise covariance function:

$$k_5(t, t') = \theta_{10}^2\delta_{t,t'}, \tag{11}$$

where $\delta_{t,t'}$ is the Kronecker delta between time points. Note that $\theta_1, \ldots, \theta_{10}$ are hyper-parameters of the GP whose values must be determined using maximum likelihood estimation.

4.4 Electricity Price Prediction

We model our price function according to characteristics found in spot market prices. According to Weron [17], a salient feature of daily prices is the periodicity: an increase in the morning (when people wake up), a decrease in the afternoon, and another increase in the evening (when people return home). We model this using a periodic covariance function with period half a day. As before, we allow for deviations from exact periodicity by multiplying the periodic covariance function with a squared exponential:

$$k_6(t, t') = \theta_{11}^2 \exp\left(-\frac{(t-t')^2}{2\theta_{12}^2} - \frac{2\sin^2(\pi(t-t'))}{\theta_{13}^2}\right). \tag{12}$$

The second feature of price movements are the erratic price fluctuations, which we model by a Matern class kernel:

$$k_7(t, t') = \sigma_6^2\left(1 + \frac{\sqrt{3}(t-t')}{\theta_{14}}\right)\exp\left(-\frac{\sqrt{3}(t-t')}{\theta_{15}}\right). \tag{13}$$

The covariance function for the price is the sum of k_3 and k_4 plus a noise term $k_8(t, t') = \theta_{16}^2\delta_{t,t'}$:

$$k_{price}(t, t') = k_6(t, t') + k_7(t, t') + k_8(t, t'). \tag{14}$$

Again, $\theta_{11}, \ldots, \theta_{15}$ are hyper-parameters of the GP whose values must be determined using maximum likelihood estimation.

5. HOME HEATING MDP

We now formalize the home heating problem as an MDP. An MDP is defined by a tuple (S, A, T, R), where S is the state space, A is the action space, $T : S \times S \times A \to \mathbb{R}_+$ is the state transition function, and $R : S \times A \to \mathbb{R}$ the reward function. We consider an MDP with a finite horizon of one day.[3] Defining the states, actions and the reward function is quite straightforward. The difficulty lies in computing the state transition probabilities, which is where the information obtained from the GPs is used.

States: The state space consists of the Cartesian product $S = \mathcal{T}^{int} \times \mathcal{T}^{ext} \times \mathcal{P} \times \mathit{TIME}$, where $\mathcal{T}^{int}$ and $\mathcal{T}^{ext}$ are the sets of internal and external temperatures, respectively, $\mathcal{P}$ the set of prices, and *TIME* the set of time steps for one day. Both, the prices and the temperatures are discretized, which is quite natural for the prices (in Cents), and also for the temperatures, since humans cannot notice the difference between two temperatures given a small enough level of granularity (e.g., between 22.0 and 22.5 degrees Celsius). To simplify the exposition, we denote the state $s = (T^{int}, T^{ext}, p, t)$ as $s_t = (T_t^{int}, T_t^{ext}, p_t)$.

Actions: The action space is $A = \{0, 1/(N_A - 1), 2/(N_A - 1), \ldots, 1\}$, where N_A is the number of actions available. For example, if $N_A = 2$, then $A = \{0, 1\}$, which corresponds to the heater being off or on, respectively, i.e., setting $h_t = 0$ or $h_t = 1$.

Reward Function: The reward function is simply the user's utility function, i.e., the user's value for a certain internal temperature T_t^{int} minus the cost of heating:

$$R(s_t, h_t) = \left(a - b(T^* - T_t^{int})^2 - h_t r_h p_t\right) \cdot \Delta t. \tag{15}$$

State Transition Function: The state transition function is a function that specifies, for every triple $(s_t, s_{t+1}, a) \in S \times S \times A$, the probability of arriving at state s_{t+1} if action h_t is taken in state s_t:

$$T(s_{t+1}, s_t, h_t) = Pr\left((T_{t+1}^{int}, T_{t+1}^{ext}, p_{t+1}) | (T_t^{int}, T_t^{ext}, p_t), h_t\right). \tag{16}$$

Although computing these probabilities might seem daunting at first, we can greatly simplify this task by making a few observations. First, note that T_{t+1}^{ext} and p_{t+1} are independent of T_t^{int} and h_t; and also of p_t and T_t^{ext}, respectively.[4] Furthermore, T_{t+1}^{int} does not depend on p_t. Exploiting these independencies, we can write:

$$T(s_{t+1}, s_t, h_t) = Pr\left(T_{t+1}^{int} | T_t^{int}, T_t^{ext}, h_t\right) Pr\left(T_{t+1}^{ext} | T_t^{ext}\right) Pr\left(p_{t+1} | p_t\right). \tag{17}$$

Note that the state transition for the internal temperature is deterministic by assumption and can be derived via Equation (2). However, the evolution of the external temperature and the price is stochastic. Therefore, we derive the transition probabilities $Pr(T_{t+1}^{ext} | T_t^{ext})$ and $Pr(p_{t+1} | p_t)$ using the probabilistic information gathered from the GPs.

5.1 Transition Probabilities for External Temperatures and Prices

We now describe how to derive the transition probabilities for the external temperatures. The approach is completely analogous for the electricity prices.

Recall that the GP gives us a predictive distribution for $\widehat{\mathbf{T}}^{ext} = (\widehat{T}^{ext}(\hat{t}_1), \ldots, \widehat{T}^{ext}(\hat{t}_{|TIME|}))$ that is a multivariate normal distribution

$$Pr\left(\widehat{\mathbf{T}}^{ext} | \mathcal{D}\right) \sim \mathcal{N}(\mathbf{m}, \mathbf{K}). \tag{18}$$

[3] We use a finite-horizon MDP for two reasons. First, the thermal effects of heating and heat leakage manifest themselves within minutes to hours. Therefore, optimizing the heating now will not greatly affect the heating for tomorrow or even further away. Second, an infinite horizon MDP would assume a stationary model of the external temperature and electricity prices. However, it is much better to predict the external temperature and electricity prices using day-ahead forecasts, thus dropping the stationarity assumption.

[4] Of course, the price may in practice also depend on the external temperature because weather conditions influence demand for energy and therefore, if many people have to heat a lot at the same time, prices may increase. However, we ignore this dependency to simplify the exposition.

Algorithm 1: Home Heating Algorithm

Input: utility function u
Variables: internal temperature T_t^{int}, external temperature T_t^{ext}, price p_t, optimal heating policy π^*
begin
  foreach *day* **do**
    $\widehat{P} \leftarrow GP.predictPrices()$
    $\widehat{T} \leftarrow GP.predictExternalTemperature()$
    $M \leftarrow$ **new** $MDP(u, \widehat{P}, \widehat{T})$
    $M.computeTransitionProbabilities()$
    $\pi^* \leftarrow M.computeOptimalPolicy()$
    for *t=1 to # of time steps per day* **do**
      $(T_t^{int}, T_t^{ext}, p_t) \leftarrow M.updateEnvironment()$
      $M.heatOptimally(\pi^*, T_t^{int}, T_t^{ext}, p_t)$

For the state transition function we need to compute conditional probability distributions of the form

$$Pr\left(\widehat{T}^{ext}(\hat{t}_i) = T | \widehat{T}^{ext}(\hat{t}_{i-1}) = \widetilde{T}, \mathcal{D}\right) \tag{19}$$

for all $i = 2, \ldots, |TIME|$ and $T, \widetilde{T} \in \mathcal{T}^{ext}$. We perform these computations in two steps: First, we compute the conditional distribution of $\widehat{T}^{ext}(\hat{t}_i)$ given $\widehat{T}^{ext}(\hat{t}_{i-1})$. Second, we integrate the conditional distribution to obtain a discrete conditional probability distribution.

Step 1: The conditional distribution can be computed as follows:

$$Pr\left(\widehat{T}^{ext}(\hat{t}_i) | \widehat{T}^{ext}(\hat{t}_{i-1}) = \widetilde{T}, \mathcal{D}\right) \sim \mathcal{N}(m_{cond}, \sigma_{cond}), \text{ with} \tag{20}$$

$$m_{cond} = m_i + \frac{\mathbf{K}_{i,i-1}}{\mathbf{K}_{i-1,i-1}} \cdot (\widetilde{T} - m_{i-1}), \text{ and}$$

$$\sigma_{cond} = \mathbf{K}_{i,i} - \frac{(\mathbf{K}_{i,i-1})^2}{\mathbf{K}_{i-1,i-1}}.$$

Step 2: We then integrate the conditional distribution over the interval $[T - \alpha, T + \alpha]$, where α is half of the discretization size in the temperature space, to obtain our final discretized transition function for the external temperature:

$$Pr\left(T_{t_i^*}^{ext} = T | T_{t_{i-1}^*}^{ext} = \widetilde{T}\right) = \int_{T-\alpha}^{T+\alpha} Pr\left(T_{t_i^*}^{ext} = y | T_{t_{i-1}^*}^{ext} = \widetilde{T}, \mathcal{D}\right) dy. \tag{21}$$

For example, if the discretization is $\mathcal{T}^{ext} = \{0, 1, 2, \ldots\}$ and we would like to compute the probability that the external temperature is 6°C after being 5°C, then

$$Pr\left(\widehat{T}^{ext}(\hat{t}_i) = 6 | \widehat{T}^{ext}(\hat{t}_{i-1}) = 5\right) = \int_{5.5}^{6.5} Pr\left(\widehat{T}^{ext}(\hat{t}_i) = y | \widehat{T}^{ext}(\hat{t}_{i-1}) = 5, \mathcal{D}\right) dy.$$

Finally, we normalize all probabilities computed this way to obtain a correct conditional probability distribution.

5.2 Computing an Optimal Policy

Now that we have constructed all components of the MDP, we can compute an optimal policy using dynamic programming (DP). Note, that at every iteration of the DP algorithm, one must take care to only include the reachable states in the time dimension (i.e., only consider states that are one time step earlier). The resulting optimal policy π^* corresponds to the optimal value function V^* that solves the Bellman optimality equation:

$$V^*(s) = \max_a \left(R(s, a) + \sum_{s'} Pr(s' | s, a) V^*(s') \right). \tag{22}$$

Thus, the optimal policy prescribes the action that maximizes the sum of the one-step reward and the expected utility going forward, assuming that the optimal policy is followed in the future. A summary of the whole heating algorithm is provided in Algorithm 1.

6. EXPERIMENTS

We evaluate our MDP-based heating algorithm via two simulation experiments. In Experiment I, we consider a simple heater that is either switched on or off. In Experiment II, we consider a heater that can work at any level between zero and maximum power.

We consider two different pricing scenarios: *times-of-use pricing* and *real-time pricing*. Times-of-use pricing models a situation in which the electricity provider sets fixed prices for certain specified (and fixed) periods of the day. Under real-time pricing the electricity price changes according to real-time market conditions. Because the actual demand and supply of energy depends on many factors (e.g., available utilities and weather conditions), real-time prices can only be predicted with a high level of uncertainty.

6.1 Experiment I: MDP vs. MIP

For Experiment I, we consider a heat pump that can only be switched on or off. We compare our MDP-based algorithm against three benchmark algorithms: a conventional thermostat that implements a simple rule-based heating policy, and a mixed integer program (MIP) that comes in two versions: one that aims to minimize heating costs, and another one that maximizes the user's utility.

Conventional Thermostat. A conventional thermostat tries to keep the room temperature around a set temperature T_{set} by implementing the following rule:

$$h_t^{thermostat} = \begin{cases} 0 & \text{if } T_{t-1}^{int} > T^{set} + \Delta T \\ 1 & \text{if } T_{t-1}^{int} < T^{set} - \Delta T \\ h_{t-1}^{thermostat} & \text{otherwise} \end{cases} \tag{23}$$

Mixed-Integer Program. Our second benchmark algorithm is a MIP, introduced by Rogers et. al [10].[5] The MIP minimizes the heating costs, subject to the constraint that the cumulative discomfort does not exceed a maximum discomfort level. Discomfort is measured as a quadratic loss function as in the utility function defined in Equation (3). We let $h_t \in \{0, 1\}$ denote the decision variables, c_t the cost of heating, d_t the discomfort, and D_{max} the maximum discomfort level.[6] The whole MIP can be stated as:

$$\min \sum_t h_t c_t \tag{24}$$

$$\text{s.t.} \quad Q_t = h_t r_h COP - \lambda \cdot (T_t^{int} - T_t^{ext})$$

$$T_{t+1}^{int} = T_t^{int} + \frac{Q_t}{c_{air} \cdot m_{air}} \Delta t$$

$$d_t = (T^* - T_t^{int})^2$$

$$\sum_t d_t \leq D_{max}$$

Note that this approach implicitly assumes that prices and external temperatures are known in advance. In particular, only the mean predictions from the GP are used instead of the whole distribution.

We also consider a modified version of this MIP, where the objective function is changed to now maximize the user's utility, which is defined analogously to the reward function of the MDP (see Equation (15)). Additionally, we drop the constraint that the discomfort should not fall below a certain target discomfort level.

[5]We thank the authors of [10] for providing their CPLEX code.
[6]D_{max} is set to the level of discomfort the user would experience if a conventional thermostat were run instead of the MIP.

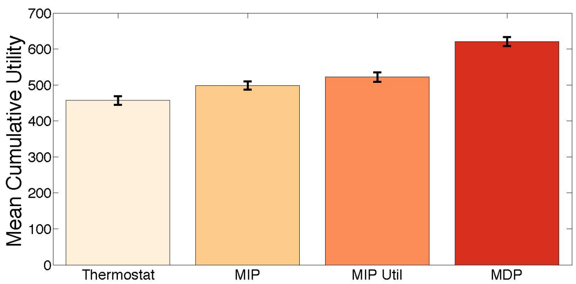


Figure 2: Results for real-time pricing. The bars denote (from left to right): the thermostat, the MIP that minimizes cost, the MIP that maximizes utility, and the MDP. Performance is measured as the mean cumulative utility, averaged over the month of January 2013.

6.1.1 Experimental Set-up

We let every algorithm heat sequentially for 31 days (i.e., for one month). Each day consists of 144 time steps, which corresponds to 10 minute time intervals. We use weather data for Zurich from December 2012 to January 2013, provided by the Swiss national meteorological service MeteoSwiss, which contains hourly forecasts for Zurich as well as actual measurements for one specific location in Zurich. We use the data from December to train the GPs, and the one from January for the actual experiment.[7]

In the times-of-use pricing scenario, there are three tariffs: 10, 20, and 40 cents/kWh. The prices are known a priori to all algorithms. In the real-time pricing scenario, the prices are generated using a GP that has the same model as described in section 4.4 to produce synthetic, but realistic pricing data. The GPs used for prediction are trained on the (synthetic) price data from December.

For the user's utility function, we set the parameters $a = 8/\Delta t$ and $b = 1/\Delta t$. These values are chosen such that values and costs in the reward function of the MDP (see Equation (15)) are approximately of the same size, corresponding to a user for whom, at typical prices, comfort and costs have comparable magnitudes.

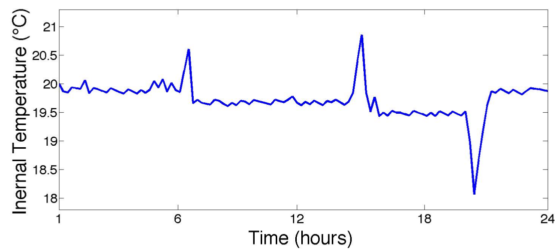
The dimensions of the house are $1,000\,m^3$, the mass of air in the house $m_{air} = 1,205\,kg$, the leakage rate $\lambda = 90\,W/kg$, the heat capacity $c_{air} = 1,000\,J/kg/K$. The power of the heater is $r_h = 1,500\,W$ with a $COP = 2.5$. The values are adopted from [10] and correspond to a small, well insulated home.

For the MDP, we discretize the temperature in steps of $0.5°C$ and the prices in steps of 5 cents. This set-up results in an MDP with approximately 1 million states that can be solved optimally in a few seconds on a standard PC. We use IBM ILOG CPLEX to solve the MIP and adopt the same approach as in [10] restricting CPLEX to run for 5 minutes per problem instance. The solver produces iteratively improving solutions. Thus, if the MIP does not terminate within the 5 minutes, it returns the best solution found so far.

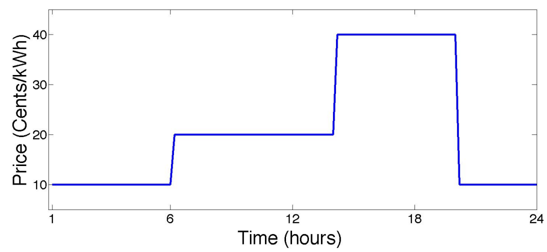
6.1.2 Results and Discussion

Figure 2 shows the results for the real-time pricing scenario (the results for the times-of-use scenario are qualitatively the same). We report the average cumulative utility (and standard errors) achieved by the different algorithms. We see that the MDP-based algorithm provides significantly higher average utility compared to all other approaches, improving the utility by more than 15%.

[7]The root mean squared errors of the GP and the meteorological service are $RMSE_{GP} = 1.29$ and $RMSE_{MF} = 1.61$, respectively. The approximately 20% improvement in accuracy makes sense since the GP can adapt to the peculiar climatic conditions (e.g., trees that provide shade) at the specific location whereas the forecast does not include this information.



(a)



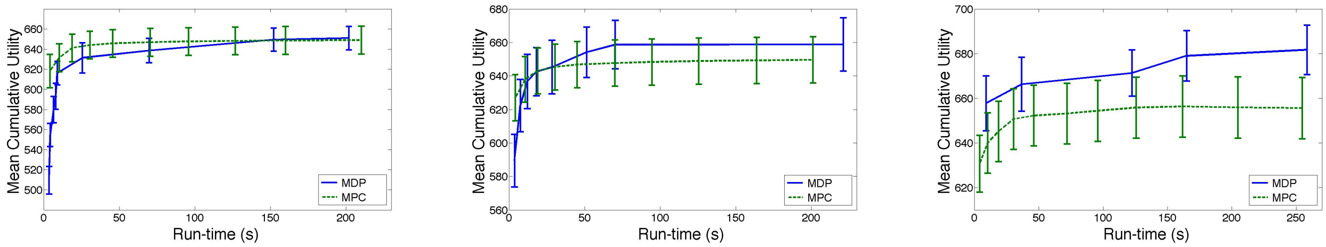
(b)

Figure 3: (a) Temperature profile over the course of one day, when running the MDP. (b) The corresponding times-of-use tariff.

There are several observations to discuss. First, we see that the conventional thermostat performs worst. This makes sense, because it completely neglects the cost component of the user's utility function. Second, the cost-minimizing MIP performs slightly better than the thermostat. Recall that it computes cost-minimizing heating plans that do not to exceed a certain level of discomfort. Third, we see that the utility-maximizing MIP performs better than the cost-minimizing MIP, which demonstrates that cost-minimization subject to comfort constraints only imperfectly approximates the maximization of the user's utility. However, even the utility-based MIP is still significantly worse than the MDP. This is because the MIP implicitly assumes that the predictions for the external temperatures and for the prices are perfectly accurate. If this assumption is not correct, then the MIP leads to sub-optimal decisions, e.g., not pre-heating when prices are low, or not saving energy when the outside temperature is about to rise, which leads to significantly higher discomfort or costs, compared to our MDP-based algorithm.

To illustrate the MDP-based approach, Figure 3(a) provides an example of the internal temperature profile that results from executing the MDP-based heating policy, while the corresponding times-of-use prices are shown in Figure 3(b).[8] By tracing the temperature curve over the course of the day, we gain insights into how the MDP optimizes the trade-off between comfort and costs. We see that just before the price goes up at the 6-hour mark, the MDP preheats a little bit, exploiting the low prices. It then uses less energy than before, consequently leading to a slightly lower temperature. Just before the next price increase at the 14-hour mark, it pre-heats again, exploiting the 20 cents/KWh price. Over the next six hours it uses even less energy than in the previous eight hours, leading to an even lower temperature. Just before the price goes back to 10 cents/KWh, the MDP essentially stops heating (to conserve costs in the high price regime), which leads to a momentary drop in temperature. Once the low price regime is reached, normal heating resumes, and the temperature goes back to the original level.

[8]Note that to produce the graph in Figure 3(a), we considered the scenario from Experiment II where the heater can be set to different levels between zero and maximum power. In particular, we used an MDP with 25 actions instead of just *on/off*, because the resulting temperature curve more cleanly illustrates the MDP policy.



(a) $b = 2$ (b) $b = 1$ (c) $b = 0.1$

Figure 4: A comparison between MDPs and MPCs. The graphs show the mean cumulative utilities for the MDP (solid blue line) and the MPC (dotted green line), for different values of the sensitivity parameter b (a smaller b corresponds to a more cost-sensitive user).

6.2 Experiment II: MDP vs.MPC

For Experiment II, we assume that the heater can work at any level between zero and maximum power. We compare our MDP-based algorithm (now with more than two actions) to an approach that uses model predictive control [18], which has proven successful in the heating domain [4, 5].

Model Predictive Control. MPCs are online algorithms that iteratively solve an optimization problem for a given time horizon to find the best sequence of (continuous) control actions, but only apply the first action to the system. After each time step, the system state is observed and a new optimization problem is solved given the new state. Thus, the time horizon is shifted one time step into the future. Applied to the home heating problem, this means that at every time step, we solve the following optimization problem:

$$\max_{h_t} \sum_t \Big(a - b(T^* - T_t^{int})^2 - h_t r_h p_t\Big) \cdot \Delta t, \text{ s.t.} \tag{25}$$

$$Q_t = h_t r_h COP - \lambda \cdot (T_t^{int} - T_t^{ext})$$

$$T_{t+1}^{int} = T_t^{int} + \frac{Q_t}{c_{air} \cdot m_{air}} \Delta t$$

$$h_t \in [0, 1].$$

This optimization problem is a quadratic program, which can be solved very efficiently. As in Experiment I, the time horizon is set to 24 hours. The predictions for the external temperature and prices are computed via GPs, in the same manner as was done for the MIPs. However, in contrast to the MIP-based approach, the GPs are updated using the new measurements of the external temperature and the current price made available at the end of each time step.

This particular version of an MPC is called *certainty equivalent MPC (CE MPC)*, because the external temperature and the electricity prices are set to the values predicted by the GPs, ignoring probabilistic effects. However, this loss of information is countered by the fact that the MPC works in an online fashion.

Note that MPCs work similarly to MDPs in the sense that MDPs *solve* the Bellman optimality equation (see Equation (22)) for a discretized version of the problem, and MPCs *approximate* the Bellman optimality equation for a continuous version of the problem. However, it is also informative to consider in more detail how the two methods differ. First, while the MDP computes an optimal policy that provides the optimal action for *every* state, the CE MPC only yields a policy for the states it believes it will encounter, given its model, the initial conditions, and the current predictions. Second, the bulk of the MDP computations are performed offline (i.e., at the beginning of the day), while the computations for the MPC (i.e. updating predictions, solving the optimization problem) must be repeated every time step. Third, the run-time of the MDP (given a fixed problem size) grows polynomially with the number of time steps, actions, and states. Thus, if we increase the discretization granularity for all three components simultaneously, then the run-time grows cubically. For the MPC, the trade-off between run-time and performance is less pronounced since only the time discretization matters. Finally, MDPs offer a rich language to model sequential decision making under uncertainty, whereas MPCs only have limited ability to model probabilistic environments.[9]

6.2.1 Experimental Setup

For both, the MDP and the MPC, we need to make a trade-off between computational complexity and performance: the finer the discretizations (time, actions, and states for the MPD; time for the MPC), the better the performance, but also the higher the computational burden. In this section, we study this trade-off in detail.

The basic experimental setup is similar to the one used in the real-time pricing scenario of Experiment I. We run every algorithm for 18 days and report average cumulative utility. However, we use different discretizations for the MDP and the MPC corresponding to different run-times. For the MPC, we vary the number of time steps from 24 to 192. For the MDP, we increase the number of time steps from 24 to 192, and the number of actions from 2 to 20, and report the highest utility achieved for a particular run-time.

We also vary the sensitivity parameter b of the utility function (see Equation (15)). We consider three values for b that corresponds to three different types of users: a value of $b = 2$ corresponds to a comfort-sensitive user; a value of $b = 1$ corresponds to an approximately equal weighting of comfort and cost; a value of $b = 0.1$ corresponds to a cost-sensitive user.

6.2.2 Results and Discussion

Figure 4 shows the results of Experiment II. The solid blue line and the dotted green line correspond to the MDP and the MPC, respectively. The graphs plot the average cumulative utility per day (on the y-axis) versus the average time spent to compute the optimal heating policy for one day (on the x-axis).

First, we see that for $b = 2$ and $b = 1$, there is no statistically significant difference between the expected utility achieved by the MDP and the MPC, except at very low run-times (less than 10 seconds), where the MPC outperforms the MDP. For cost-sensitive users ($b = 0.1$), the MDP leads to higher expected utility than the MPC, and this difference is statistically significant for run-times larger than 150 seconds. This results makes sense: we expect the

[9] There also exist *stochastic MPCs* that can handle some forms of uncertainty. At the same time, there also exist more sophisticated methods for handling continuous state and/or action spaces for MDPs that avoid some of the limitations of a discretized state or action space. Furthermore, one could also consider online (rollout) methods for the MDP, which could speed up the computations and which would also allow the MDP to update the GP predictions based on the new information in each time step. All of these considerations are left to future research.

MDP to be better at saving costs because its probabilistic model enables it to better account for the stochastic prices and temperatures.

Overall, we see that the performance of both algorithms improves a lot in the beginning as the computational complexity is increased, but that the rate of improvement quickly diminishes. This effect is particularly strong for the MDP, which makes sense, because the MDP is severely limited at very low run-times (with a low level of discretization in three dimensions).

7. CONCLUSION AND FUTURE WORK

In this paper, we have studied the adaptive home heating problem under weather and price uncertainty. We have presented a general technique that uses the predictive distributions obtained from GP regressions to construct the state transition probabilities of a corresponding MDP. Applied to the heating domain, the solution to the resulting home heating MDP constitutes a sequentially optimal heating policy that accounts for all available probabilistic information. Via simulations, we have demonstrated that in a scenario where the heater is limited to being switched on or off, our approach outperforms all benchmark algorithms from the literature. For another scenario, where a heater can run at any level between zero and maximum power, we have compared our MDP-based approach against an MPC-based approach. In particular, we have studied the resulting trade-off between computational run-time and performance for MDPs and MPCs. Our results indicate that both algorithms lead to very similar performance, except at very low run-times, where the MPC is slightly better. However, for price-sensitive users, the MDP eventually leads to significantly higher expected utility than the MPC, because it is better to able to account for the stochastic nature of energy prices and outside temperatures.

One important advantage of our MDP-based solution is that it naturally lends itself towards incorporating our prior work on preference elicitation in the home heating domain [13]. For this paper, we have assumed that we already have a good model of the user's utility function. In practice, however, this model must be learned over time, while the thermostat already optimizes the heating policy. Towards this end, our future work will involve extending our MDP model to explicitly incorporate the preference elicitation decisions. Because the user's utility is never fully revealed to the thermostat, this leads to a partially-observable MDP (i.e., POMDP) [1]. In our future research, we will work towards this vision of a smart thermostat that acts optimally on the user's behalf by carefully eliciting the user's preferences while simultaneously computing an optimal heating policy that maximizes the user's expected utility.

Acknowledgments

We would like to thank Siddhartha Ghosh for insightful discussions and the anonymous reviewers for helpful comments. This research was supported by the Swiss National Science Foundation (SNSF).

8. REFERENCES

[1] C. Boutilier. A POMDP Formulation of Preference Elicitation Problems. In *Proceedings of National Conference on Artificial Intelligence (AAAI)*, 2002.

[2] P. Cramton and A. Ockenfels. Economics and Design of Capacity Markets for the Power Sector. Working Paper, University of Maryland, May 2011.

[3] D. J. Leith, M. Heidl, and J. V. Ringwood. Gaussian Process Prior Models for Electrical Load Forecasting. In *IEEE International Conference on Probabilistic Methods Applied to Power Systems*, 2004.

[4] Z. Y. L. McLaughlin, L. Jia, M. C. Murphy-Hoye, A. Pratt, and L. Tong. Modeling and Stochastic Control for Home Energy Management. In *Proceedings of the Power and Energy Society (PES) General Meeting*, 2012.

[5] F. Oldewurtel, A. Parisio, C. N. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth. Energy Efficient Building Climate Control Using Stochastic Model Predictive Control and Weather Predictions. In *American Control Conference (ACC), 2010*, pages 5100–5105. IEEE, 2010.

[6] M. A. Osborne, S. Roberts, A. Rogers, S. Ramchurn, and N. R. Jennings. Towards Real-time Information Processing of Sensor Network Data Using Computationally Efficient Multi-output Gaussian Processes. In *Proceedings of the 7th international conference on Information processing in sensor networks*, pages 109–120. IEEE Computer Society, 2008.

[7] J. Palmer and I. Cooper. Great Britain's Housing Energy Fact File. Departmant of Energy and Climate Change (DECC), 2011.

[8] S. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings. Putting the "Smarts" into the Smart Grid: A Grand Challenge for Artificial Intelligence. *Communications of the ACM*, 55(4):86–97, 2012.

[9] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

[10] A. Rogers, S. Maleki, S. Ghosh, and J. Nicholas R. Adaptive Home Heating Control Through Gaussian Process Prediction and Mathematical Programming. In *Proceedings of the 2nd International Workshop on Agent Technology for Energy Systems (ATES)*, 2011.

[11] A. Rogers, S. D. Ramchurn, and N. R. Jennings. Delivering the Smart Grid: Challenges for Autonomous Agents and Multi-agent Systems Research. In *Proceedings of Conference on Artificial Intelligence (AAAI)*, 2012.

[12] J. Scott, A. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar. PreHeat: Controlling Home Heating Using Occupancy Prediction. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp)*, 2011.

[13] M. Shann and S. Seuken. An Active Learning Approach to Home Heating in the Smart Grid. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*. 2013.

[14] U.N. Framework Convention on Climate Change. Kyoto Protocol. http://unfccc.int/kyoto_protocol/items/2830.php, December 1997.

[15] U.S. Energy Information Administration. Energy Consumption in Homes by End Users. http://www.eia.gov/consumption/residential, March 2013.

[16] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings. Agent-based Micro-storage Management for the Smart Grid. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2010.

[17] R. Weron. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. HSC Books. Hugo Steinhaus Center, Wroclaw University of Technology, 2006.

[18] J.B. Rawlings and D.Q. Mayne. *Model Predictive Control: Theory and Design*. Nob Hill Publishing, Madison WI, 2009.