A data–based approach to power capacity optimization

Giorgio Manganini[1], Alessandro Falsone[2], Jan Siroky[3], Maria Prandini[2]

Abstract— We consider a power capacity optimization problem where a consumer has to decide the amount of electrical power capacity to purchase for the following year, which includes an amount that is constant over the year (yearly capacity), and an additional surplus per month (monthly capacity). The cost per power unit of the yearly capacity is lower than that of the monthly capacity. A high violation cost is paid when the actual power consumption in a month exceeds the pre-allocated capacity. Given that future power consumption is uncertain, we propose a solution that consists in the minimization of the average expected cost, which includes also the violation costs. By replacing the average with its empirical mean, we can compute an approximate solution to the original problem with a pre-defined level of accuracy by extracting a sufficiently large number of power consumption realizations, which is here set via the uniform convergence of empirical means theory. Extractions are obtained based on a stochastic model that is built from available historical data. The effectiveness of the approach is shown on a real case study.

I. INTRODUCTION

Many EU states have established the Capacity Remuneration Mechanisms (CRMs) in their electricity markets. These mechanisms aim to ensure that the electricity system has sufficient capacity available to meet the electricity demand at all times, and they depend on medium- to long-term projections of peak demand and the long-term projection of supply. Reducing peak electricity usage and limiting the necessity of expensive marginal capacity plans has long been a goal of the power sector [1], [2].

Capacity includes generation and other energy supply sources such as storage or demand response [3]. Efforts to try to encourage demand-side response are growing, see for example [4]. Tariff-based economic instruments tend to be the dominant approach to demand response in the residential sector [5], and incentive-based approaches are used in the commercial sector, particularly in US [3]. Further capacity remuneration mechanisms have been devised [6].

To allow the electrical energy providers to efficiently schedule their production – possibly keeping part of their generation capacity (power plants) in reserve so as to compensare demand peaks, if needed, and ensure stability of the electrical system – big consumers, such as factories, hospitals, sport facilities, shopping malls, business buildings,

are requested to set their capacity reservation each year for the next year. This is the so-called electrical power capacity optimization problem. In this paper, we focus on one particular mechanism that is used in the Czech Republic. The capacity reservation for a customer is defined on a monthly basis and consists of the purchase of a yearly base capacity and a monthly extra capacity. A penalty price is applied when such reserved maximum capacity is exceeded. The usual practice is that the energy specialist responsible for power capacity optimization tends to purchase a higher reserved capacity than needed in order to avoid any penalty. This, however, may result in a costly solution, especially if looking at a long term horizon. The problem is that future power consumption can only be estimated, and the expected violation of the capacity reservation should be appropriately accounted for when deciding the yearly and monthly capacities.

In this work we propose a solution that consists in the minimization of the average expected cost, which includes the planned yearly and monthly capacity costs, as well as the uncertain costs due to violation. Given that such an average cost function cannot be computed exactly, we adopt an approximate solution by replacing the average with its empirical mean. By using results from statistical learning theory [7] and, in particular, the uniform convergence of empirical means, we provide an estimate N of the number of stochastic realizations of the yearly power consumption that are needed to get a solution with a cost that is close to the optimal one of a certain a-priori defined amount. Since N is large, we propose to build first a stochastic model from historical data via a functional data analysis approach, [8], and then extract from it the N future consumption profiles.

II. PROBLEM FORMULATION AND PROPOSED SOLUTION

We consider a scenario in which, at the end of every year, (medium-big) customers must provide its monthly maximum electrical power consumption profile for the next year to the electrical grid operator (or to the electricity provider), so that he can efficiently schedule his production plants. The consumption profile that each customer needs to provide is a twelve months profile composed of two contribution: a minimum electrical power capacity x_y throughout the year and a monthly additional capacity x_m, where $m \in \{1, \ldots, 12\}$ denotes the month of the year.

The cost of buying the total electrical power capacity $x_y + x_m$ for month m is given by $c_y x_y + c_m x_m$, where $c_m > c_y > 0$ for all $m = 1, \ldots, 12$ (note that $c_{m'}$ is not necessarily equal to c_m for $m' \neq m$). If the customer maximum power consumption Q_m is higher than the allocated power capacity

$x_y + x_m$, then the customer will incur an additional cost

$$v(x_y, x_m, Q_m) = \gamma[Q_m - x_y - x_m]_+, \qquad (1)$$

in the bill of month m, where $[\,\cdot\,]_+ = \max\{0, \cdot\}$ and γ is the unitary cost for the non-allotted power consumption, which satisfies $\gamma \gg c_m > c_y$ for all $m = 1, \ldots, 12$. The customer overall power capacity allocation cost is then given by

$$\sum_{m=1}^{12} \big(c_y x_y + c_m x_m + v(x_y, x_m, Q_m)\big). \qquad (2)$$

If the monthly maximum power consumption Q_m is known in advance, then, the power allocation problem that each customer needs to solve consists in choosing x_y and x_m, $m = 1, \ldots, 12$, so as to minimize the overall cost (2), i.e.:

$$\min_{x_y, \{x_m\}_{m=1}^{12}} \sum_{m=1}^{12} \big(c_y x_y + c_m x_m + v(x_y, x_m, Q_m)\big) \qquad (3)$$

$$\text{subject to:} \quad x_y, x_m \geq 0, \quad m = 1, \ldots, 12.$$

Substituting (1) in (2), one can see that the cost function (2) is non-linear but still convex, which means that problem (3) can be efficiently solved. Furthermore, by introducing additional variables h_m, $m = 1, \ldots, 12$, problem (3) can be reformulated as

$$\min_{x_y, \{x_m, h_m\}_{m=1}^{12}} \sum_{m=1}^{12} \big(c_y x_y + c_m x_m\big) + \gamma \sum_{m=1}^{12} h_m \qquad (4)$$

$$\text{subject to:} \quad Q_m - x_y - x_m \leq h_m$$
$$x_y, x_m, h_m \geq 0$$
$$m = 1, \ldots, 12$$

which is a linear program.

Taking a closer look at problem (4), one can notice that $x_m \geq$ and $h_m \geq 0$ have the same role in the constraint $Q_m - x_y - x_m \leq h_m$. However, h_m has a higher contribution than x_m in the cost function since $\gamma > c_m$. Therefore, h_m, $m = 1, \ldots, 12$, will be set to zero when solving problem (4), that is then equivalent to

$$\min_{x_y, \{x_m\}_{m=1}^{12}} \sum_{m=1}^{12} (c_y x_y + c_m x_m) \qquad (5)$$

$$\text{subject to:} \quad Q_m - x_y - x_m \leq 0$$
$$x_y, x_m \geq 0$$
$$m = 1, \ldots, 12.$$

Figure 1 represents a solution to (5) obtained by setting Q_m equal to the maximum power consumption experienced by some customer in month m of year 2014, and then solving problem (5) to compute x_y and x_m for the forthcoming year.

Although this seems a sensible solution, it completely fails to capture the uncertain nature of the power consumptions Q_m, $m = 1, \ldots, 12$, and therefore it cannot provide any guarantees about how much will the customer actually pay for the electricity the next year.

Indeed, Q_m appearing in problem (3) is an uncertain quantity and hence problem (3) is ill-posed. If we now suppose
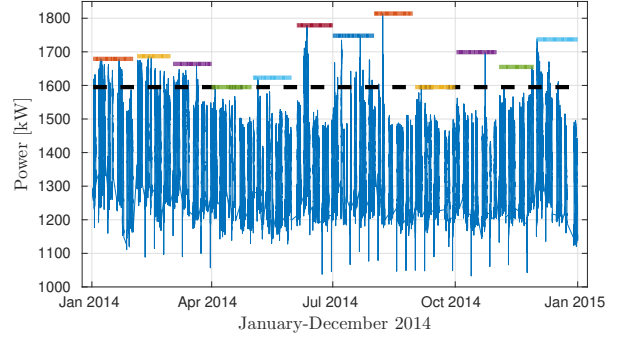

Fig. 1. Power consumption profile experienced during the year 2014 by some costumer (blue line) and optimal values of x_y (dashed line) and x_m (colored solid lines) for problem (5) when Q_m is set equal to the maximum of the blue line within month m. Data are sampled every 15 minutes.

that the power consumption profiles are stochastic, we can take into account the uncertainty on Q_m by considering the average of the cost in (3), which leads to the following optimization program

$$\min_{x_y, \{x_m\}_{m=1}^{12}} \mathcal{C} + \sum_{m=1}^{12} \mathbb{E}[v(x_y, x_m, Q_m)] \qquad (6)$$

$$\text{subject to:} \quad x_y, x_m \geq 0, \quad m = 1, \ldots, 12,$$

where we set $\mathcal{C} = \sum_{m=1}^{12} \big(c_y x_y + c_m x_m\big)$.

Problem (6) cannot be solved exactly when the probability distribution of Q_m is not known. And even in the case when it is known, obtaining an analytical description of $\mathbb{E}[v(x_y, x_m, Q_m)]$ as a function of x_y and x_m is generally not trivial. In this paper we propose to use extractions of Q_m to approximate $\mathbb{E}[v(x_y, x_m, Q_m)]$ up to a desired level of accuracy. Specifically, we consider the following sample program

$$\min_{x_y, \{x_m\}_{m=1}^{12}} \mathcal{C} + \sum_{m=1}^{12} \frac{1}{N} \sum_{i=1}^{N} v(x_y, x_m, Q_m^{(i)}) \qquad (7)$$

$$\text{subject to:} \quad x_y, x_m \geq 0, \quad m = 1, \ldots, 12,$$

where $Q_m^{(1)}, \ldots, Q_m^{(N)}$ are N independent samples of Q_m, $m = 1, \ldots, 12$.

Let $J(x)$ with $x = [x_y \; x_1 \; \cdots \; x_{12}]^\top$ denote the cost function of (6) and let $x^\star$ and $\hat{x}^\star$ be the solutions to (6) and (7), respectively.

Sampling methods for estimating the mean of a function of a random quantity are very well-known and popular in the engineering practice. Note that $\mathbb{E}[v(x_y, x_m, Q_m)]$ that we are estimating depends on the variables x_y and x_m that we want to optimize. We then need a uniformly convergent approximation of $\mathbb{E}[v(x_y, x_m, Q_m)]$ to obtain a solution $\hat{x}^\star$ whose cost $J(\hat{x}^\star)$ is close to the optimal cost $J(x^\star)$.

Resorting on the notion of uniform convergence of empirical mean we can state the following result.

Theorem 1: Select $\varepsilon \in (0, e/(2\log_2 e))$ and a confidence parameter $\delta \in (0, 1)$. If N is set so as to satisfy

$$N \geq \frac{4608}{\varepsilon^2} \left[\ln \frac{96}{\delta} + d \ln \frac{192e}{\varepsilon} + d \ln \ln \frac{192e}{\varepsilon} \right] \qquad (8)$$

with $d = 4\log_2(4e)$, then

$$\mathbb{P}\left\{\{Q_m^{(i)}\}_{i=1}^{N_m} : J(\hat{x}^\star) < J(x^\star) + 2\varepsilon\gamma\mathcal{Q}\right\} \geq 1 - \delta, \quad (9)$$

where $\mathcal{Q} \in \mathbb{R}$ is such that $Q_m \leq \mathcal{Q}$ for all $m = 1,\ldots,12$. ∎

Theorem 1 ensures that, if the number of realization is appropriate, then the average cost that the customer will pay applying $\hat{x}^\star$ will not exceed the optimal one by more than a $2\varepsilon\gamma\mathcal{Q}$ amount, with a certain confidence $1 - \delta$. Note that both the quality of the results and the confidence level are design parameters that can be arbitrarily set. The higher is the required accuracy and confidence (low ε and δ), the larger will be N. In particular, N scales logarithmically with δ which can be set small enough (say $\delta = 10^{-6}$) so as to make the statement on the accuracy practically deterministic without increasing too much N.

III. PROOF OF THEOREM 1

Knowing that $Q_m \leq \mathcal{Q}$ and by the positivity constraints in (6), we can safely assume $(x_y, x_m) \in X = [0, \mathcal{Q}]^2$ for all $m = 1,\ldots,12$. Denote as

$$\mathcal{F}_m = \left\{f_m : f_m(Q_m) = \frac{v(x_y, x_m, Q_m)}{\gamma\mathcal{Q}}\right\}_{(x_y, x_m) \in X} \quad (10)$$

the family of functions for which we want to construct an approximation of $\mathbb{E}[f_m(Q_m)]$ by sampling. Note that, by definition, $f_m : [0, \mathcal{Q}] \to [0, 1]$. Let $Q_m^{(i)}$, $i = 1,\ldots,N_m$ be independent samples extracted from the distribution of Q_m and define

$$\hat{\mathbb{E}}_{N_m}[f_m] = \frac{1}{N_m} \sum_{i=1}^{N_m} f_m(Q_m^{(i)}) \quad (11)$$

the sample counterpart of $\mathbb{E}[f_m(Q_m)]$. The family $\mathcal{F}_m$ is said to have the property of Uniform Convergence of Empirical Means (UCEM) to their true values if, for any $\varepsilon > 0$,

$$\mathbb{P}\left\{\{Q_m^{(i)}\}_{i=1}^{N_m} : \sup_{f_m \in \mathcal{F}_m} \left|\hat{\mathbb{E}}_{N_m}[f_m] - \mathbb{E}[f_m(Q_m)]\right| > \varepsilon\right\} \quad (12)$$

tends to 0 as $N_m \to \infty$. This is indeed the case if the family of functions $\mathcal{F}_m$ has finite Pollard-dimension, [7]. In fact, for any $\varepsilon < e/(2\log_2 e)$, (12) can be upper bounded by

$$8\left(\frac{16e}{\varepsilon}\ln\frac{16e}{\varepsilon}\right)^d e^{-N_m\varepsilon^2/32}, \quad (13)$$

where d is the Pollard-dimension of $\mathcal{F}_m$, [7]. Note that the result holds irrespectively of the distribution of Q_m. The problem then boils down to find the Pollard-dimension of $\mathcal{F}_m$. To this purpose, we need to associate to $\mathcal{F}_m$ a collection $\mathcal{G}_m$ of indicator functions. Let $c \in [0, 1]$ and define the function

$$\tau_m(Q_m, c, x_y, x_m) = \frac{Q_m - x_y - x_m}{\mathcal{Q}} - c, \quad (14)$$

the set

$$A_m = \{(Q_m, c, x_y, x_m) : \tau_m(Q_m, c, x_y, x_m) \geq 0\}, \quad (15)$$

and the family

$$\mathcal{G}_m = \{\mathcal{I}_{A_m}(Q_m, c, x_y, x_m)\}_{(x_y, x_m) \in X}, \quad (16)$$

$\mathcal{I}_{A_m}(\cdot)$ being the indicator function on the set A_m. Moreover, $\mathcal{G}_m$ can be equivalently described as

$$\mathcal{G}_m = \{H(f_m(Q_m) - c)\}_{f_m \in \mathcal{F}_m}, \quad (17)$$

$H(z)$ being the Heaviside function, i.e. $H(z) = 1$ when $z \geq 0$ and $H(z) = 0$ otherwise. To see the equivalence between (16) and (17) it suffice to notice that

$$H(f_m(Q_m) - c) = 1 \iff f_m(Q_m) \geq c$$
$$\iff \frac{[Q_m - x_y - x_m]_+}{\mathcal{Q}} \geq c$$
$$\iff \frac{Q_m - x_y - x_m}{\mathcal{Q}} \geq c$$
$$\iff \tau_m(Q_m, c, x_y, x_m) \geq 0$$
$$\iff \mathcal{I}_{A_m}(Q_m, c, x_y, x_m) = 1,$$

where the second implication is due to the definitions of $f_m(Q_m)$ and $v(Q_m, x_y, x_m)$ in (10) and (1), respectively, the third implication is due to the fact that $c \geq 0$, and the fourth and fifth implications come from (14) and (15), respectively.

Thanks to [7, Lemma 10.1] we can link the Pollard-dimension of $\mathcal{F}_m$ with the Vapnik Chervonenskis (VC)-dimension of $\mathcal{G}_m$ as expressed in (17). Moreover, from [9], [10] we know that the VC-dimension of $\mathcal{G}_m$ as expressed in (16) is upper bounded by $4\log_2(4e)$. Therefore (12) can be upper bounded as

$$8\left(\frac{16e}{\varepsilon}\ln\frac{16e}{\varepsilon}\right)^{4\log_2(4e)} e^{-N_m\varepsilon^2/32}. \quad (18)$$

Equation (18) ensures that (12) tends to 0 as $N_m \to \infty$ for one month m. To get a relative error smaller than ε for $\mathbb{E}\left[\sum_{m=1}^{12} v(x_y, x_m, Q_m)\right]$, which is the quantity appearing in our cost function in (6), we need to consider the following chain of inequalities

$$\mathbb{P}\left\{\sup_{\substack{f_m \in \mathcal{F}_m \\ m=1,\ldots,12}} \left|\sum_{m=1}^{12} \hat{\mathbb{E}}_{N_m}[f_m] - \mathbb{E}[f_m(Q_m)]\right| > \varepsilon\right\}$$
$$\leq \mathbb{P}\left\{\sum_{m=1}^{12} \sup_{f_m \in \mathcal{F}_m} \left|\hat{\mathbb{E}}_{N_m}[f_m] - \mathbb{E}[f_m(Q_m)]\right| > \varepsilon\right\}$$
$$\leq \sum_{m=1}^{12} \mathbb{P}\left\{\sup_{f_m \in \mathcal{F}_m} \left|\hat{\mathbb{E}}_{N_m}[f_m] - \mathbb{E}[f_m(Q_m)]\right| > \frac{\varepsilon}{12}\right\}$$
$$\leq \sum_{m=1}^{12} 8\left(\frac{192e}{\varepsilon}\ln\frac{192e}{\varepsilon}\right)^{4\log_2(4e)} e^{-N_m\varepsilon^2/4608}. \quad (19)$$

Setting $N_m = N$ for all $m = 1,\ldots,12$, (19) can be easily inverted to find N. In fact, if we choose N so as to satisfy (8), then

$$\mathbb{P}\left\{\sup_{\substack{f_m \in \mathcal{F}_m \\ m=1,\ldots,12}} \left|\sum_{m=1}^{12} \hat{\mathbb{E}}_N[f_m] - \mathbb{E}[f_m(Q_m)]\right| > \varepsilon\right\} \leq \delta. \quad (20)$$

Multiplying by $\gamma\mathcal{Q}$ both sides inside the probability in (20), adding and subtracting $\mathcal{C}$ inside the absolute value, and considering the complementary event, we get

$$\mathbb{P}\left\{\sup_x \left|\hat{J}(x) - J(x)\right| \leq \varepsilon\gamma\mathcal{Q}\right\} \geq 1 - \delta, \qquad (21)$$

where $\hat{J}(x)$ denote the cost function of (7). By definition of minimizer, $\hat{x}^\star$ satisfies

$$\hat{J}(\hat{x}^\star) \leq \hat{J}(x^\star) = \hat{J}(x^\star) - J(x^\star) + J(x^\star), \qquad (22)$$

and, by (21), we have

$$\mathbb{P}\left\{J(\hat{x}^\star) - \hat{J}(\hat{x}^\star) \leq \varepsilon\gamma\mathcal{Q} \wedge \hat{J}(x^\star) - J(x^\star) \leq \varepsilon\gamma\mathcal{Q}\right\} \geq 1 - \delta. \qquad (23)$$

Using (22) in (23), (9) follows, thus concluding the proof.

IV. STATISTICAL MODELING

The approach introduced in Section III makes use of a certain number of realization of the maximum power consumption Q_m, which may exceed the number of available power consumption profile historical data. We then need to be able to generate extractions of the power consumption from its probability distribution.

In the common practice, uncertainty on power consumption is often characterized by means of some probability distribution with fixed parameters, which are assumed to be known. However, this assumption is restrictive and unrealistic since the monthly power consumption of a building is determined by several factors and its behavior would be hardly captured by a known distribution. Alternatively, one could model the building and the disturbances affecting its power consumption as in [11], where real consumption processes in office buildings are modeled by composing three modules of occupancy in office rooms, the usage of electric appliances, the operation of the Heating, Ventilation and Air Conditioning (HVAC) system and their mutual interrelationship. This however, requires a strong effort for modeling the system and then identifying its parameters.

We then suggest to adopt a data-based black-box approach to determine a stochastic power consumption model. Specifically, we introduce a statistical tool for building a stochastic model of the random variables Q_m starting from historical power consumption data only. This way, one does not need any expertise for modeling the power consumption mechanisms and can extract realizations of Q_m without any explicit characterization of its distribution.

Functional Data Analysis (FDA) is a (fairly recent) branch of statistics that extends the tools of multivariate statistics to random experiments whose outcomes are functions of a continuous variable, [8]. The typical example of sample element in FDA is the height of a person as a function of its age. In this paper we denote a sample element as $q_m(t)$ and it represents the daily power consumption profile in month m as a function of the time of the day $t \in [0, T]$, with $T = 24$ hours. Clearly $q_m(t)$ is a random quantity that depends on a continuous variable: time t. If we create a model for $q_m(t)$, then a sample $Q_m^{(i)}$ of the random quantity

Q_m we are interested in can be obtained by extracting D_m realizations of daily profiles $q_m^{(d,i)}(t)$, $d = 1, \ldots, D_m$, and computing

$$Q_m^{(i)} = \max_{d=1,\ldots,D_m} \max_{t \in [0,T]} q_m^{(d,i)}(t), \qquad (24)$$

where D_m denotes the number of days in month m.

Our goal is therefore to get a compact representation of $q_m(t)$ based on historical data so that we can sample realizations of Q_m based on (24). To this end, we employ the Functional Principal Component Analysis (FPCA) technique, [8, Chapter 8]. FPCA is the functional counterpart of the well-known principal component analysis (PCA) in the multivariate statistics literature.

Let $\{q_m^{(i)}(t), i = 1, \ldots, N_m, t \in [0; T]\}$ be our collection of historical daily profiles in the m-th month. We neglect the day index because days of the same month are considered as different realizations of the same random quantity that is the daily power consumption in month m. If, for example, $m = 1$, then we have 31 daily profiles for each year in the dataset. Denote with $\langle f, g \rangle = \int_0^T f(s)g(s)ds$ the inner product of two functions $f(\cdot)$ and $g(\cdot)$ on $[0, T]$ and with $\mathcal{L}^2([0, T])$ the space of functions $f(\cdot)$ on $[0, T]$ satisfying $\|f\|_2^2 = \langle f, f \rangle < \infty$. Let $\bar{q}_m(t) = \frac{1}{N_m}\sum_{i=1}^{N_m} q_m^{(i)}(t)$ be the average daily power consumption profile during month m and $\mathcal{S} : \mathcal{L}^2([0, T]) \to \mathcal{L}^2([0, T])$

$$\mathcal{S}\xi(t) = \frac{1}{N_m}\sum_{i=1}^{N_m}\left[(q_m^{(i)}(t) - \bar{q}_m(t))\int_0^T(q_m^{(i)}(s) - \bar{q}_m(s))\xi(s)ds\right]$$

the sample variance operator.

The k-th (functional) principal component for the daily profile of the m-th month $\xi_{m,k}(t)$ is defined as the solution of the following optimization program

$$\max_{\xi \neq 0} \frac{\langle \xi, \mathcal{S}\xi \rangle}{\langle \xi, \xi \rangle} \qquad (25)$$
$$\text{subject to:} \quad \langle \xi_{m,s}, \xi \rangle = 0, \quad s = 1, \ldots, k - 1.$$

Solving (25) amounts to find the eigenfunction associated with the k-th largest eigenvalue of the $\mathcal{S}$ operator. Note that in (25) the dependence of all quantities from t has been omitted for ease of notation.

Note that in (25) the k-th principal component is required to be orthogonal to the previous $k-1$ ones. This ensures that $\xi_{m,1}(t), \ldots, \xi_{m,k}(t)$ is an orthonormal basis of a subspace of the $\mathcal{L}^2([0, T])$ space. We can therefore approximately represent our data points as a linear combination of such basis functions, i.e.

$$q_m^{(i)}(t) \approx \bar{q}_m(t) + \sum_{k=1}^{K} \vartheta_{m,k}^{(i)}\xi_{m,k}(t), \qquad (26)$$

where the number $K \in \{1, \ldots, N_m - 1\}$ of basis function is a design parameter and is selected so as to describe a certain percentage of the total variance of the dataset.

From (26) it is clear that every datum $q_m^{(i)}(t)$ is associated to a set of coefficients $\boldsymbol{\vartheta}_m^{(i)} = [\vartheta_{m,1}^{(i)}, \ldots, \vartheta_{m,K}^{(i)}]^\top$. An important aspect of the (F)PCA is that it chooses the

Fig. 3. Historical (red circles) and generated (box plots) monthly maximum power consumption.
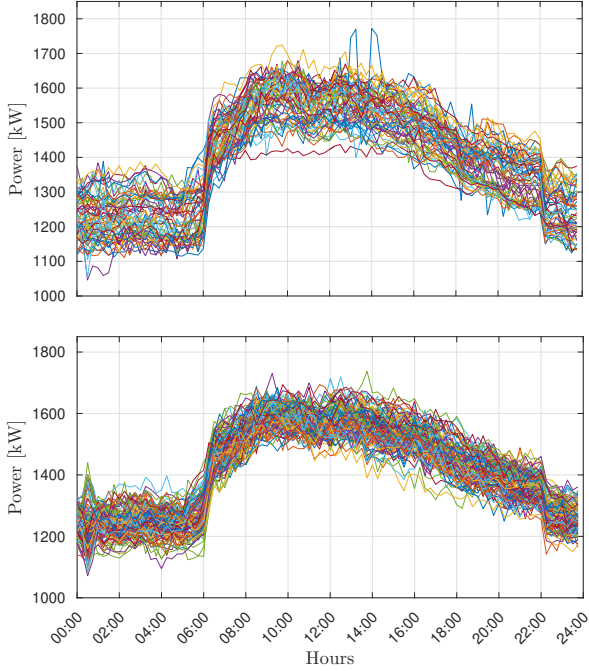

Fig. 2. January daily profiles: historical (top) and artificial (bottom).

basis functions so as to make $\vartheta_m^{(i)}$ as close as possible to samples extracted to a K-variate Gaussian distribution with independent components. This latter point can be exploited to generate a lot more realizations of Q_m. Specifically, letting μ_m and Σ_m be respectively the sample mean and covariance matrix estimated from $\vartheta_m^{(i)}$, $i = 1, \ldots, N_m$, we can draw new samples $\vartheta_m^{(j)}$ from $\mathcal{N}(\mu_m, \Sigma_m)$, $\mathcal{N}(\mu_m, \Sigma_m)$ denoting the K-variate Gaussian distribution with mean μ_m and covariance matrix Σ_m, then use (26) to get a sample from $q_m(t)$ and, finally, a sample from Q_m using (24).

Note that the dependence of ξ from t in (25) makes (25) an infinite dimensional program. To actually solve (25) one can either approximate each datum $q_m^{(i)}(t)$, $i = 1, \ldots, N_m$, with a linear combination of basis functions, *e.g.* B-spline, Fourier, Wavelet, etc. or suitably discretize the continuous variable t. In this paper we adopt this latter approach to compute the first K principal components $\xi_{m,1}(t_j), \ldots, \xi_{m,K}(t_j)$ for all $m = 1, \ldots, 12$, where $t_j = (1/4)j$ denotes the sampling instants within the interval $[0, T]$, with $j = 0, \ldots, 96$ (i.e., the power consumption is sampled every quarter hour).

In Figure 2 we report some historical daily profiles for January from our dataset (top), and some profiles generated using the FPCA procedure for the same month (bottom). As it can be seen from the picture the FPCA procedure is able to generate realistic profiles of the daily power consumption.

V. NUMERICAL RESULTS

The proposed approach has been applied to a real case study of a building in Prague.

The following parameters have been employed in the numerical experiments. The yearly and monthly cost for purchasing power capacity are, respectively, $c_y = 159,183 \ [CZK/kW]$ and $c_m = 175,908 \ [CZK/kW]$, for
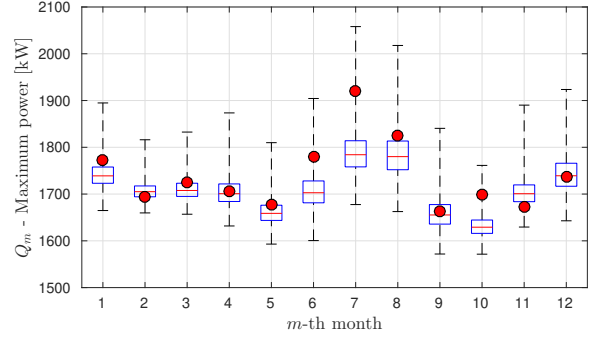
all $m = 1, \ldots, 12$, while the unitary cost for power violation is $\gamma = 636,732 \ [CZK/kW]$ (with a ratio cost/violation of $\gamma/c_y = 4$). A data-base of historical power-profiles spanning three years (2014 to 2016) of measurement has been employed for the statistical modeling phase. $K = 35$ principal components $\xi_{m,1}(\cdot), \ldots, \xi_{m,K}(\cdot)$ for all $m = 1, \ldots, 12$ have been considered in the FPCA procedure in order to generate N new realizations $\{Q_m^{(i)}\}_{i=1}^N$. According to Theorem 1, setting $\varepsilon = 0.1$ and $\delta = 10^{-3}$ in (8) gives an $N = 7.3 \cdot 10^7$. Unfortunately, this was beyond our memory capabilities, hence we used a reduced number of $N = 10^5$ samples to solve (7), and we here present the obtained results.

Statistical properties of the realizations of the maximum power consumption drawn as discussed in Section IV are summarized in Figure 3. Red circles represent the maximum power consumption in m-th month based on historical measurements available in the dataset, and box plots describe the empirical distribution of corresponding generated realizations: red horizontal bars denote the median, blue boxes represent the 25-th and 75-th percentiles, and whiskers represent minimum and maximum values. It is interestingly to notice that most of historical power picks lie in the high probable region of the identified distribution, and all months are contained between the minimum and maximum values. Figure 3 can be therefore considered as an empirical evidence of the capabilities of the FPCA in characterizing the distribution of Q_m.

Based on these generated power profiles, problem (7) is solved so as to obtain the optimal solution for the yearly and monthly power capacity c_y and c_m for the next year power capacity planning. As the reader can notice by comparing the "average" solution in Figure 4 with the "trivial" one in Figure 1, the uncertainty characterizing the next year power consumption, modeled as the random variable Q_m, led to a more conservative solution than considering it as a deterministic quantity, which cannot provide any insight about the possible future power violation cost.

A detailed analysis on how the violation cost γ might change the optimal solution of the problem (7) has been conducted, and the results are showed in Figures 5 and 6 for different values of the penalization factor γ/c_y.

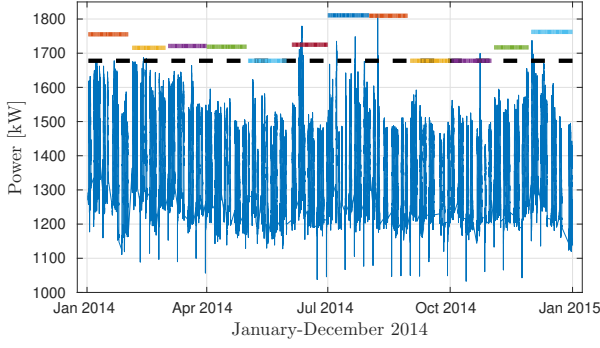As long as the γ/c_y increases, the optimal values of $\hat{x}_y^\star$

Fig. 4. Historical power profile (blue line) and optimal values of $\hat{x}_y^\star$ (dashed line) and $\hat{x}_y^\star + \hat{x}_m^\star$ (colored lines) for problem (7).

Fig. 5. Influence of the penalization factor γ/c_y in (7) on the purchased power capacity $\hat{x}_y^\star + \hat{x}_m^\star$.

and $\hat{x}_m^\star$ suggest to buy additional power capacity, since it is better not to incur in a likely violation event which would be associated to a very high violation cost. On the contrary, the more the penalization factor decreases, the lesser power capacity needs to be allocate and purchased, given that a possible violation would not correspond to a high violation cost. These dynamics can be clearly observed in Figure 5.

By changing the violation cost, we do not only change the amount of purchased electrical capacity, but also the probability of a violation event. Given a solution $\hat{x}^\star$ and a yearly realization $Q = \{Q_m\}_{m=1}^{12}$ of the power consumption, let us compactly represent the violation cost when exciting the reserved capacity as

$$V(\hat{x}^\star, Q) = \sum_{m=1}^{12} v(\hat{x}_y^\star, \hat{x}_m^\star, Q_m).$$

Then we have a violation if $V(\hat{x}^\star, Q) > 0$. Figure 6 shows the connection between the effective cost of purchasing the electrical power $\mathcal{C}(\hat{x}^\star)$ in (7) and the probability of violation $\hat{\mathbb{P}}_N[V(\hat{x}^\star, Q) > 0]$ estimated using N realizations of Q for validation. This probability monotonically decreases as the penalization factor increases, because a higher penalization cost causes an augmented risk associated to an event of violation, which can be only mitigated by reducing the frequency, that is the probability, of such an event. Consequently, the effective cost for power purchase increases as the penalization factor increases, meaning that it is convenient to buy more and more electrical power capacity (this minimizes the event of a violation), and the value of the objective function decreases since the cost associated to a violation decreases.

VI. CONCLUSIONS

We addressed power capacity optimization via a data-based approach that rests on the integration of the UCEM theory and the FPCA technique. The approach appears promising. However, the obtained bound on the number of samples needed for a uniformly accurate estimate of the mean cost function appears too conservative and alternative solutions should be conceived.
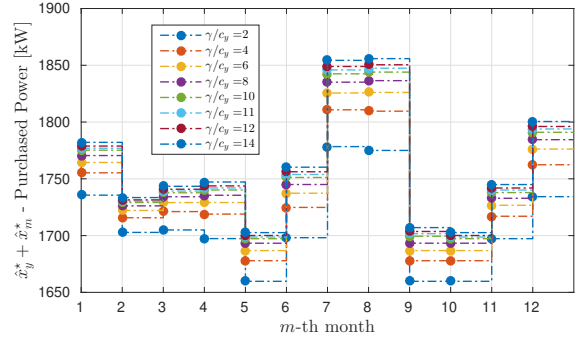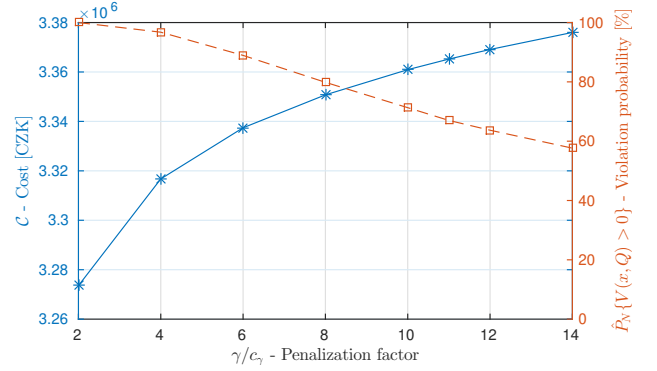

Fig. 6. Trade-off between the cost of purchasing electrical power (without considering any penalization) and the probability of violation.

REFERENCES

[1] A. Faruqui, D. Harris, and R. Hledik, "Unlocking the 53 billion savings from smart meters in the eu: How increasing the adoption of dynamic tariffs could make or break the eus smart grid investment," *Energy Policy*, vol. 38, no. 10, pp. 6222–6231, 2010.

[2] A. Faruqui, S. Sergici, and A. Sharif, "The impact of informational feedback on energy consumptiona survey of the experimental evidence," *Energy*, vol. 35, no. 4, pp. 1598–1608, 2010.

[3] M. H. Albadi and E. El-Saadany, "A summary of demand response in electricity markets," *Electric power systems research*, vol. 78, no. 11, pp. 1989–1996, 2008.

[4] Y. Liu, J. T. Holzer, and M. C. Ferris, "Extending the bidding format to promote demand response," *Energy Policy*, vol. 86, pp. 82–92, 2015.

[5] R. Daniell, "Creating the right environment for demand side response: Next steps," *Office of Gas and Electricity Markets (OFGEM)*, vol. 16, 2013.

[6] P. Bradley, A. Coke, and M. Leach, "Financial incentive approaches for reducing peak electricity demand, experience from pilot trials with a uk energy provider," *Energy Policy*, vol. 98, pp. 108–120, 2016.

[7] V. Mathukumalli, *A Theory of Learning and Generalization: With Applications to Neural Networks and Control Systems*. Springer-Verlag Berlin, 1997.

[8] J. O. Ramsay, *Functional data analysis*. Wiley Online Library, 2006.

[9] M. Karpinski and A. Macintyre, "Polynomial bounds for vc dimension of sigmoidal neural networks," in *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, 1995, pp. 200–208.

[10] ——, "Polynomial bounds for vc dimension of sigmoidal and general pfaffian neural networks," *Journal of Computer and System Sciences*, vol. 54, no. 1, pp. 169–176, 1997.

[11] C. Sandels, D. Brodn, J. Widn, L. Nordstrm, and E. Andersson, "Modeling office building consumer load with a combined physical and behavioral approach: Simulation and validation," *Applied Energy*, vol. 162, pp. 472 – 485, 2016.