

Received June 29, 2020, accepted July 9, 2020, date of publication July 13, 2020, date of current version July 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009060

Simulating Tariff Impact in Electrical Energy Consumption Profiles With Conditional Variational Autoencoders

MARGAUX BRÉGÈRE^{1,2,3} AND RICARDO J. BESSA⁴, (Senior Member, IEEE)

¹EDF Research and Development, 91120 Palaiseau, France

²Laboratoire de mathématiques d'Orsay, Université Paris-Sud, CNRS, Université Paris-Saclay, 91400 Orsay, France

³INRIA - Département d'Informatique de l'École Normale Supérieure, PSL Research University, 75006 Paris, France

⁴INESC Technology and Science (INESC TEC), 4200-465 Porto, Portugal

Corresponding author: Ricardo J. Bessa (ricardo.j.bessa@inesctec.pt)

This work was supported in part by the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme, and in part by the National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia, within project ESGRIDS – Desenvolvimento Sustentável da Rede Elétrica Inteligente/SAICTPAC/0004/2015-POCI-01-0145-FEDER-016434.

ABSTRACT The implementation of efficient demand response (DR) programs for household electricity consumption would benefit from data-driven methods capable of simulating the impact of different tariffs schemes. This paper proposes a novel method based on conditional variational autoencoders (CVAE) to generate, from an electricity tariff profile combined with weather and calendar variables, daily consumption profiles of consumers segmented in different clusters. First, a large set of consumers is gathered into clusters according to their consumption behavior and price-responsiveness. The clustering method is based on a causality model that measures the effect of a specific tariff on the consumption level. Then, daily electrical energy consumption profiles are generated for each cluster with CVAE. This non-parametric approach is compared to a semi-parametric data generator based on generalized additive models. Experiments in a publicly available data set show that, the proposed method presents comparable performance to the semi-parametric one when it comes to generating the average value of the original data (13% difference in root mean square error). The main contribution from this new method is the capacity to reproduce rebound and side effects in the generated consumption profiles. Indeed, the application of a special electricity tariff over a time window may also affect consumption outside this time window. Another contribution is that the proposed clustering approach is capturing the reaction to a tariff change. When compared to a clustering method with classical features (min, max and average consumption), the improvement in the Calinski-Harabasz index was 128% for consumers associated with tariff changes.

INDEX TERMS Deep learning, clustering, simulation, demand response, smart grids, energy consumption.

NOMENCLATURE

i	Household, with \mathcal{I} an household set
Y	Power consumption half-hourly profile
p	Tariff, with $p \in \mathcal{P} = \{\text{Low, Normal, High}\}$
τ	Temperature
h	Half hour
H	Number of half hours in a day
w	Type-of-day
κ	Position-inside-the-year
t, s	Days
T	Number of days in the entire data set
T_0	Number of days in the training set
k	Number of household clusters (C_1, \dots, C_k)
x, X	Vectors of exogenous variables
μ	Standard deviation of the power consumption

σ	Mean of the power consumption
Z	Variable of CVAE latent space (decoder inputs)
d	Dimension of CVAE latent space
Σ	Covariance matrix
E	Noise vector
α	Effect of w in the semi-parametric generator
ξ	Effect of p in the semi-parametric generator
N	Number of generated samples

I. INTRODUCTION

The deployment of smart meters, which provides access to new sources of information like 5-15 minutes resolution electrical energy consumption, makes it possible to envisage the development of new customers services [1]. For example, electricity demand response (DR) policies aim at modifying customers' energy consumption behavior (see [2] for an overview) to enable higher integration levels of renewable energy sources.

The associate editor coordinating the review of this manuscript and approving it for publication was Elizete Maria Lourenco.

Most of these DR schemes rely on changes in electricity prices, which can take the form of seasonal tariffs, super-peak time-of-use, real-time pricing, critical peak pricing, etc. [3]. Recent works (see among others [4] and [5]) proposed online learning algorithms to optimize these price incentives, considering human preferences and satisfaction level. However, the responsiveness to a tariff change may change from a consumer to another. By clustering consumers according to their tariff responsiveness, an electricity supplier can send different signals depending on the cluster to which they belong, and further improve DR management. For instance, for a given temperature, day of the week, etc., the electricity supplier defines an hourly electricity tariff profile to send to some consumers clusters.

A energy consumption data simulator is very useful to conduct an *ex-ante* assessment of the algorithms that set tariff profiles (i.e., ensure that they induce the right behavior from consumers) or to study the business models of different DR models [6] or to implement data-driven DR strategies such as contextual bandit [5]. This simulator should be able to randomly generate energy consumption profiles for different combinations of exogenous variables and tariff profiles, with consumers clustered according to their tariff responsiveness. The present paper proposes a novel method, based on conditional variational autoencoders (CVAE), which aims to randomly generate daily energy consumption profiles conditioned by a specific electricity tariff combined with weather and calendar variables.

The remainder of this paper is organized as follows. Section II conducts a literature review of the clustering and data generation methods applied in the energy domain and identifies the main contributions from this work. In section III, the data set used throughout the rest of the paper is presented. The structure of our contribution is to first provide a clustering method, in Section IV. Then, the CVAE approach used to generate energy consumption profiles is presented and discussed in Section V. In order to evaluate the proposed method, Section VI introduces a benchmark data generator based on semi-parametric models often used for energy consumption forecasting. Section VII presents a comparison of the two generators and simulations that illustrate the interest of our approach. Section VIII summarizes the main conclusions and identifies potential for future work.

The reproducibility of this research was ensured by applying the methodology to the open data set ‘‘SmartMeter Energy Consumption Data in London Households’’ from UK Power Networks [7], where price incentives were sent to users via their smart meters, and by making the CVAE code available in a GitHub repository.¹

II. LITERATURE DISCUSSION AND CONTRIBUTIONS

A. CLUSTERING METHODS

Different clustering approaches were already proposed in the literature to segment consumers according to their energy consumption behavior. Generally, they relied on the

construction of individual features from the average/total consumption and demographic factors. With the recent smart meter deployment, individual consumption records at higher temporal resolutions are now available and allow to consider energy consumption time series in consumers segmentation.

Therefore, more complex features may be extracted and used to cluster consumers with classical algorithms. Among others, Chicco *et al.* compared the results obtained by using various unsupervised clustering algorithms (i.e., modified follow-the-leader, hierarchical clustering, *k*-means, fuzzy *k*-means) to group together customers with similar consumption behavior [8]; Le Ray and Pinson proposed an adaptive and recursive clustering method that creates typical load profiles updated with newly collected data [9]; Rodrigues *et al.* described an online hierarchical clustering algorithm, which was applied to cluster energy consumption time series in a load forecasting task [10]; Fidalgo *et al.* described a clustering approach based on simulated annealing that tries to reconcile billing processes that use 15 min meter data and monthly total consumption and derive typical profiles for consumers classes [11]; Sun *et al.* proposed a copula-based mixture model clustering algorithm that captures complex dependency structures present in energy consumption profiles and detects outliers [12].

These clustering methods do not include information about the elasticity of consumers to tariff changes. However, recent research developed mathematical and statistical models for modeling price responsiveness from domestic consumers. Ganesan *et al.* applied a causality model to the Low Carbon London data set in order to rank consumers according to their responsiveness to tariff changes, and outperformed correlation-based metrics [13]. Saez-Gallego and Morales applied inverse optimization to improve the accuracy of load forecasting when aggregating a pool of price-responsive consumers and considering the effect of calendar and weather variables [14]; Le Ray *et al.* applied a clinical testing approach (based on a test and a control group) to assess whether or not loads of households participating in the EcoGrid EU DR program are price-responsive [15]; Mohajeryami *et al.* proposed an economic model to explain the consumption shift between peak and off-peak hours that maximizes customer’s utility function [16].

These works are closely linked to the forecast of consumers reactions to DR policies, but, to our knowledge, were never combined with (or embedded in) clustering techniques for consumer segmentation or used to simulate daily consumption profiles according to their price-responsiveness.

B. DATA GENERATION METHODS

The generation of energy consumption profiles for households is not new and it was already covered by different authors in the literature. Capasso *et al.* proposed a bottom-up approach based on the aggregation of individual appliance consumption in order to produce a household consumption profile [17]. A Monte Carlo simulation model was proposed

¹github.com/MargauxBregere/power_consumption_simulator

to combine behavioral data (home activities, availability at home from each member, etc.) and engineering functions (appliance mode of operation, technological penetration, etc.) with associated probability distributions. Park *et al.* proposed a platform, exploiting SystemC language for event-driven simulation, which simulates the behavior of individual appliances and smart plugs [18]. Both works did not consider weather-dependent appliances (e.g., heating, ventilating and air conditioning - HVAC) or the effect of price signals.

Physically-based models for appliances (including HVAC) are also proposed in [19], combined with heterogeneous Markov chain for activity patterns, to simulate households energy consumption. A similar approach was followed in [20], but using individual appliance consumption data. A set of physical models for appliances are proposed in [21], implemented in MATLAB Simulink, and can simulate optimal on/off decisions of household appliances. Gottwalt *et al.* described a simulation engine for households with two modules: (a) bottom-up approach that generates consumption data for each appliance by combining statistical data about appliance use and resident presence at home; (b) optimization of appliances schedule in order to find the optimal load shift according to time-based tariffs [22]. Iwafune *et al.* proposed a Markov chain Monte Carlo method for simulating electric vehicle driving behaviors, which enables an evaluation of the DR potential when combined with domestic photovoltaic panels [23].

The aforementioned methodologies assume that information about individual appliances (usage patterns, energy consumption, etc.) and behavioral data is available, instead of just using the total household consumption collected by the smart meter. One exception is [24], which describes a methodology based on an elasticity coefficient (approximated by a Gaussian distribution) to estimate indices that characterize the impact of real-time prices in the consumption pattern, such as proportion of maximum load decrease, proportion of peak-valley difference of load decrease, etc. The method consisted in an empirical rule-based calculation of transferred consumption between periods, which was only applied to aggregated consumption of an electric power system and not to households.

C. CONTRIBUTIONS

The major contributions from this paper are described in the following paragraphs.

The CVAE-based generator of daily energy consumption profiles, in contrast to the methods revised in Section II-B, only relies in data collected by smart meters for the total household consumption and exogenous variables such as tariff profile, weather and calendar variables. Compared to [18]–[20], it is fully data-driven and does not require physical models for individual appliances and consumer behavior data.

Moreover, in comparison to [24], the proposed method is non-parametric and estimates changes in consumption profiles by applying a deep learning model without empirical

assumptions about load shifting, showing a high capacity to learn behavioral changes when consumers experience different tariff schemes. In statistical literacy, the proposed method corresponds to sampling random vectors from a given joint density function, which was also explored in the renewable energy forecasting literature to generate temporal trajectories from conditional marginal probability distributions (see [25] and [26] for wind energy trajectories forecast with Gaussian copula and generalized adversarial networks correspondingly). In this work, we are sampling random vectors (i.e., coherent energy consumption profiles) conditioned by tariff, weather and calendar variables. It is important to note that CVAE were recently applied in [27] to learn specific representations for atypical conditions discovery (e.g., holidays) in daily electrical consumption, but not explored for synthetic data generation.

As a complementary contribution, a novel semi-parametric data generator, based on generalized additive models, is proposed as a benchmark model. Its numerical performance highlights the main advantage offered by the CVAE-based approach, which is the capacity to take into account and reproduce the rebound (the fall or rise in consumption shifts to another time of the day when a special tariff is applied over a period) and side (the fall or rise induced by a special tariff lasts longer – for High tariff – or less long – for Low tariff – than the period in which the tariff is actually applied) effects in the generated consumption profiles.

Finally, the proposed clustering methodology extends the clustering algorithm from in [28] in order to include the causal model between tariff and energy consumption. Thus, in contrast to the methods revised in Section II-A, this clustering algorithm gathers consumers according to their (total) consumption behavior and tariff-responsiveness.

III. DATA SET DESCRIPTION AND PREPROCESSING

As a case-study for this work, we consider the open data set published by UK Power Networks and containing energy consumption (in kWh per half-hour) of around 5 000 households throughout 2013 [7]. A sub-group of approximately 1 100 customers was subjected to a dynamic Time of Use (ToU) tariff. The tariff values, among High (67.20 p/kWh), Low (3.99 p/kWh), or Normal (11.76 p/kWh), and the (half-hourly) intervals of day where these prices are applied, were announced day-ahead via the smart meter or text message. All non-ToU customers were on a flat rate tariff of 14.228 p/kWh and we refer to them as Standard (Std) customers. The report of Schofield *et al.* (see [29]) provides a full description of this experimentation and an exhaustive analysis of results.

The data set contains tariffs and energy consumption records, for each client, at half-hourly intervals. Only ToU customers with more than 95% of data available (1 007 clients) are kept and the same number of Std clients are sampled to build a control group. We denote by \mathcal{I}_{ToU} the set of ToU households and by \mathcal{I}_{Std} the set of Std ones. The missing values in the time series were filled by linear

TABLE 1. Summary of the data set division and use for the CVAE and semi-parametric models.

	Training Set (75% of the original data)	Testing Set (remaining data)
Households clustering (Section IV)	✓	✓
CVAE model training (Section V)	✓	
CVAE hyper-parameters calibration (Subsection V-B)		✓
Semi-parametric model training (Section VI)	✓	
Numerical experiments (Section VII)		✓

interpolation, using the previous and next interval records for small gaps and the days preceding and following for longer periods of missing data. Finally, for each household, we also compute the average energy consumption, its minimum, and its maximum as well as the half-hour of the daily peak and of the daily trough, for the hot months (from April to September) and for the cold months (the others).

Since weather has a strong impact on energy consumption, we added half-hourly data points of air temperature in London obtained from hourly public observations [30] by linear interpolation. Thus, for each household $i \in \mathcal{I}_{ToU} \cup \mathcal{I}_{Std}$, for any day t of year 2013, we get three 48-vectors denoted by $Y_t^1(i), \dots, Y_t^{48}(i), p_t^1, \dots, p_t^{48}$, and $\tau_t^1, \dots, \tau_t^{48}$, which are energy consumption profiles, tariff for ToU consumers and temperature respectively. From now on, $H = 48$ represents the number of consumption readings per day. Since a smoothed temperature – that models the thermal inertia of buildings – is likely to improve forecasts (see among others, [31] and [32]), a daily smoothed temperature $\bar{\tau}_t$ is introduced (see Appendix A for further details). Energy consumption also depends on calendar variables such as the type-of-day and position-inside-the-year. Thus, two additional variables were created (see [33] for more details): (i) binary variable w_t that takes 0 on weekends and 1 on week days; (ii) κ_t , a continuous calendar variable modeling the position inside the year, which increases linearly from 0 (on January, 1.) to 1 (on December, 31.).

The final data set (presented in Table 2) contains, for each of the 2 014 households (half Std, half ToU), $T = 365$ observations of the energy consumption, tariff, and temperatures profiles, the smoothed temperature, the type-of-day, and the position-inside-the-year.

This data set is split in two sub-sets: a training set which contains about 75% of the original data – days are randomly sampled from those of 2013 – and a testing set made of the remaining data points. A perfect design of the experiments would require four data sets but the size of the original data led us to exclude this possibility. As the household clustering is a prior knowledge for the creation of the data generators (we create a generator per cluster), the entire data is used to cluster the clients. The (non-parametric and semi-parametric) data generators are optimized on the training set. The testing set is used to calibrate CVAE-based data generators and to choose the best combination of exogenous variables to give in input. Moreover, the best CVAE among several executions of the training process (CVAEs may converges to local minima) is selected thanks to this testing set. Finally, it also permits to compare the two approaches, non-parametric and

semi-parametric, in the experiments of Section VII. To simplify notation, we re-indent the observations of the original data set: observations from 1 to $T_0 = 273$ form the training set, and the ones from $T_0 + 1$ to $T = 365$ form the testing set. The data set division and use is summarized in Table 1.

IV. CLUSTERING OF HOUSEHOLD CONSUMERS

A. CAUSALITY MODEL

To measure the impact of the tariff on the energy consumption, a causality model similar to the one proposed by Ganezan *et al.* (see [13]) is considered. The finite set of available tariff is denoted by $\mathcal{P} = \{\text{Low, High, Normal}\}$ and its cardinal by $|\mathcal{P}|$. For each household and each tariff, a daily profile of the mean and the standard deviation of its energy consumption will be computed. For an household i , at an half-hour h , the random variable $Y^h(i)$ refers to the individual energy consumption of household i . It depends on the chosen tariff $p \in \mathcal{P}$ but also on the exogenous variables gathered in a vector $x^h = (\tau_t^h, \bar{\tau}_t, w_t, \kappa_t)$.

Here, the aim is to estimate, for each tariff p and for each half-hour h , the expectation and the standard deviation of the random variable $Y^h(i) | \mathcal{P} = p$. Thanks to T observations $Y_t^h(i), x_t^h$, and p_t^h , with $t \in \{1, \dots, T\}$, of energy consumption, tariffs, and exogenous variables, respectively, a model that gives, for the tariff p and the exogenous variables x^h , a forecast of the expected consumption at h when tariff p is picked, is trained. In the original model, the authors used kernel regression and then an approach based on bootstrapping to provide an estimation of the standard deviation (see [13] for further details). In this work, for any exogenous variable x_t^h and tariff p_t^h , the random energy consumption $Y_t^h(i)$ is assumed to be Gaussian of mean $\mu_i(x_t^h, p_t^h)$ and standard deviation $\sigma_i(x_t^h, p_t^h)$ and that these mean and standard deviation depend on additive smooth predictors. They are estimated with generalized additive models (GAM), see [34] – full calculations are detailed in Appendix B. Therefore, for any tariff p , the trained model provides these estimations, that are denoted by $\hat{\mu}_i(x_t^h, p)$ and $\hat{\sigma}_i(p, x_t^h)$. Then, an approximation of the impact of a tariff change is computed with the two following quantities:

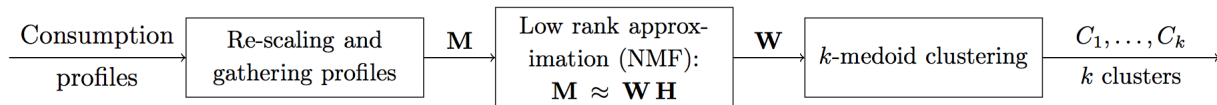
$$\mathbb{E}[Y^h(i) | \mathcal{P} = p] \approx \frac{1}{T} \sum_{t=1}^T \hat{\mu}_i(x_t^h, p) \quad (1)$$

and

$$\sqrt{\text{Var}[Y^h(i) | \mathcal{P} = p]} \approx \frac{1}{T} \sum_{t=1}^T \hat{\sigma}_i(x_t^h, p). \quad (2)$$

TABLE 2. Summary of the variables provided and created for each household i of the data set.

Variable	Description	Notation
Energy consumption	Daily energy consumption profile at half-hourly intervals	$Y_t^1(i), \dots, Y_t^H(i)$
Tariff (for ToU consumers)	Daily electricity price profile at half-hourly intervals	p_t^1, \dots, p_t^H
Temperature	Daily London air temperature profile at half-hourly intervals	$\tau_t^1, \dots, \tau_t^H$
Smooth temperature	Computed from past temperatures	$\bar{\tau}_t$
Type-of-day	1 from Monday to Friday, 0 for weekends	w_t
Position-inside-the-year	Linear value between 0 (January, 1.) and 1 (December, 31.)	κ_t

**FIGURE 1.** Scheme of the proposed clustering method.

For simplicity of notation, these approximations associated with an household $i \in \mathcal{I}_{ToU} \cup \mathcal{I}_{Std}$, are denoted by $\mu_i^h(p)$ and $\sigma_i^h(p)$, respectively. Vectors $\mu_i^1(p), \dots, \mu_i^H(p)$ will be used to cluster the consumers whereas vectors $\sigma_i^1(p), \dots, \sigma_i^H(p)$ will not be used until later, in Section VI for the creation of the benchmark data generator. Actually, they will not be directly useful, but a similar approach will be applied to compute the standard deviation per tariff of the energy consumption of a consumer cluster, namely by replacing household i by a group of households.

B. CLUSTERING METHOD

The proposed method used to cluster the households according to their consumption profile is very similar to the one used in [28]. In this section, \mathcal{I} will refer indifferently to \mathcal{I}_{ToU} or to \mathcal{I}_{Std} . For any household, $i \in \mathcal{I}$, the causality model described in the previous section provides, for each tariff $p \in \mathcal{P}$, a daily energy consumption profile, namely H mean energy consumption $\mu_i^1(p), \dots, \mu_i^H(p)$. As the focus is more on the shape of the profiles, rather than on the amount of consumed electricity, the profiles of an household i are first rescaled with its average consumption associated with a base tariff, namely Normal tariff.

Then, these profiles are concatenated in a matrix $\mathbf{M} \in \mathcal{M}_{|\mathcal{I}| \times H|\mathcal{P}|}$ that gathers all the households. The dimension of \mathbf{M} is reduced with a non-negative matrix factorization (NMF): with r a small integer, \mathbf{M} is approximated by $\mathbf{W}\mathbf{H}$, where \mathbf{W} and \mathbf{H} are $|\mathcal{I}| \times r$ and $r \times H|\mathcal{P}|$ -non-negative matrices, respectively. As soon as this approximation is good enough, line i of the matrix \mathbf{W} is sufficient to reconstruct household i profiles (with the knowledge of matrix \mathbf{H} - which is not used for the clustering). Thus, for each household i , from the $H|\mathcal{P}|$ -vector $(\mu_i^1(p), \dots, \mu_i^H(p))_{p \in \mathcal{P}}$, r features are extracted: line i of \mathbf{W} . With this low dimension representation of households in \mathbb{R}^r , k -medoids clustering algorithm provides the k clusters C_1, \dots, C_k , using `KMedoid` function implemented in the Python-library `sklearn_extra`. The diagram in Figure 1 sums up the steps of the procedure described here in a summarized way and detailed in Appendix C.

C. EVALUATION OF THE HOUSEHOLDS CLUSTERING

Three different clustering approaches of the households of \mathcal{I}_{ToU} and of \mathcal{I}_{Std} , with $k = 4$ clusters, are compared. The first one is a random clustering: an integer between 0 and $k - 1$ is randomly assigned to each household. The second one relies on classical features used to define an household profile: the minimum, maximum, and average consumption in winter and in summer, the peek-hour, and the off-hour (average instant of maximum and minimum consumption). From these rescaled features, k -medoid algorithm is used to cluster the households. The third approach is the one proposed in this paper and described in the previous section. For a cluster C_ℓ , and for any day t and half-hour h , we will, from now on, consider the average energy consumption $Y_t^h(C_\ell) = 1/|C_\ell| \sum_{i \in C_\ell} Y_t^h(i)$, where $Y_t^h(i)$ is the energy consumption record associated with household i .

Figure 2 depicts, for the three clustering approaches applied on ToU households, the weekly profile of the average energy consumption of each cluster $Y_t^h(C_\ell)$ (to the left) and the normalized energy consumption (to the right), namely the weekly profile of $Y_t^h(C_\ell) / (\frac{1}{TH} \sum_{s=1}^T \sum_{j=1}^H Y_s^j(C_\ell))$. Classical features allow to discriminate households depending on the amount of electricity they consume but does not really catch daily or weekly behavior. Conversely, profile types clearly come off with the proposed method.

The Calinski-Harabasz index, (see [35]) is a variance ratio criterion, that evaluate the relevance of the clustering. By denoting $Y(i)$ the vector that contains some of the consumption records associated with household i , and by $Y(C_\ell)$ the one with the average consumption records of cluster C_ℓ and by $Y(\mathcal{I})$ the average consumption records of all households, the score S_{CH} is defined as the ratio of inter-clusters variances and intra-cluster variances:

$$S_{CH} = \frac{(|\mathcal{I}| - k) \text{Var}(C_1, \dots, C_k)}{(k - 1) \sum_{\ell=1}^k \text{Var}(C_\ell)} \quad (3)$$

with

$$\text{Var}(C_1, \dots, C_k) = \sum_{\ell=1}^k \|Y(C_\ell) - Y(\mathcal{I})\|^2 \quad (4)$$

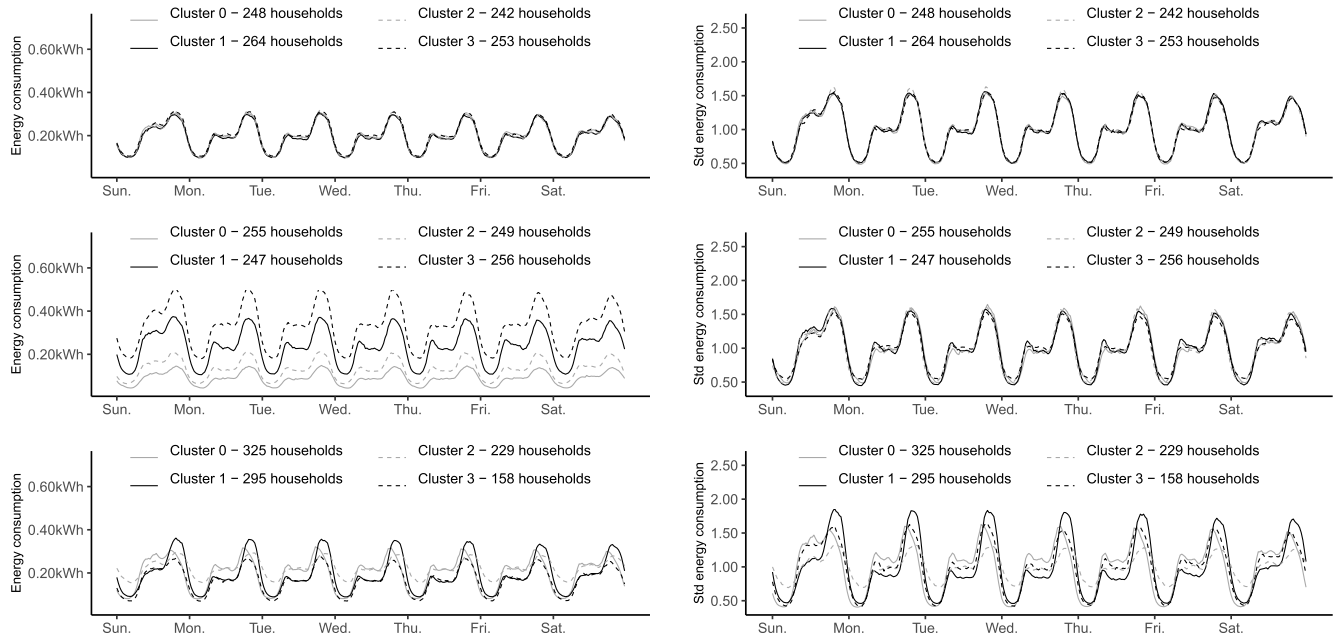


FIGURE 2. Daily profile of ToU cluster energy consumption (left, in kWh) and standardized energy consumption (right) for a random clustering (top), for clustering based on “classical features” (middle), and the clustering method proposed in Section IV (bottom).

TABLE 3. Calinski-Harabasz score for a random clustering (“Rd”), for a clustering based on classical features (“Features”), and the clustering method proposed in Section IV (“NMF”) computed for different consumption record series.

	Rd ToU	Features ToU	NMF ToU	Rd Std	Features Std	NMF Std
Non-standardized	4 627	40 764	13 432	5 014	5 088	12 971
Standardized	4 950	6 870	14 088	4 834	4 741	14 934
Special tariff - standardized	5 151	7 033	16 070	4 904	4 694	15 460

and

$$\text{Var}(C_\ell) = \frac{1}{|C_\ell|} \sum_{i \in C_\ell} \|Y(i) - Y(C_\ell)\|^2. \quad (5)$$

where $\text{Var}(C_\ell)$ is the intra-cluster variance of C_ℓ and $\text{Var}(C_1, \dots, C_k)$ is the inter-clusters variance.

To compute this score, three different vectors Y are considered. First, all the records of the data set are taken into account, namely the records of the entire year 2013; therefore, in Equations (4) and (5), the vector $Y(i)$ is equal to $(Y_1^1(i), Y_1^2(i), \dots, Y_T^H(i))$. Then we look at the normalized energy consumption records, so $Y(i) = (Y_1^1(i), Y_1^2(i), \dots, Y_T^H(i)) / (\frac{1}{TH} \sum_{t=1}^T \sum_{h=1}^H Y_t^h(i))$. Finally the normalized records associated with the sending of incentive signals are selected: only the normalized records associated with tariff Low or High are kept and the others are removed. The results are presented in Table 3, where we observe a higher score on non-standardized records for the “classical features” clustering, which is totally coherent with the curves of Figure 2. The proposed clustering method seems efficient for catching households behavior. Indeed it gets the higher score for standardized records. Moreover, the score is even higher when we select only records associated with special tariff and this increase is more important for ToU

consumers that for Std ones. This presumes that the clustering is not only catching a global behavior but also the reaction to a tariff change.

It is important to mention that since we want to simulate energy consumption of quite large sub-groups of households (between one and five hundreds households), we did not investigated the optimal number of clusters k (i.e., it was fixed to 4).

In the following sections we present the two data-driven methods that simulate energy consumption profiles associated with the clusters of \mathcal{I}_{ToU} obtained with the method described above. For both approaches, we will train a data generator per cluster. So from now on and for simplicity of notation, a record Y_t^h will refer to $Y_t^h(C_\ell)$, where C_ℓ designs any clusters of set \mathcal{I}_{ToU} .

V. ENERGY CONSUMPTION PROFILE GENERATION WITH CONDITIONAL VARIATIONAL AUTOENCODER

The training set made of the T_0 observations $(Y_1, X_1), (Y_2, X_2), \dots, (Y_{T_0}, X_{T_0})$ is considered. For a day t , $Y_t = (Y_t^1, \dots, Y_t^H)$ is the H -dimension vector which corresponds to the daily profile of the half-hour energy consumption of a household cluster. The vector X_t gathers calendar, weather, and tariff information of day t , which will be detailed further.

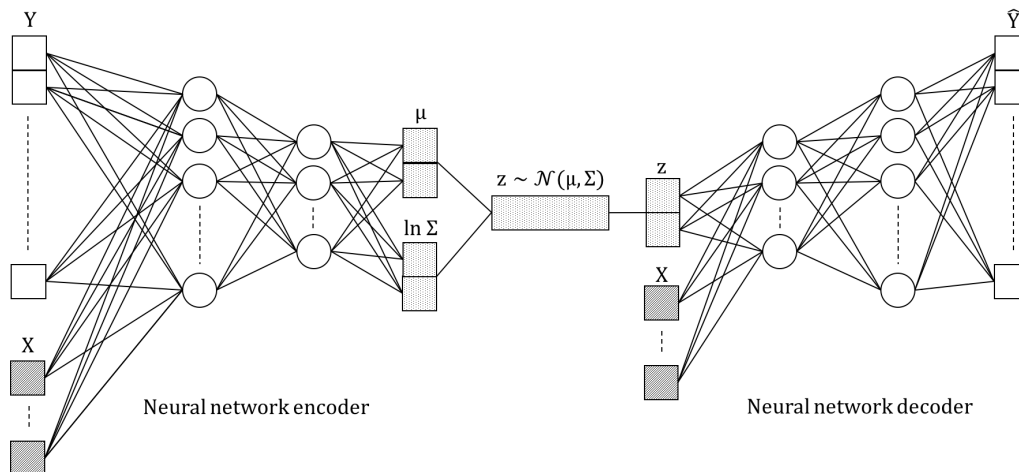


FIGURE 3. Diagram of a conditional variational autoencoder.

A. CONDITIONAL VARIATIONAL AUTOENCODER

1) DESCRIPTION

The proposed method to generate energy consumption profiles uses the conditional version of variational autoencoders (VAE), which are generative models introduced by Kingma and Welling in 2013 (see [36] for further details). It is important to mention that in this work we use generative modeling, instead of discriminative modeling (e.g., support vector machines, k-nearest neighbor), since we are interested in modeling the joint probability $p(Y, X)$ and not the conditional probability $p(Y|X)$. A generative model aims at learning the true distribution of the training data set and sample new data points.

Autoencoders were mostly used for dimensionality reduction or feature learning (see, among others [37] and [38]). They consist of two neural networks: an “encoder” E and a “decoder” D . An autoencoder learns a low dimension representation of a set of H -dimension data points by training both networks at the same time. Indeed, the encoder transforms the H -dimension vectors into d -dimension vectors (with $d \ll H$) and the decoder tries to rebuild initial vectors from the encoder outputs. Considering $Z = E(Y)$ as the d -dimension output of the encoder for the H -dimensional input Y and $D(Z)$ as the H -dimension output of the decoder for the d -dimension input Z , the autoencoder is trained to minimize the following “reconstruction loss”

$$\begin{aligned}
 L_{AE} &= \frac{1}{T_0} \sum_{t=1}^{T_0} \|Y_t - \hat{Y}_t\|^2 \\
 &= \frac{1}{T_0} \sum_{t=1}^{T_0} \|Y_t - D(E(Y_t))\|^2, \quad (6)
 \end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm. Therefore, a data point Y can be represented in a d -dimension latent space by $E(Y)$.

In the autoencoder framework, there is no constraint on this latent space and the only guarantee is that the representation

$Z = E(Y)$ can be decoded in the original signal $D(Z) \approx Y$. Moreover, we have no idea what the decoded variable $D(Z)$ would look like for a value of $Z \notin \{E(Y_1), \dots, E(Y_{T_0})\}$. Thus, there is no guarantee on the shape of the latent space. Without regularization term, for any $d \geq 1$, by increasing the number of neurons in both the encoder and the decoder networks, we can create an autoencoder with enough degrees of freedom to fully overfit the data, which points out the need for a regularization term. In VAEs, the introduction of a penalty on the latent space implicitly makes the strong assumption that the distribution of data points $E(Y)$ is close to a given prior distribution. This prior is often set to the standard normal distribution, which we also do in our experiments. From now on, the encoder encodes the distribution of $Z|Y$, which is wanted close to $\mathcal{N}(0, I_d)$. We consider that $Z|Y \sim \mathcal{N}(\mu(Y), \Sigma(Y))$, where $\mu(Y)$ and $\Sigma(Y)$ are the encoder outputs. The outputs \hat{Y}_t of the decoder are now $D(Z_t)$, where the random variable Z_t is sampled from a d -multivariate Gaussian of mean $\mu(Y_t)$ and covariance matrix $\Sigma(Y_t)$, which are the encoder outputs. With $D_{KL}(P||Q)$ as the Kullback-Leibler divergence from Q to P , the VAE is trained by minimizing the following loss

$$\begin{aligned}
 L_{VAE}(\eta) &= \frac{1}{T_0} \sum_{t=1}^{T_0} \|Y_t - \hat{Y}_t\|^2 \\
 &\quad + \eta \frac{1}{T_0} \sum_{t=1}^{T_0} D_{KL}(\mathcal{N}(\mu(Y_t), \Sigma(Y_t)) || \mathcal{N}(0, I_d)). \quad (7)
 \end{aligned}$$

The first term corresponds to the reconstruction error and the second one is a regularization penalty on the latent space. The coefficient η balances these two terms. Calculations from [36] are recalled in Appendix D. They show how, under some assumptions on the existence of a representation of the data in a d -dimensional latent space, minimizing this loss corresponds to conjointly maximizing the likelihood of the observations with the density induced by the data generation

process and minimizing an approximation error in the latent space.

Finally, conditional variational autoencoders (CVAE) [39] are an extension of VAE where a vector of exogenous variables X is given as input to both the decoder and the encoder. Adding this conditional information may improve the reconstruction. Figure 3 depicts a scheme of the CVAE architecture used in the experiments. The encoder takes as input a daily energy consumption profile Y (so namely a H -vector gathering the half-hourly records of energy consumption) and an exogenous vectors X (with calendar, weather, and tariffs information) and outputs the d -dimension vectors μ and $\ln \Sigma$ (it is usual to consider a log-transformation, see [27]). The vector $\ln \Sigma$ is also of dimension d . Indeed, only the diagonal of the covariance matrix Σ is encoded since both approaches (diagonal and full-matrix) were tested and there was no major difference on the reconstruction loss (obviously the regularization term is higher for a full covariance matrix). Since considering a full-matrix (which is symmetric) increases the dimension of encoders outputs (from $2d$ to $d(d + 1)/2$) and the CVAE converges slower, we decided to keep a diagonal matrix to encode the covariance matrix Σ . The random variable Z is then sampled and given to the decoder as well as the vector of exogenous variables X . Finally, the decoder outputs \hat{Y} .

Once the CVAE is trained, the decoder is isolated and used to generate data. For any day s , it is enough to sample a random variables $Z_s \sim \mathcal{N}(0, I_d)$ in the latent space and give it as input to the decoder, combined with a vector of exogenous variables X_s (that could be taken from the original data set or eventually created). Then, the decoder generates a H -vector \hat{Y}_s that corresponds to a new randomly generated daily consumption profile, for the day s and the contextual variables X_s .

2) IMPLEMENTATION DETAILS

The CVAE were implemented by using the software libraries `Tensorflow` and `Keras` in `Python` programming language. The architecture of a CVAE is defined by the latent dimension d as well as the number of layers and units in encoder and decoder neural networks. We use dense layers which are deeply connected neural network layers. Once the architecture of the CVAEs is set, hyperparameters are chosen: the neural activation functions, the initialization method for neural weights and the parameter η , defined in Equation (7), that balances the two terms of the loss. The choice of the architecture and hyper-parameters calibration is detailed in Section V-B.

In order to optimize a CVAE, so namely to compute weights and bias for each neural of both the encoder and the decoder, the loss is minimized by using the Adam optimizer (see [40]), an extension of stochastic gradient descent method, which is commonly used in deep learning and already implemented in `Keras`. Note that the learning rate of this optimizer is also an hyper-parameter to set before training CVAEs.

Finally, the energy consumption records are rescaled to get values between 0 and 1 by computing the maximum Y_{\max} and minimum Y_{\min} of the energy consumption observed on the train period. The generated value are re-scaled to get coherent profile, mostly between Y_{\min} and Y_{\max} .

We recall that the data described in section III was divided into two data sets: the training set contains 75% for the observations (sampled randomly from the complete data set) and is used to train the CVAE (see Table 1); the testing set, made with the remaining daily observations, is used to calibrate hyper-parameters (see Section V-B). Finally, as CVAE may converge into local minima, many CVAE are trained and the testing set is also used to select the best one (see Section V-D).

B. HYPER-PARAMETERS CALIBRATION

The process described below will be applied for each of the cluster defined in Section IV, for which a half-hourly energy consumption profile for each day of 2013 is available.

1) METHODOLOGY

To perform CVAEs hyperparameter calibration we opt for a grid search approach that is simply an exhaustive searching through a manually specified subset of the hyperparameter space. This optimization is guided by the performance metric detailed below, which is simply an evaluation on a held-out validation set. For each set of parameters, namely for each point of the grid, we train a CVAE and test it according to the procedure described below. Once the CVAEs have converged, (we stop the convergence process when the loss is not decreasing any more), we compute the mean squared error (which corresponds to the reconstruction loss) on the testing set made of the observations Y_{T_0+1}, \dots, Y_T :

$$\text{MSE} = \frac{1}{T - T_0} \sum_{t=T_0+1}^T \|Y_t - \hat{Y}_t\|^2 \quad \text{where} \quad \hat{Y}_t = D(Z_t),$$

with

$$Z_t \sim \mathcal{N}(\mu(Y_t), \Sigma(Y_t)). \quad (8)$$

The architecture and hyperparameters of the CVAE hat reaches to lowest MSE are kept.

2) RESULTS

We tested different values from 1 to 20 for the latent dimension d and reached a final value of 4, which is coherent with the results in [27] for the daily energy consumption in France. Moreover, we also performed a principal component analysis (PCA) on the consumption data and found that 4 components were enough to explain more 80% of the variance in the data. We tested CVAEs with one or two hidden layers of 10, 15, 20 or 25 units per layer and concluded that an architecture with a hidden layer of at least 15 neurons performed much better than smaller architectures. We continued to increase the number of layers or the number of neurons per layer,

but without improvement in the MSE. Moreover, the number of iterations necessary before convergence increased. So we decided to keep a single hidden layer of 15 units for both the encoder and the decoder.

Concerning the activation function of the neurons; rectified linear unit (ReLU), linear, and sigmoid functions were tested and there was no doubt that the best performance was obtained with a ReLU activation function.

For the initialization of the network weights, we compared various `Keras` initializers (Glorot uniform, HE normal, Lecun normal, Zeros, Ones) and a manual initialization with PCA (as described in [41]). We noticed that the weights initialization does not have a strong impact on the results and therefore the Glorot uniform initializer was selected [42].

For the regularization parameter η that balances the two terms (reconstruction and regularization) in the loss function, various strategies to tune its value already exist. For example, [43] showed that a constant $\eta > 1$ may outperform classical VAE (defined with $\eta = 1$). Moreover, [44] and [45] considered a moving parameter that gradually increases from 0 to 1 across iterations, linearly and according to a sigmoid, respectively. We tried the three approaches and opted for a constant regularization parameter equal to 10. Finally, we tested various learning rates for Adam optimizer but did not notice major variations in the performance, so we set it to 10^{-3} .

C. CONDITIONAL VARIABLES PREPROCESSING

We tried various combinations of the exogenous variables described in Table 2 and selected the one with the lowest MSE on the testing set. For a day t , the conditional vector X_t gathers the variables described below.

Without loss of generality, prices are categorical variables (Low, Normal or High), so, for an day t and an half-hour h , the prices p_t^h are encoded into two binary variables $\mathbf{1}_{p_t^h=\text{Low}}$, and $\mathbf{1}_{p_t^h=\text{High}}$ (if these two variables are null in the same time, the tariff is Normal). The position-inside-the-year $\kappa_t \in [0, 1]$ and the binary variable w_t for the type-of-day are also considered.

Taking into account the half-hourly temperature $\tau_t^1, \dots, \tau_t^H$ significantly improves the MSE on the testing set, but the dimension of the conditional variables vector is then quite high. We tried to reduce the dimension of the temperature profile and obtained better results. A PCA was performed on the vectors made of all temperatures at day t (half-hourly records and smoothed temperature). Three components were enough to explain 98% of the variance. Therefore, we only keep the three components provided by the PCA and re-scale them between 0 and 1 to provide the variables $\tilde{\tau}_t^1, \tilde{\tau}_t^2, \tilde{\tau}_t^3$. Then, they are considered as conditional variables (the daily temperature profiles $(\tau_t^1, \dots, \tau_t^H)$ are not taking into account anymore).

Therefore, for a day t , the vector of conditional variables X_t is made of the binary variables $w_t, \mathbf{1}_{p_t^1=\text{Low}}, \dots, \mathbf{1}_{p_t^H=\text{Low}}$, and $\mathbf{1}_{p_t^1=\text{High}}, \dots, \mathbf{1}_{p_t^H=\text{High}}$ and of the continuous variables $\tilde{\tau}_t^1, \tilde{\tau}_t^2, \tilde{\tau}_t^3$, and κ_t that lie in $[0, 1]$.

D. SIMULATOR CREATION

Finally, we emphasize that CVAEs may converge into local minima. To avoid it, each CVAE is trained 50 times and the one with the lowest MSE on testing set is selected. For each of the cluster presented in Section IV, we thus get a CVAE that takes as inputs the daily energy consumption profile $Y_t = (Y_t^1, \dots, Y_t^H)$ of the considered cluster (which is rescaled during the training process) and the conditional vector X_t described above. Then, the decoder is isolated and enables the generation of new data. Indeed, for a new vector $X_{t'}$ at a day t' , which can either be created or extracted from the data test, we sample a vector $Z_{t'} \sim \mathcal{N}(0, I_d)$ and give these two vectors as inputs of the decoder, which outputs a daily energy consumption profile. The quality of the generated data is evaluated in two situations. First, samples for the conditional vectors X_{T_0+1}, \dots, X_T associated with the training set are generated. Thus, we will measure the ability of the data generators to forecast energy consumption (we will see that we can deduce a foretasted density from the generated samples). Secondly, we will create new vectors X_t for which we modify the variables $\mathbf{1}_{p_t^h=\text{Low}}$, and $\mathbf{1}_{p_t^h=\text{High}}$ in order to measure the impact of tariff changes. These results are presented in Section VII and compare them with data generated according to a semi-parametric data generator presented below.

VI. SEMI-PARAMETRIC GENERATOR: ADDITIVE MODEL

The following semi-parametric method based on generalized additive models (GAM), see [34], is proposed to generate new daily consumption profile data. GAMs form a powerful and efficient semi-parametric approach to model electricity consumption (see, among others, [46]) as a sum of independent exogenous variable effects. Here, we assume that there exists a class of functions \mathcal{F} , such that, for a given half-hour h and an instance t , with x_t^h a vector of exogenous variables and p_t^h the tariff, the energy consumption expectation satisfies

$$\mathbb{E}[Y_t^h] = f^h(x_t^h, p_t^h), \quad f^h \in \mathcal{F}. \quad (9)$$

After estimating the functions f^h (we detail further the set \mathcal{F} and how GAMs may approximate these functions), we could compute the residuals and try to fit a model on them. They are centered, but a time dependence is observed, so adding a independent white noise to each forecast will not provide realistic profiles. It is important to note that the same problem can be found in renewable energy uncertainty forecasting and the need to generate scenarios (or trajectories) with inter-temporal dependency structure for multi-period stochastic optimization (see [25] for more details).

In this paper, we propose an approach based on a conjoint estimation of both mean and variation of the energy consumption. Then, we tried to used Gaussian copula to create trajectories, applying the methods proposed by Pinson *et al.* [25]. We faced an important problem: as soon as the function f^h is not very well-estimated, the residuals variance comes, in majority, from the estimation error. More precisely, a bad

estimation of the expected consumption leads to an increase of the estimated standard deviation.

As the focus is on generating realistic a profile (and not necessary on having the best forecast in expectation), the standard deviation used to simulate data must reflect the variability observed in energy consumption data. Thanks to the causality model of Section IV-A, that is now fitted on cluster consumptions (and not on individual ones), we can estimate the standard deviation of the noise as a function of the tariff and the half-hour h . We recall that we denote by $\sigma^h(p)$ the approximation of the standard $\sqrt{\text{Var}[Y^h(i) | P = p]}$ deviation associated with the half-hour h and the tariff p – see Equation (2). It is used to normalize the residuals, which should then be centered and of variance 1 (but not independent). Finally, we consider the standardized residual vectors and compute an estimation of their correlation matrix Σ . We can now generate new data points this way:

$$\begin{bmatrix} Y_t^1 \\ \vdots \\ Y_t^H \end{bmatrix} = \begin{bmatrix} f^1(x_t^1, p^1) \\ \vdots \\ f^H(x_t^H, p^H) \end{bmatrix} + (\sigma^1(p^1), \dots, \sigma^H(p^H))^T E_t,$$

where

$$E_t \sim \mathcal{N}(0, \Sigma). \tag{10}$$

Functions $(f^h)_{1 \leq h \leq H}$ are estimated with GAMs and the exogenous vector x_t^h gathers the temperature of the instance at the considerate half-hour τ_t^h , the smoothed temperature $\bar{\tau}_t$, the position in the year κ_t , the binary variable w_t , which is equal to 1 if the day considered is a week day and 0 otherwise. For each half-hour h , we set the same underlying GAM:

$$f^h(x_t^h, p_t^h) = s_\tau^h(\tau_t^h) + s_{\bar{\tau}}^h(\bar{\tau}_t) + s_\kappa^h(\kappa_t) + \alpha^h w_t + \xi_{\text{Low}}^h \mathbf{1}_{p_t^h=\text{low}} + \xi_{\text{High}}^h \mathbf{1}_{p_t^h=\text{High}}. \tag{11}$$

Therefore, \mathcal{F} is the set of functions that can be written this way. The s_τ^h , $s_{\bar{\tau}}^h$, and s_κ^h functions are catching the effect of the temperatures and of the yearly seasonality. They are approximated by cubic splines, i.e. C^2 -smooth functions made up of sections of cubic polynomials joined together at points of a grid (the knots). Fixing the number of knots k and their position is sufficient to determine a linear basis of dimension k in which these functions can be projected. The `mgcv` R-package allows to estimate the coordinates of the splines in their basis and the coefficients α^h , ξ_{Low}^h , and ξ_{High}^h that catch day of the week and tariff effects. Appendix E provides details on the estimation of the correlation matrix Σ , which makes it possible to model the correlations between the consumption profiles of two half-hours of the same day, whereas keeping a variance of the residuals that varies according to the half-hour and the price.

VII. EVALUATION OF THE DATA GENERATORS

A. EVALUATION METRICS

By generating lots of energy consumption profiles from the simulators, an estimation of their densities can be obtained.

Therefore, we use some proper scoring scores from probabilistic forecast evaluation to assess the quality of our generators. The three scores detailed below allow to evaluate the data generated on the testing period and compare both generators. For a day t of the testing set, from the vector of exogenous variables X_t , both generators output H -random vectors that are assumed to be drawn from an underlying distribution \widehat{F}_t . These distributions approximate the true and unknown H -dimensional distributions F_t from which the observation (Y_t^1, \dots, Y_t^H) is actually drawn. We generate $N = 200$ samples $\widehat{Y}_t^{(1)}, \dots, \widehat{Y}_t^{(N)}$ for each generator. From these H -random vectors, we can approximate the three scores described below, that measure the adequacy between the observation vectors Y_t and the distributions \widehat{F}_t .

First of all, for a distribution F , and a vector of observation y , the root mean squared error is considered: $\text{RMSE}(F, y) = \|\mathbb{E}[Y] - y\|$, where Y is a random vectors distributed according to F . The first score is thus the RMSE between the expectation of the distribution \widehat{F}_t (which we approximate with empirical mean of the generated samples) and the observation Y_t :

$$\text{RMSE}(\widehat{F}_t, Y_t) \approx \left\| \frac{1}{N} \sum_{i=1}^N \widehat{Y}_t^{(i)} - Y_t \right\|. \tag{12}$$

Here, the expectation of the distribution \widehat{F}_t is actually seen as a forecast of the energy consumption Y_t . But to evaluate the quality of \widehat{F}_t , a criterion including the variance and shape of the densities is necessary.

The two other scores are proper scoring rules used to evaluate weather ensembles or temporal trajectories generated by a statistical method (e.g., copula model). The energy score, introduced by Gneiting and Raftery [47], generalizes the univariate continuous ranked probability score (CRPS) and is defined as

$$\text{EN}(F, y) = \mathbb{E}[\|Y - y\|] - \frac{1}{2} \mathbb{E}[\|Y - Y'\|], \tag{13}$$

where Y and Y' are two independent random vectors that are distributed according to F . This score is approximated by splitting the generated samples in two groups $\widehat{Y}_t^{(1)}, \dots, \widehat{Y}_t^{(N/2)}$ and $\widehat{Y}_t^{(N/2+1)}, \dots, \widehat{Y}_t^{(N)}$:

$$\text{EN}(\widehat{F}_t, Y_t) \approx \frac{2}{N} \sum_{i=1}^{N/2} \|\widehat{Y}_t^{(i)} - Y_t\| - \frac{1}{N} \sum_{n=1}^{N/2} \|\widehat{Y}_t^{(i)} - \widehat{Y}_t^{(N/2+i)}\|. \tag{14}$$

Moreover, for the first three days of the testing set (that are actually the first three days of 2013), 20 samples generated by the simulators for one of the 4 clusters, their empirical means (computed on all the samples) and the corresponding observations Y_t are plotted in Figure 5. Plots of each cluster can be found in Figure 7 of Appendix F.

Scheuerer and Hamill have shown that the ability of energy score to detect correctly correlations between the components

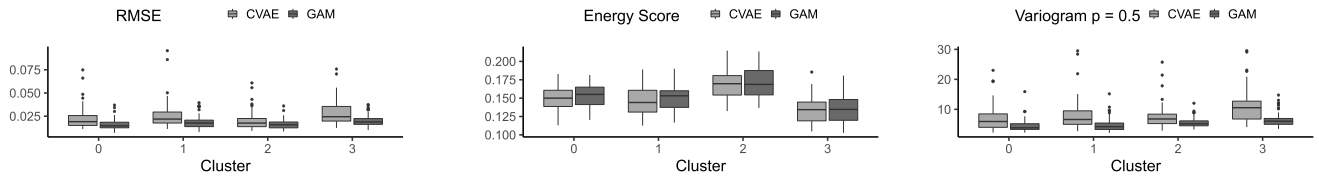


FIGURE 4. Boxplots. From left to right Root Mean Squared Error (RMSE), Energy Score and variogram for $p = 0.5$ evaluated for each day of the test period.

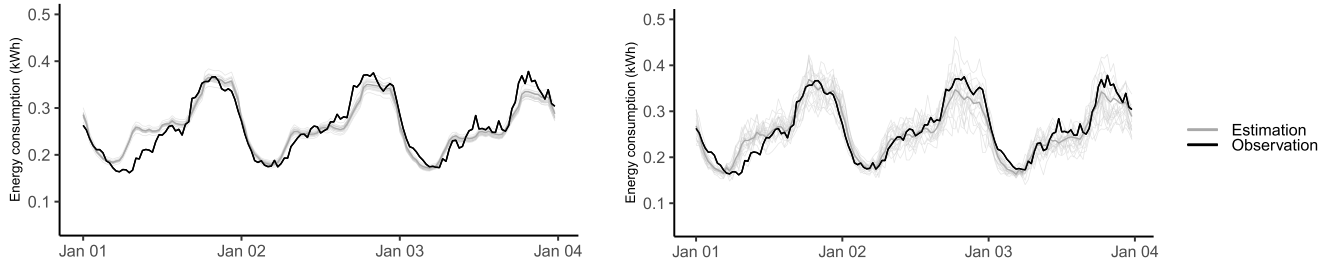


FIGURE 5. Electrical energy consumption profiles generated for a cluster of 158 households, over the first three days of the testing period, by using the CVAE (left) and GAM (right) simulators; 20 energy consumption profiles (light grey lines) and mean profile calculated on 200 samples (bold grey line) were obtained by giving to the two generators the exogenous variables observed over this period. Black line: observed profile.

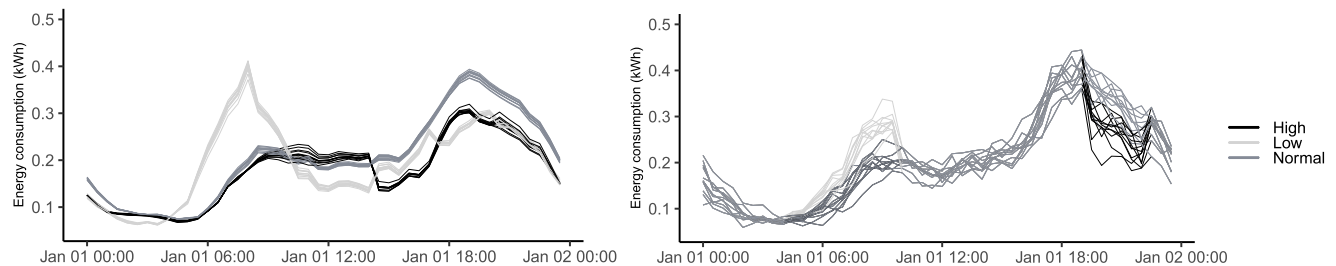


FIGURE 6. Electrical energy consumption profiles generated for a cluster of 158 households subjected to a dynamic Time of Use tariff for the first day of the data set test, by using the CVAE (left) and GAM (right) simulators; Dark grey lines: 20 energy consumption profiles, obtained by giving, to the two simulators, a Normal tariff for every half-hour, as well as weather and calendar variables observed over this period. Black lines: same plots but with a High tariff in the evening and Normal tariff otherwise. Light grey lines: same plots but with a Low tariff in the early morning and Normal tariff otherwise.

of the multivariate distribution was limited (see [48] for further details). To remedy, they introduced the variogram score of order p :

$$VG_p(F, y) = \sum_{h, h'=1}^H \left(|y^h - y^{h'}|^p - \mathbb{E}[|Y^h - Y^{h'}|^p] \right)^2, \quad (15)$$

where Y is a random vectors distributed according to F . On simulated data, they compared the performance of different scores (including the energy score) with the variogram scores for various p . This score is approximated with:

$$VG_p(\hat{F}_t, Y_t) \approx \sum_{h, h'=1}^H \left(|Y_t^h - Y_t^{h'}|^p - \frac{1}{N} \sum_{i=1}^N \left| (\hat{Y}_t^{(i)})^h - (\hat{Y}_t^{(i)})^{h'} \right|^p \right)^2. \quad (16)$$

We emphasize that for all the scores above, the smaller the value, the better the forecast.

B. NUMERICAL RESULTS

For each cluster and each day t of the testing set, we compute, for both generators (CVAE-based and GAM-based) the three scores (thanks to the 200 generated samples). Results are represented by boxplots in Figure 4.

It is quite difficult to discriminate significantly both generators from these scores, but some conclusions may still be drawn. First, RMSE bloxplots and plots suggest that GAM-based generators work better than those that use CVAE when it comes to generating the average value of the original data (which is approximated by the empirical mean of the samples). However, the energy score is slightly lower for the non-parametric approach (namely for CVAE-based simulator) than for the semi-parametric one (GAM-based simulator). Thus, the method that consists in adding a noise term to a forecast in expectation may have some limits whereas CVAEs seem to catch correctly the distributions of daily energy consumption.

Experiments of [48] highlight that, when the estimation of the average value of the original data is incorrect (namely when the expectation of F differs from the expectation of y in Equation (15)), variogram scores increase. Moreover, a too

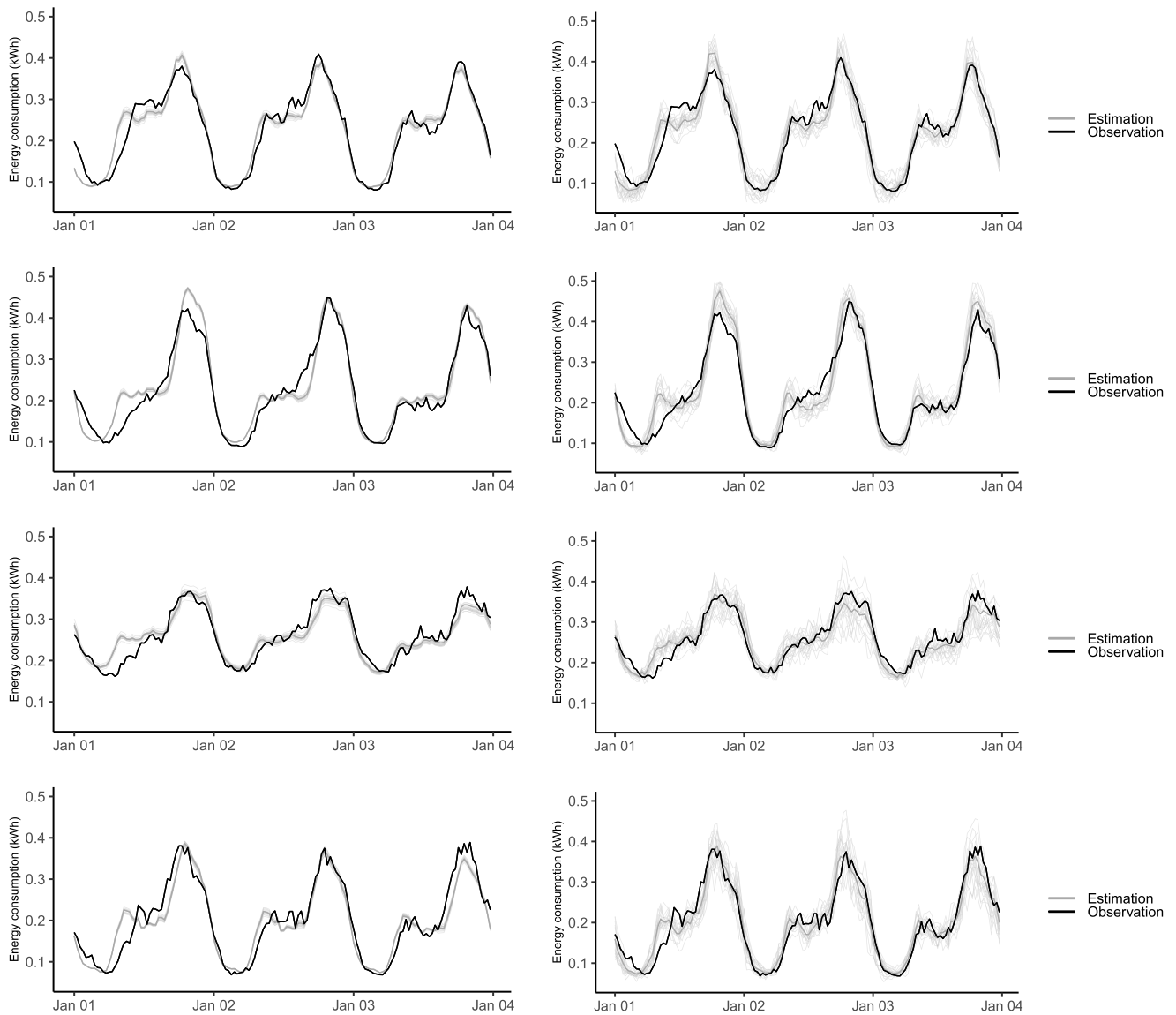


FIGURE 7. Electrical energy consumption profiles generated the CVAE-based generator (left) and GAM-based generator (right) for every cluster over the first three days of the testing period. Light grey lines: 20 energy consumption profiles and empirical mean profile calculated on 200 samples (in bold) were obtained by giving, to the two generators, the exogenous variables observed over this period. Black line: observed profiles.

low or a too high variance – when the variance of F differs from the one of y – also increases variogram. Given the variogram scores and the plots, we conclude that CVAE-based generators face an estimation of expected energy consumption worst than the semi-parametric generator but provide also samples with a too low variance. Conversely, GAM-based generators provide sample with too much variance, which also leads to a quite high variogram score.

Moreover, in the CVAE approach, consumption values from an half-hour to another are very correlated, when in the semi-parametric one, consumption profiles are more erratic. Observations suggest that the real variances and correlations lie somewhere in between. The semi-parametric method is very sensitive to the standard deviation $\sigma^h(p)$ estimations.

Thus, over-estimating these variances, provide, for sure, very different samples, which may be also very erratic. Concerning CVAE-based generator, the variance of the samples could manually be increased by generating the decoder inputs according to $\mathcal{N}(0, \sigma^2 I_d)$ with $\sigma > 1$.

Finally, we emphasize that in the semi-parametric approach, the variance depends only on the tariff and on the half-hour, whereas in the CVAE, all exogenous variables are taking into account. Moreover, the next section presents some strong advantages of the CVAE generator.

C. IMPACT OF THE TARIFF

In these last experiments, for a day t of the testing set, three different conditional vectors X_t^{Normal} , X_t^{Low} and X_t^{High}

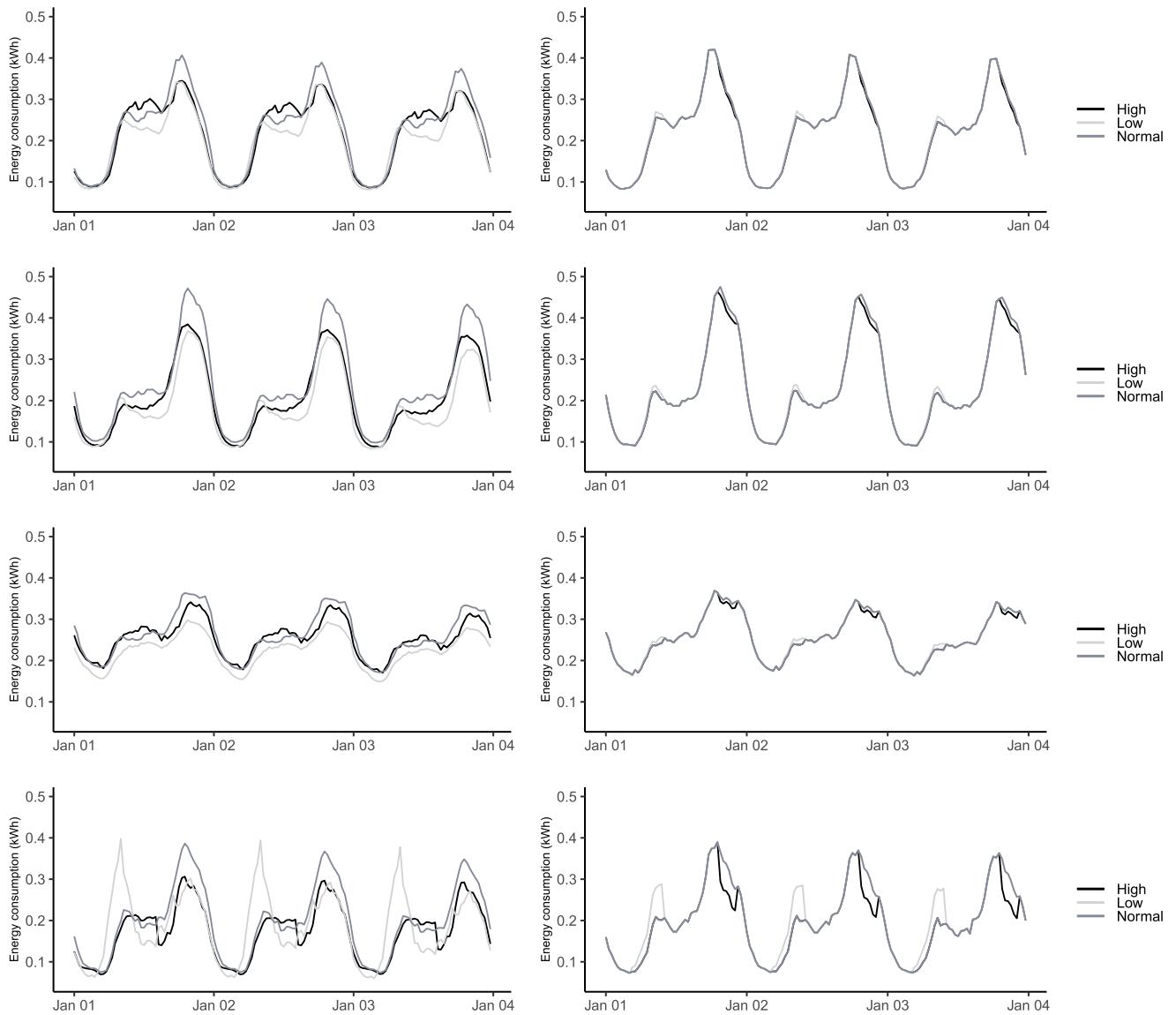


FIGURE 8. Electrical energy mean profiles (computed over 200 samples) generated the CVAE-based generator (left) and GAM-based generator (right) for every cluster over the first three days of the testing period. These profiles were obtained by giving, to the two simulators, a Normal tariff for every half-hour and the weather and calendar variables observed over this period. Black lines: same plots but with a High tariff in the evening and Normal tariff otherwise. Light grey lines: same plots but with a Low tariff in the early morning and Normal tariff otherwise.

are considered. The tariff is Normal for all the day long for X_t^{Normal} . For the vector X_t^{Low} , Low tariff applies from 4:30 to 9:30 a.m., and Normal one otherwise, finally, tariff is Normal except from 7:30 to 10 p.m. where it is High for X_t^{High} . For all other components, namely for the calendar and weather variables, X_t^{Normal} , X_t^{Low} , and X_t^{High} are equal to X_t . For the first day of the testing period, 20 samples generated by the generators for one of the four clusters are plotted in Figure 6. Empirical means (computed with all the samples) are plotted for the first three days of the testing period and for each cluster in Figure 8 of Appendix F.

For both data generators, an increase of the consumption when tariff Low is applied and a decrease when the tariff is

High are observed. For the GAM-based generator, the effect of the tariff is very interpretable, it is actually measured by coefficients ξ_{Low}^h and ξ_{High}^h of equation (11). This model makes actually this assumption that the tariff effect only depends on the half-hour. Moreover, matrix Σ models the correlations between the energy consumption at two half hours of the same day; this implicitly assumes that these correlations do not change according to the applied tariff profile. Conversely, CVAE-based generator does not have this assumption and the effect of a tariff may differ from a day to another.

Moreover, two effects that cannot be modelled by the semi-parametric approach are observed. First, the fall of the energy consumption occurs a little bit before the effective

establishment of a special tariff High and continues a little after it is stopped. Thus, the effect of the High tariff exceeds the time window in which the special tariff is actually applied. This is called a side effect. Secondly, in comparison to a day of Normal tariff, when tariff Low is applied in the morning, there is a drop of the consumption in the afternoon and evening. Similarly, we observe a little increase of the consumption in the afternoon when the tariff is High during the evening. Therefore, the fall or rise in consumption shifts to another time of the day when a special tariff is applied over a time window. This is called a rebound effect. These side and rebound effects are well known behaviors of consumers and it is very valuable that the generator detects them. The main drawback of this non-parametric generator is the generation of non-intuitive consumption profiles when the input is a tariff profile never observed in the training set, like an entire day of High tariff for example. This shows that the method has a limited generalization capacity. Enlarging the data set, especially the variety of price signals, would eliminate this limitation. On the other hand, for a full day of tariff High, the semi-parametric model generates samples with an energy consumption below the typical one for each half-hour, which is unrealistic since electricity uses cannot be delayed indefinitely.

Figure 8 of Appendix F shows that tariff-responsiveness vary from a cluster to another, i.e., rebound or side effects are not always observed and the amount of electricity over or under consumed also depends on the considered cluster. These results fully illustrate the motivation behind the use of the causality model to cluster consumers.

VIII. CONCLUSIONS

This paper proposed a data-driven and non-parametric methodology, based on CVAE, for generating synthetic energy consumption profiles for households enrolled in a DR program with different tariff schemes. The results for the largest data set publicly available (released by UK Power Networks) show that the proposed non-parametric generator captures correctly the effect of the exogenous variables and performs almost as well as the benchmark semi-parametric generator to generate the mean value of the original data. Besides and above all, the whole point of the CVAE approach comes from its ability to capture the effect of a daily tariff profile on the daily consumption profiles. Indeed, unlike the semi-parametric generator that only captures the effect of a special tariff for the half-hours affected by this tariff change, the generator built from a decoder of a CVAE provides daily consumption samples for a daily tariff policy, including rebound and side effects. Moreover, for the same conditional variables as inputs, the generated samples differ from one group of consumers to another. Thus, the proposed clustering approach divides correctly the households according to their responsiveness to a tariff profile.

Finally, to deal with the lack of variability in the sent tariff profile of the original data set, we could imagine an online

data generator: when a new tariff profile is sent, the observed consumption is integrated in the training set and the data generator is updated. The use of transfer learning methods could also improve the realism of the generated data. This machine learning field focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.

Therefore, by combining data sets of consumer responsiveness to various DR programs (i.e. by combining diverse knowledge of electricity demand in the face of tariff changes), a data set with a higher variability in the sent tariff profiles may be obtained. These data generators could be very useful to test potential future DR policies, before deploying such solutions in consumer households. Another topic of interest is the extension of the proposed model to consider privacy of the smart meter measurements and where recent research in privacy-preserving machine learning is a promising approach [49].

APPENDIX

A. EXPONENTIAL SMOOTHING

This section describes the construction of the exogenous variable $\bar{\tau}_t$, which is a smoothed air temperature (that models the thermal inertia of buildings, with the a -exponential smoothing. To do so, London temperatures are considered as a 1-dimensional half-hourly time series $\tau_1^1, \dots, \tau_t^H, \tau_2^1, \dots$ and not anymore as H -dimensional profiles. For any day $t \in \{1, \dots, T\}$, and any half-hour $h \in \{1, \dots, H\}$ the smoothed temperature is defined by

$$\bar{\tau}_t^h = \begin{cases} \tau_1^1 & \text{if } t = 1 \text{ and } h = 1 \\ (1 - a)\tau_t^h + a\bar{\tau}_t^{h-1} & \text{if } h \neq 1 \\ (1 - a)\tau_t^1 + a\bar{\tau}_{t-1}^H & \text{else,} \end{cases} \quad (17)$$

where the smoothing parameter a is in $[0, 1]$. After testing several values, we set $a = 0.998$. Then, for a given day $t \in \{1, \dots, T\}$, $\bar{\tau}_t$ is simply the daily average smoothed temperature:

$$\bar{\tau}_t = \frac{1}{H} \sum_{h=1}^H \bar{\tau}_t^h. \quad (18)$$

B. CAUSALITY MODEL

In the following, we detail models that were used to compute, for each household i and for each tariff $p \in \mathcal{P}$, the daily profile of the mean and the standard deviation of household i energy consumption. We recall that these profiles are then used to cluster households. Therefore, we have to estimate, for each half-hour h , the expectation and the standard deviation of the random variable $Y^h(i) | P = p$. To do so, we train a model that gives, for the tariff p and the exogenous variables x^h , a forecast of the expected consumption at h when tariff p is selected and a forecast of the standard deviation of this consumption. For any exogenous variable x_t^h and tariff p_t^h , the random energy consumption $Y_t^h(i)$, of household i at the half hour h of the day t , is assumed Gaussian of mean $\mu_i(x_t^h, p_t^h)$ and standard deviation $\sigma_i(x_t^h, p_t^h)$.

Moreover, we assume that these mean and standard deviation:

$$\mu^{i,h}(x_t^h, p_t^h) = \mathbb{E}[Y_t^h(i)] \quad (19)$$

and

$$\sigma^{i,h}(x_t^h, p_t^h) = \sqrt{\text{Var}[Y_t^h(i)]}, \quad (20)$$

depend on additive smooth predictors. Generalized additive models (GAM – see [34]) may model electricity consumption (see [46]) as a sum of independent exogenous variable effects. Here, they are used to estimate conjointly, for any half-hour h of a day t and any tariff $p \in \mathcal{P}$, both $\mu^{i,h}(x_t^h, p)$ and $\sigma^{i,h}(x_t^h, p)$. These approximations are denoted by $\widehat{\mu}^i$, $h(x_t^h, p)$ and $\widehat{\sigma}^{i,h}(p, x_t^h)$, respectively. For each half-hour h , we set the same underlying models:

$$\begin{aligned} \mu^{i,h}(x_t^h, p_t^h) &= s_\tau^{i,h}(\tau_t^h) + \xi_L^{i,h} \mathbf{1}_{p_t^h=\text{Low}} \\ &\quad + \xi_N^{i,h} \mathbf{1}_{p_t^h=\text{Normal}} + \xi_H^{i,h} \mathbf{1}_{p_t^h=\text{High}} \end{aligned} \quad (21)$$

$$\begin{aligned} \sigma^{i,h}(x_t^h, p_t^h) &= \gamma_L^{i,h} \mathbf{1}_{p_t^h=\text{Low}} \\ &\quad + \gamma_N^{i,h} \mathbf{1}_{p_t^h=\text{Normal}} + \gamma_H^{i,h} \mathbf{1}_{p_t^h=\text{High}}. \end{aligned} \quad (22)$$

where $s_\tau^{i,h}$, the function catching the effect of the temperature, is approximated by a cubic spline. Fixing the number of knots k and their positions is enough to determine a linear basis of dimension k in which this function can be projected. The `mgcv` R-package permits to estimate the coordinates of the spline in its basis and all the coefficients $\xi_L^{i,h}$, $\xi_N^{i,h}$, $\xi_H^{i,h}$, $\gamma_L^{i,h}$, $\gamma_N^{i,h}$, and $\gamma_H^{i,h}$ defined in Equation (22), which catch tariff effect. We highlight that models fitted on variances are linear. Both models (on mean and standard deviation) are estimated simultaneously, by setting the model family parameter of the `gam` function to the Gaussian location-scale model family.

Once the function and coefficients have been estimated (we write $\widehat{s}^{i,h}$ for the estimation of $s^{i,h}$ and so on), for any tariff p , the estimations $\widehat{\mu}^{i,h}(x_t^h, p)$ and $\widehat{\sigma}^{i,h}(p, x_t^h)$ are computed:

$$\begin{aligned} \widehat{\mu}^{i,h}(x_t^h, p) &= \widehat{s}_\tau^{i,h}(\tau_t^h) + \widehat{\xi}_L^{i,h} \mathbf{1}_{p=\text{Low}} \\ &\quad + \widehat{\xi}_N^{i,h} \mathbf{1}_{p=\text{Normal}} + \widehat{\xi}_H^{i,h} \mathbf{1}_{p=\text{High}} \end{aligned} \quad (23)$$

$$\begin{aligned} \text{and } \widehat{\sigma}^{i,h}(p, x_t^h) &= \widehat{\gamma}_L^{i,h} \mathbf{1}_{p=\text{Low}} \\ &\quad + \widehat{\gamma}_N^{i,h} \mathbf{1}_{p=\text{Normal}} + \widehat{\gamma}_H^{i,h} \mathbf{1}_{p=\text{High}}. \end{aligned} \quad (24)$$

Finally, the approximations of tariff impact are provided by:

$$\mathbb{E}[Y^{i,h} | P = p] \approx \frac{1}{T} \sum_{t=1}^T \widehat{\mu}^{i,h}(p, x_t^h) \quad (25)$$

$$\text{and } \sqrt{\text{Var}[Y^h | P = p]} \approx \frac{1}{T} \sum_{t=1}^T \widehat{\sigma}^{i,h}(p, x_t^h). \quad (26)$$

C. CLUSTERING METHOD

Here, the three steps of the clustering method proposed in Section IV are detailed.

1) SCALING AND GATHERING PROFILES

For an household $i \in \mathcal{I}$, for all $p \in \mathcal{P}$, the daily expected consumption profile $\mu_i^1(p), \dots, \mu_i^H(p)$ is considered. We assume that there is a base tariff $p_0 \in \mathcal{P}$ that corresponds to a signal of no incentive, namely Normal tariff. We consider the quantity $\bar{\mu}_i = \frac{1}{H} \sum_{h=1}^H \mu_i^h(p_0)$ that is an approximation of the average daily expected consumption of household i under no DR program.

Then, all the profiles of household i are rescaled by this quantity and, for each tariff $p \in \mathcal{P}$, the daily consumption profiles under tariff p of all the households $i \in \mathcal{I}$ are gathered in a matrix $\mathbf{M}(p) \in \mathcal{M}_{|\mathcal{I}| \times H}$. Finally, the matrix $\mathbf{M} \in \mathcal{M}_{|\mathcal{I}| \times H | \mathcal{P}|}$ is created by binding by column matrices $\mathbf{M}(p)$, so

$$\mathbf{M}_{i,h}(p) = \frac{\mu_i^h(p)}{\bar{\mu}_i}$$

and

$$\mathbf{M} = \left(\mathbf{M}(1) \mid \dots \mid \mathbf{M}(P) \right). \quad (27)$$

a: LOW RANK APPROXIMATION

Since we are interested in energy consumption, all the coefficients of \mathbf{M} are non-negative – we will write $\mathbf{M} \geq 0$ and say that this matrix is non-negative. To reduce dimension of non-negative matrices, the factorization method proposed by [50] and [51] that uses non-negativity constraints is considered. The integer $r \ll \min(|\mathcal{I}|, H|\mathcal{P}|)$ that will ensure a reduction of the dimension is fixed (we chose $r = 5$ in our case study). The non-negative matrix factorization (NMF) approximates matrix of profiles \mathbf{M} with \mathbf{WH} by minimizing the euclidean distance between both matrices under the constraint that \mathbf{W} and \mathbf{H} are non-negative matrices of size $|\mathcal{I}| \times r$ and $r \times H|\mathcal{P}|$, respectively. Function NMF of the Python-library `sklearn.decomposition` allows to approximate \mathbf{W} and \mathbf{H} with a coordinate descent solver. For simplicity of notation, \mathbf{W} is confounded with its approximation. Thus, for any household $i \in \mathcal{I}$, we get r features, namely the i^{th} line of matrix \mathbf{W} , that we denote by \mathbf{W}_i . in the following.

2) k-MEDOID CLUSTERING

Now, the vectors \mathbf{W}_i . allow to cluster households in k clusters.

In k -means clustering, the center of a given cluster is simply the average between the points of this cluster. Since it can be influenced by extreme value, k -means algorithm is sensitive to outliers. Conversely, k -medoid algorithm chooses data points to represent clusters, which makes it more robust and favors a clustering where clusters have sizes of the same order. This algorithm was introduced by [52] with the L^1 -norm. Here, we use it with the Euclidean distance and the best clustering C_1^*, \dots, C_k^* is the one that minimizes the following

criteria:

$$\{C_1^*, \dots, C_k^*\} \in \underset{\{C_1, \dots, C_k\}}{\operatorname{argmin}} \sum_{\ell=1}^k \sum_{i \in C_\ell} \|\mathbf{W}_i - \mathbf{W}_{C_\ell}\|^2$$

with

$$C_\ell \in \underset{i \in C_\ell}{\operatorname{argmin}} \sum_{j \in C_\ell} \|\mathbf{W}_i - \mathbf{W}_j\|^2, \quad (28)$$

where $\|\cdot\|$ is the Euclidean norm. The clusters are computed by using KMedoid function implemented in the Python-library sklearn_extra.

D. VARIATIONAL AUTOENCODER

The calculations below are an adaptation of the ones proposed by [36] to our case-study. The generation of the data is assumed to follow a two-steps process: firstly, a variable Z was sampled from a standard Gaussian and then, Y was sampled from the distribution $p_{\theta^*}(\cdot|Z)$. The decoder, parametrized by θ , can model this process: with $Z \sim \mathcal{N}(0, I_d)$ as input, it generates the variable Y , conditionally to Z , by sampling it from $p_{\theta}(\cdot|Z)$, which is an approximation of the true distribution $p_{\theta^*}(\cdot|Z)$. In our generation process, we will denote by $q_Y(Z)$ the approximation made by the encoder of the density of $Z|Y$. The variational autoencoder is trained in a way that q_Y is the Gaussian of mean $\mu(Y)$ and covariance matrix $\Sigma(Y)$, where $\mu(Y)$ and $\Sigma(Y)$ are the outputs of the encoder for the input Y . For $Y \in \mathbb{R}^H$, by using Bayes' theorem and the variables Z sampled from the encoder distribution q_Y , the log-marginal likelihood $p_{\theta}(Y)$ satisfies

$$\begin{aligned} \ln p_{\theta}(Y) &= \mathbb{E}_{Z \sim q_Y} [\ln p_{\theta}(Y)] \\ &= \mathbb{E}_{Z \sim q_Y} \left[\ln \frac{p_{\theta}(Y|Z)p_{\theta}(Z)}{p_{\theta}(Z|Y)} \right] \\ &= \mathbb{E}_{Z \sim q_Y} \left[\ln \frac{q_Y(Z)}{p_{\theta}(Z|Y)} + \ln \frac{p_{\theta}(Z)}{q_Y(Z)} + \ln p_{\theta}(Y|Z) \right] \\ &= \operatorname{D}_{\text{KL}}(q_Y(Z) \parallel p_{\theta}(Z|Y)) - \operatorname{D}_{\text{KL}}(q_Y(Z) \parallel p_{\theta}(Z)) \\ &\quad + \mathbb{E}_{Z \sim q_Y} [\ln p_{\theta}(Y|Z)]. \end{aligned} \quad (29)$$

The first term corresponds to the error made by approximating the distribution $p_{\theta}(\cdot|Y)$ with q_Y . Thus to conjointly maximizing the log-likelihood and minimizing this approximation error, the loss

$$\operatorname{D}_{\text{KL}}(q_Y(Z) \parallel p(Z)) - \mathbb{E}_{Z \sim q_Y} [\ln p_{\theta}(Y|Z)], \quad (30)$$

has to be minimized. The two parts of the equation above are known as the regularization term and the reconstruction term, respectively. We recall that q_Y is the Gaussian distribution of mean $\mu(Y)$ and of covariance matrix $\Sigma(Y)$ and that we assume $Z \sim \mathcal{N}(0, I_d)$, so the regularization term is the Kullback–Leibler divergence between $\mathcal{N}(\mu(Y), \Sigma(Y))$ and a standard d -multidimensional normal distribution. Moreover, we highlight that if the decoder samples $Y|Z$ from a distribution of the exponential family,

$$p_{\theta}(Y|Z) = a(Y)b(Z) \exp(\eta(Z)T(Y)), \quad (31)$$

with θ gathering the functions a, b, η , and T . Then, the second term is explicit. But, for a given Z , the decoder outputs a unique vector $D(Z) = \hat{Y}$, so inferring the previous distribution is a tough task. Nevertheless, assuming that $Y|Z$ is a multivariate Gaussian of mean $D(Z)$ and with a known covariance matrix $\sigma^2 I_d$, a very simple expression of the regularization term is obtained:

$$-\ln p_{\theta}(Y|Z) = \frac{1}{2\sigma^2} \|Y - D(Z)\|_2^2 - \ln(2\pi^{H/2}\sigma). \quad (32)$$

Therefore, the loss defined in Equation (30) can be re-written:

$$\frac{1}{2\sigma^2} \|Y - \hat{Y}\|_2^2 + \operatorname{D}_{\text{KL}}(\mathcal{N}(\mu(Y), \Sigma(Y)) \parallel \mathcal{N}(0, I_d)). \quad (33)$$

Under all the assumptions above, and given the independent observations Y_1, \dots, Y_{T_0} , to obtain the generative process that best models the real one, we will thus consider the loss

$$L_{\text{VAE}}(\eta) = \frac{1}{T_0} \sum_{t=1}^{T_0} \left(\|Y_t - \hat{Y}_t\|^2 + \eta \operatorname{D}_{\text{KL}}(\mathcal{N}(\mu(Y_t), \Sigma(Y_t)) \parallel \mathcal{N}(0, I_d)) \right). \quad (34)$$

We recall that the vectors \hat{Y}_t are the outputs of the decoder $D(Z_t)$, where the random variable is sampled from a d -multivariate Gaussian of mean $\mu(Y_t)$ and covariance matrix $\Sigma(Y_t)$. This loss is conjointly maximizing the log-likelihood of the observation with the data generation process distribution:

$$\ln p_{\theta}(Y_1, \dots, Y_{T_0}) = \sum_{t=1}^{T_0} \ln p_{\theta}(Y_t) \quad (35)$$

and minimizing the approximation error

$$\sum_{t=1}^{T_0} \operatorname{D}_{\text{KL}}(q_{Y_t}(Z) \parallel p_{\theta}(Z|Y_t)). \quad (36)$$

It is important to underline that the previous calculations are still valid when all the distributions are conditioned by exogenous variables.

E. CORRELATION MATRIX OF SEMI-PARAMETRIC GENERATOR

Here the estimation of the matrix Σ , which is used to generate profiles with correlations between temporal intervals of the same day, is detailed. If the model defined by Equation (10) was true, residuals $Y_t^h - f^h(x_t^h, p_t^h)$ should be Gaussian of mean 0 and standard deviation $\sigma^h(p_t^h)$. Thus the vector of standardized residuals $e_t = (e_t^h)_{1 \leq h \leq H}$ is considered, where

$$e_t^h = \frac{Y_t^h - f^h(x_t^h, p_t^h)}{\sigma^h(p_t^h)}. \quad (37)$$

Assuming the model above, the covariance matrix Σ of vectors e_1, \dots, e_{T_0} should have 1 on the diagonals and all other coefficients between -1 and 1 . To deal with our imperfect

modeling and avoid again the problem of high standard deviation coming from the estimation error, Σ is approximated by the empirical correlation matrix of vectors e_1, \dots, e_{T_0} . From the T_0 observations e_1, \dots, e_{T_0} , which are assumed independent, of the H -dimensional random vector $e = (e^1, \dots, e^H)$, the coefficients of the $H \times H$ -correlation matrix Σ are defined by

$$\Sigma_{i,j} = \frac{\text{cov}(e^i, e^j)}{\sqrt{\text{Var}(e^i) \text{Var}(e^j)}},$$

where

$$\text{cov}(e^i, e^j) = \mathbb{E}(e^i e^j) - \mathbb{E}(e^i) \mathbb{E}(e^j). \quad (38)$$

We point out that in the case of random variables e^1, \dots, e^H of standard deviation 1 (we assume it in the semi-parametric simulator described in Section VI), covariance and correlation matrices are equal. In there, $\Sigma_{i,j}$ is estimated by replacing covariances and variances of random variables e^i and e^j by their empirical estimations:

$$\text{cov}(e^i, e^j) \approx \frac{1}{T_0 - 1} \sum_{t=1}^{T_0} (e_t^i - \bar{e}^i)(e_t^j - \bar{e}^j) \quad (39)$$

and

$$\text{Var}(e^i) \approx \frac{1}{T_0 - 1} \sum_{t=1}^{T_0} (e_t^i - \bar{e}^i)^2, \quad (40)$$

with

$$\bar{e}^i = \frac{1}{T_0} \sum_{t=1}^{T_0} e_t^i.$$

F. EXPERIMENTS - GRAPHICAL RESULTS

See Figures 7 and 8.

ACKNOWLEDGMENT

Margaux Brégère is very grateful to the entire team of the Center for Power and Energy Systems at INESC TEC for their wonderful welcome in Porto. In particular, she warmly thanks Kamalanathan Ganesan who explained to her the developed causality model and José Ricardo Andrade for his help with Tensorflow and all the IT support.

REFERENCES

- [1] P. Mallet, P.-O. Granstrom, P. Hallberg, G. Lorenz, and P. Mandatova, "Power to the people!: European perspectives on the future of electric distribution," *IEEE Power Energy Mag.*, vol. 12, no. 2, pp. 51–64, Mar. 2014.
- [2] P. Siano, "Demand response and smart grids—A survey," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 461–478, Feb. 2014.
- [3] G. Dutta and K. Mitra, "A literature review on dynamic pricing of electricity," *J. Oper. Res. Soc.*, vol. 68, no. 10, pp. 1131–1145, Oct. 2017.
- [4] F. Alfaverh, M. Denä, and Y. Sun, "Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management," *IEEE Access*, vol. 8, pp. 39310–39321, 2020.
- [5] M. Brégère, P. Gaillard, Y. Goude, and G. Stoltz, "Target tracking for contextual bandits: Application to demand side management," in *Proc. 36th Int. Conf. Mach. Learn. (ICML PMLR)*, 2019, pp. 754–763.
- [6] S. S. Karlsen, M. Hamdy, and S. Attia, "Methodology to assess business models of dynamic pricing tariffs in all-electric houses," *Energy Buildings*, vol. 207, Jan. 2020, Art. no. 109586.
- [7] J. R. Schofield *et al.*, *Low Carbon London Project: Data From the Dynamic Time-of-Use Electricity Pricing Trial, 2013*, UK Data Service, SN: 7857, 2015. Accessed: Feb. 2020. [Online]. Available: data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households
- [8] G. Chicco, R. Napoli, and F. Piglion, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 933–940, May 2006.
- [9] G. Le Ray and P. Pinson, "Online adaptive clustering algorithm for load profiling," *Sustain. Energy, Grids Netw.*, vol. 17, Mar. 2019, Art. no. 100181.
- [10] P. P. Rodrigues, J. Gama, and J. P. Pedroso, "Hierarchical clustering of time-series data streams," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 5, pp. 615–627, May 2008.
- [11] J. N. Fidalgo, M. A. Matos, and L. Ribeiro, "A new clustering algorithm for load profiling based on billing data," *Electric Power Syst. Res.*, vol. 82, no. 1, pp. 27–33, Jan. 2012.
- [12] M. Sun, I. Konstantelos, and G. Strbac, "C-vine copula mixture model for clustering of residential electrical load pattern data," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2382–2393, May 2017.
- [13] K. Ganesan, J. Tomé Saraiva, and R. J. Bessa, "On the use of causality inference in designing tariffs to implement more effective behavioral demand response programs," *Energies*, vol. 12, no. 14, p. 2666, Jul. 2019.
- [14] J. Saez-Gallego and J. M. Morales, "Short-term forecasting of price-responsive loads using inverse optimization," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4805–4814, Sep. 2018.
- [15] G. Le Ray, E. M. Larsen, and P. Pinson, "Evaluating price-based demand response in practice—With application to the EcoGrid EU experiment," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2304–2313, May 2016.
- [16] S. Mohajeryami, I. N. Moghaddam, M. Doostan, B. Vatani, and P. Schwarz, "A novel economic model for price-based demand response," *Electr. Power Syst. Res.*, vol. 135, pp. 1–9, Jun. 2016.
- [17] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, "A bottom-up approach to residential load modeling," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 957–964, May 1994.
- [18] S. Park, H. Kim, H. Moon, J. Heo, and S. Yoon, "Concurrent simulation platform for energy-aware smart metering systems," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1918–1926, Aug. 2010.
- [19] M. Muratori, M. C. Roberts, R. Sioshansi, V. Marano, and G. Rizzoni, "A highly resolved modeling technique to simulate residential power demand," *Appl. Energy*, vol. 107, pp. 465–473, Jul. 2013.
- [20] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic electricity use: A high-resolution energy demand model," *Energy Buildings*, vol. 42, no. 10, pp. 1878–1887, Oct. 2010.
- [21] J. M. G. López, E. Pouresmaeil, C. A. Cañizares, K. Bhattacharya, A. Mosaddegh, and B. V. Solanki, "Smart residential load simulator for energy management in smart grids," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1443–1452, Feb. 2019.
- [22] S. Gottwalt, W. Ketter, C. Block, J. Collins, and C. Weinhardt, "Demand side management—A simulation of household behavior under variable prices," *Energy Policy*, vol. 39, no. 12, pp. 8163–8174, Dec. 2011.
- [23] Y. Iwafune, K. Ogimoto, Y. Kobayashi, and K. Murai, "Driving simulator for electric vehicles using the Markov chain Monte Carlo method and evaluation of the demand response effect in residential houses," *IEEE Access*, vol. 8, pp. 47654–47663, 2020.
- [24] J. Li, Y. Lei, Q. Huang, Z. Qin, and B. Chen, "Feature analysis of generalized load patterns considering active load response to real-time pricing," *IEEE Access*, vol. 7, pp. 119443–119453, 2019.
- [25] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, Jan. 2009.
- [26] Y. Chen, X. Wang, and B. Zhang, "An unsupervised deep learning approach for scenario forecasts," in *Proc. Power Syst. Comput. Conf. (PSCC)*, Jun. 2018, pp. 1–7.
- [27] A. Marot, L. Rosin, L. Crochepierre, B. Donnot, P. Pinson, and L. Boudjeloud-Assala, "Interpreting atypical conditions in systems with deep conditional autoencoders: The case of electrical consumption," in *Proc. Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases (ECML PKDD)*, 2019, pp. 638–654.
- [28] M. Brégère and M. Huard, "Online hierarchical forecasting for power consumption data," 2020, *arXiv:2003.00585*. [Online]. Available: <http://arxiv.org/abs/2003.00585>

- [29] J. Schofield, R. Carmichael, S. Tindemans, M. Woolf, M. Bilton, and G. Strbac, "Residential consumer responsiveness to time-varying pricing," Imperial College London, 'Low Carbon London' LCNF Project, London, U.K., Tech. Rep. A3, 2014. [Online]. Available: <https://spiral.imperial.ac.uk/bitstream/10044/1/21252/2/LCL%20Learning%20Report%20-%20A3%20-%20Residential%20consumer%20responsiveness%20to%20time%20varying%20pricing.pdf>
- [30] National Oceanic and Atmospheric Administration (NOAA). *Hourly/Sub-Hourly Observational Data Map*. Accessed: Feb. 2020. [Online]. Available: <https://gis.ncdc.noaa.gov/maps/ncei/cdo/hourly>
- [31] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *J. Oper. Res. Soc.*, vol. 54, no. 8, pp. 799–805, Aug. 2003.
- [32] Y. Goude, R. Nedellec, and N. Kong, "Local short and middle term electricity load forecasting with semi-parametric additive models," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 440–446, Jan. 2014.
- [33] M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi, "Optimized clusters for disaggregated electricity load forecasting," *REVSTAT-Stat. J.*, vol. 8, no. 2, pp. 105–124, Nov. 2010.
- [34] S. Wood, *Generalized Additive Models: An Introduction with R*. Boca Raton, FL, USA: CRC Press, 2006.
- [35] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Statist.-Simul. Comput.*, vol. 3, no. 1, pp. 1–27, 1974.
- [36] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. 2nd Int. Conf. Learn. Represent. (ICLR)*, 2014.
- [37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
- [38] G. E. Hinton and R. S. Zemel, "Autoencoders, minimum description length and Helmholtz free energy," in *Proc. 6th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 1994, pp. 3–10.
- [39] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Proc. 28th Conf. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 3581–3589.
- [40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015.
- [41] V. Miranda, J. da Hora Martins, and V. Palma, "Optimizing large scale problems with metaheuristics in a reduced space mapped by autoencoders—Application to the wind-hydro coordination," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 3078–3085, Nov. 2014.
- [42] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2010, pp. 249–256.
- [43] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "Beta-VAE: Learning basic visual concepts with a constrained variational framework," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, 2017, vol. 2, no. 5, p. 6.
- [44] D. Liang, R. G. Krishnan, M. D. Hoffman, and T. Jebara, "Variational autoencoders for collaborative filtering," in *Proc. World Wide Web Conf. World Wide Web (WWW)*, 2018, pp. 689–698.
- [45] S. R. Bowman, L. Vilnis, O. Vinyals, A. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," in *Proc. 20th SIGNLL Conf. Comput. Natural Lang. Learn.*, 2016, pp. 10–21.
- [46] P. Gaillard, Y. Goude, and R. Nedellec, "Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting," *Int. J. Forecasting*, vol. 32, no. 3, pp. 1038–1050, Jul. 2016.
- [47] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *J. Amer. Stat. Assoc.*, vol. 102, no. 477, pp. 359–378, Mar. 2007.
- [48] M. Scheuerer and T. M. Hamill, "Variogram-based proper scoring rules for probabilistic forecasts of multivariate Quantities*," *Monthly Weather Rev.*, vol. 143, no. 4, pp. 1321–1334, Apr. 2015.
- [49] M. Al-Rubaie and J. M. Chang, "Privacy-preserving machine learning: Threats and solutions," *IEEE Secur. Privacy*, vol. 17, no. 2, pp. 49–58, Mar. 2019.
- [50] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, Jun. 1994.
- [51] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, p. 788, 1999.
- [52] L. Kaufman and P. J. Rousseeuw, "Clustering by means of Medoids," in *Statistical Data Analysis Based on the L1 Norm*, Y. Dodge, Ed. Amsterdam, The Netherlands: North Holland, 1987, pp. 405–416.



MARGAUX BRÉGÈRE was born in France, in 1993. She received a french "Grande École" equivalent to an M.Sc. degree in engineering sciences from Supaéro, in 2017, and the M.Sc. degree in applied mathematics from University Paris-Saclay. She is currently pursuing the Ph.D. degree (third year) with the EDF Research and Development Lab, Saclay, France, at the Laboratoire de Mathématiques d'Orsay and in SIERRA team at Inria Paris, advised by Yannig Goude, Gilles

Stoltz, and Pierre Gaillard.

Her research focuses on power consumption forecasting and mainly on bandit models for power consumption control. She develops reinforcement learning algorithms to maintain the electricity consumption-production balance.



RICARDO J. BESSA (Senior Member, IEEE) was born in Viseu, Portugal, in 1983. He received the Licenciado (five-year) degree in electrical and computer engineering from the Faculty of Engineering, University of Porto (FEUP), Porto, Portugal, in 2006, the M.Sc. degree in data analysis and decision support systems from the Faculty of Economics, University of Porto (FEP), Porto, in 2008, and the Ph.D. degree from the Doctoral Program in Sustainable Energy Systems (MIT

Portugal), FEUP, in 2013.

He is currently the Coordinator of the Center for Power and Energy Systems, INESC TEC. He worked in several international projects, such as the European Projects FP6 ANEMOS.plus, FP7 SuSTAINABLE, FP7 evolVDISO, Horizon 2020 UPGRID, and InteGrid, and international collaboration with Argonne National Laboratory for the U.S. Department of Energy. His research interests include renewable energy, energy analytics, smart power systems, and electricity markets. He is currently the Technical Coordinator of the Horizon 2020 InteGrid project. He serves as an Editor for the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY.

• • •