Interpreting atypical conditions in systems with deep conditional Autoencoders: the case of electrical consumption

Antoine Marot[1], Antoine Rosin[1], Laure Crochepierre[1,2], Benjamin Donnot[1], Pierre Pinson[3], and Lydia Boudjeloud-Assala[2]

[1] Reseau Transport Electricite R&D, Paris, France
[2] Universite de Lorraine, CNRS, LORIA, F-57000 Metz, France
[3] DTU Technical University of Denmark

Abstract. In this paper, we propose a new method to iteratively and interactively characterize new feature conditions for signals of daily French electrical consumption from our historical database, relying on Conditional Variational Autoencoders. An autoencoder first learn a compressed similarity-based representation of the signals in a latent space, in which one can select and extract well-represented expert features. Then, we successfully condition the model over the set of extracted features, as opposed to simple target label previously, to learn conditionally independent new residual latent representations. Unknown, or previously unselected factors such as atypical conditions now appear well-represented to be detected and further interpreted by experts. By applying it, we recover the appropriate known expert features and eventually discover, through adapted representations, atypical known and unknown conditions such as holidays, fuzzy non working days and weather events, which were actually related to important events that influenced consumption.

Keywords: Interpretability · Autoencoder · Representation.

1 Introduction

1.1 Context

Well-established power systems such as the French power grid are experiencing a mutation with a steep rise in complexity. This is due to many factors, such as new consumer habits in the digital era with new usages relying on more controllable, individual and numerous appliances, as well as a necessary energy transition towards a greater share of renewable energy in the mix and better energy efficiencies to reduce our carbon footprint in climate change. This makes it harder to maintain the proper balance between production and consumption at all time, which is a necessary condition for power grid stability to avoid dramatic blackouts. More advanced predictive tools become necessary.

Classical load forecasting methods[1] previously heavily relied on seasonal and deterministic behaviors, modeled through expert features, but hardly grasped atypical and dynamical behaviors. Load analysis and forecasting, whether it is at individual level or a national level, is nevertheless a very dynamic field within the current energy transition era to eventually make smart grids happen, trying to overcome some remaining challenges with recent methods, as reviewed by Wang et al. [2]. Better understanding the new causal factors and the profiles of load consumption, handling their related uncertainties through probabilistic load forecasting [3], as well as dealing with bad and missing data for online predictions in real-time data streams [4], are three of the main research directions in this field. Our work will focus along the first research avenue of electrical consumption characterization, in interaction with experts, especially for new or atypical conditions that are under-represented in a dataset and have been hard to characterize until now, even when some instances were detected.

1.2 Industrial Challenge

Bank holidays, to which we will refer as "holidays" for short in this paper, have been known historical examples of such atypical conditions for electrical consumption, especially in France. A recent data challenge [4] organized by RTE was designed to address this issue of prediction under atypical conditions, with half of the test days being holidays. Machine learning models relying on xgboost or deep neural nets actually showed to perform better than RTE models overall on those atypical days. But they still each had some extreme errors on certain days and best models were different for each day tested. In addition, they were still merely black boxes, not giving many insights to the operators in charge, on the relevant factors for prediction and their effects, insights they could use to adjust the forecast with any new additional information. Eventually operators did not trust those new models on which RTE gave up for now. Trust and interpretability in models and applications are in fact prerequisites for operators responsible of a system in challenging situations: they will be asked for explanations if anything goes unexpected. More generally, beyond automatic method only, this highlights the need we will address here for renewed interpretability in models [9], through causal understanding, modeling and inference, which are essential for operators and humans to properly intervene and control any system [10].

In practice, expert operators spend a lot of time trying to identify the most similar holidays in the past to characterize and predict the consumption of a new holiday, while leaving the forecasting models predict more automatically over the typical days. Even doing so, the day-ahead forecasting error is still approximately of 1.5% Mean Absolute Percentage Error (MAPE) for holidays, reaching sometimes 3%, compare to below 1% MAPE for typical days. Predicting holidays is time-consuming because they do not have tools that give them adapted representations to study collectively those under-represented atypical signals. In addition, because new modes of consumption are appearing, atypical conditions, beyond holidays only, will be of greater importance to well predict.

[4] RTE Data Challenge: `https://dataanalyticspost.com/wp-content/uploads/2017/06/Challenge-RTE-Prevision-de-Consommation.pdf`

New tools to study and interpret them more efficiently with adapted representations are hence necessary.

1.3 Proposal

In data analysis and knowledge discovery, feature importance [5] and anomaly [7] or outlier [6] detection methods are often helpful to assist human experts. They have helped characterize and label some events for bike sharing systems [8] for instance, while limited in its depth of discovery beyond extreme events. However, they can be complemented with representation learning methods: besides looking at data statistically or individually, similarity-based representations let one investigate signal instances still collectively but also specifically and contextually. In that sense, our paper aims at highlighting the importance of learning adapted representations to let experts efficiently interpret underlying conditions in signals, even with simple feature importance and anomaly detection modules.

While deep learning methods have shown real promises in terms of predictive power, being successfully applied to power flow predictions in power systems for instance [11], they also have a potential to foster interpretability, beyond the black box effect, as illustrated by [12] in which they produce interesting clusters over representations learnt by a neural network. Indeed, deep learning can also be regarded as representation learning [13]. Word2Vec [14], and later Interligua [15], have been major illustrations of such interpretability power since in their latent representation, similarities and generic semantic relations (such as masculine and feminine or translations) between words were recovered. More generally, generative models, deep variational autoencoders in particular, are one family of representation learning models with recent interesting developments [16].

By compressing data signals in a latent representation, autoencoders (AE) implicitly capture the most salient features [17], with possible non-linear and mutual dependencies. To explicitly extract those features for interpretation, score to measure importance of existing expert features can be defined. To integrate and leverage those selected expert feature to discover deeper knowledge, we further consider Conditional Variational Autoencoders (CVAE) [18] which we review later in the method section to learn successive conditional representations. Whereas previous CVAE models mostly used as conditions simple target labels for anomaly detection, signal correction or inpainting [19] [20], one major technical contribution of our paper is, for the first time as far as we know of, to effectively learn a full conditional network module over a set of extracted features, to let expert discover new conditions in the residual latent representation.

The paper is organized as follows. First, we give an overview of the characteristics of electrical consumption with a specific focus on holidays. We then present our method based on CVAE to learn adapted representations. We define scores over features and instances to qualify those representations and extract knowledge from them. Finally, successive experiments demonstrate our ability to effectively learn such models and qualify the relevance of the representations to let expert interpret signals under unknown atypical conditions, and label them.

2 French electrical consumption: characteristics and data

National electrical consumption has been studied and its forecasting improved over the last few decades for power system operators to anticipate the required amount of energy production at every moment to match the demand in the system. Over the years, France has relied more heavily on electricity, given the development of big nuclear power plants which represent up to 75% of the total production, and incentives for electrical heating increased significantly the thermal dependency of electrical consumption. Weekday habits and temperature influence have been among commonly shared expert knowledge to predict electrical consumption. In addition, holidays have been known as atypical events within a year, shifting habits that are still hard to predict.

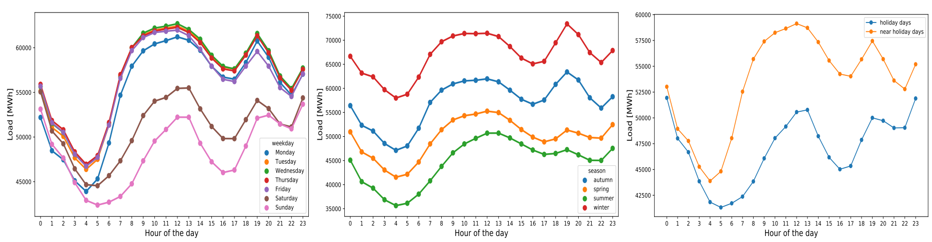


Fig. 1: From left to right, averaged daily load profiles at 30 minute resolution highlighting weekdays and seasonal patterns, and holiday atypical profile.

Common existing factors for daily electrical consumption and associated Expert Knowledge As shown on Figure 1, over the course of a day, the electrical consumption usually varies according to human activities, lowest in the middle of the night at 4am and higher during the day, with a peak either at noon or 7pm depending on the season. Within a week, the consumption is lower during the weekend, because of reduced working activities. Over the 5 weekdays, the profiles look similar but some differences can be noticed:

- The average load is less important on Monday than other week days on morning (from midnight to 2 pm). It is due to the *activity recovery effect*: after the weekend there is some inertia to retrieve a nominal activity level.
- The opposite is observed on Friday afternoon. The average load is less important than in the other weekdays (from 2 pm to midnight), as the activity tends to decrease before the weekend.

Weather, temperature in particular, are important factors as well for electrical consumption in France, with a thermal gradient of approximately 3% more consumption per 1°C less in winter. Nebulosity, wind, snow, humidity could also influence the consumption but their effect have been harder to characterize and are not yet considered in today's models. These factors are summarized on Figure 2 through a graphical model, a commonly used representation in the field of causality [10]. It sketches for a quick overview the main relations of dependence and independence between known factors, which we will actually recover through our first experiments to validate the proper learning of our CVAEs. While other methods can recover [21] those causal relationship between common factors when attributes are available in databases, we are mostly interested in events with unknown underlying conditions that might not be yet in any database and which will need additional human interpretation to be defined. These event relations are highlighted in dark, with specific characteristics that can influence the consumption differently.

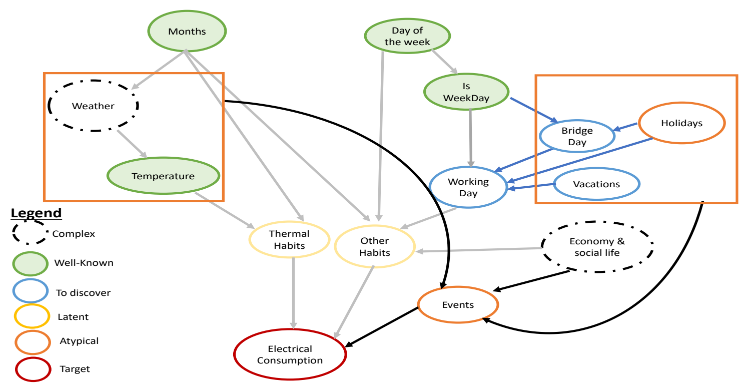


Fig. 2: Simplified graphical model. Well-known factors to integrate as expert knowledge appear as plain green, links to fuzzy atypical conditions we try to interpret appear darker, latent conditions we are also interested in are in blue.

Atypical Events One such event category which is well-known are holidays which will serve as a baseline for atypical condition discovery in our experiments. In France, there are 11 holiday days per year which sum to 55, or 3% of the data points in our dataset. They all have in common to be non-working days overall, but can differ in some ways as either days for national or religious celebration and commemorations. Most of them usually appear on fixed dates but not fixed weekdays, and some appear on fixed weekdays but with variable date spread over a month. This hence makes habits different for each holidays compared to a previous year which is even more challenging. When a holiday happens before a weekend, the behavior of consumers usually are more similar to a Saturday. On the following Saturday, the behaviors tend to be similar to a Sunday. For a holiday that happens at the beginning of the week on the contrary, it tends to have Sunday patterns. Often when a holiday happens on a Tuesday or Thursday, many people do not work on the Monday or Friday as well, but it is a fuzzier behavior that is not measured. Those are known as bridge days. Overall, those days look most similar to typical non-working day such as weekends, and some gives an opportunity for a short break, shifting the habits of a typical week.

Other events due to weather or social events can also affect the consumption. As we are looking to characterize a daily profile, weather events happening over a day or longer are likely to be influential and recovered. However, punctual social events often have shorter duration of only few hours and are less likely to affect the consumption over an entire day. Therefore, we will focus on first recovering holiday-like events, then on discovering weather-related events but we will let aside socio-economical events for now over this daily timescale.

Data The dataset covers the years 2013 to 2017 at a 30-minute resolution and was used for the RTE challenge in 2017. This sums up to 1830 daily data points. The temperature profile represents a weighted average over France computed by RTE. Table 1 gathers all the variables, binary, discrete as well as continuous, that will be used in our experiments. Daily Electrical Consumption Profile is our target variable of study. Temperature profile, day of the week, month of the years and holidays are possible features to characterize our electrical consumption. No missing data is reported, apart from the hour change event at the end of winter, which results in a fictitious additional hour with no data. The data at national level is considered clean since it has been used in production for many years with data quality processes, and was further used for the challenge.

Description	**Dimensions**	**Type**	**Formula**
Daily Consumption Profile	48	Quantitative	L_i, $i = 1, .., 48$
Daily Temperature Profile	48	Quantitative	T_i, $i = 1, .., 48$
Day of the week indicator (OH)	7	Categorical	W_i, $i = 1, ..., 7$
Month indicator (OH)	12	Categorical	M_i, $i = 1, .., 12$
Holiday indicator (OH)	1	Binary Unbalanced	H_i, $i = 0, 1$

Table 1: Summary table of variables in our dataset used in cvae. Categorical variables are One Hot (OH) encoded.

In the next two sections, we present our method to first learn dense representations with autoencoder models, and their conditional extension over extracted features, and later assess the quality of those representations with scores we define to retrieve expert features and extract new ones.

3 CVAE to learn conditional similarities over features

In this section we explain and motivate the choice of the model we use throughout the experiments performed in the last section. All these experiments share a common objective: representing the input data $\boldsymbol{x}$, daily consumption profile, by a more compact vector $\boldsymbol{z} = z(\boldsymbol{x})$. Especially, we want the representation $\boldsymbol{z}$ to reflect a notion of proximity: if two different days $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are encoded by respectively $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, and that $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are close together, in the sense of the l_2 norm, then we expect that the days represented by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ share some common features. Several methods can perform this transformation in the first place but not necessarily as deep by iteratively considering extracted features and adapting the representation. We choose to focus on Variational Autoencoder (VAE), first introduced in [27] and further on their Conditional extension to learn new specific representations z given some conditions c that we denote by $z_{|c}$.

Autoencoders An Autoencoder is a relatively simple model introduced in [22]. It consists of two parts, one call "encoder" Q (and parametrized by parameters θ) that will transform input $\boldsymbol{x}$ into its latent representation $\boldsymbol{z} = Q_\theta(\boldsymbol{x})$. It can be learned jointly in a completely unsupervised way with a decoder P (parametrized by vector ϕ) which takes the representation $\boldsymbol{z}$ and who's aim is to transform it back into the original representation $\boldsymbol{x}$. If we denote by $\hat{x}$ the output of the decoder, i.e. $\hat{x} = P_\phi(z) = P_\phi(Q_\theta(x))$, then the model is trained by minimizing the "reconstruction loss", which is a similarity measure between $\hat{x}$ and x. Most of the time Q and P are represented by deep neural networks. The Autoencoder has some drawbacks: no constraints are set on the latent representation $\boldsymbol{z}$, the only guarantee is that it can be decoded into the original signal x. Thus it is not always relevant to deduce some properties from the distance between two projections z_1 and z_2. This problem is partially solved by the VAE.

Variational Autoencoders VAEs aim at learning a parametric latent variable model by maximizing the marginal log-likelihood of the training data $\{x(i)\}_{1\leq i\leq N}$. Compared to the Autoencoder, it introduces a penalty on the latent space. This latent space z is seen as probabilistic distribution, and must be close to a given prior distribution. It is part of the "variational" literature and is nowadays mostly used to generate data. In this paper, we are interested in the property of the latent space z and will not use the generating capabilities of VAE. Adding this constraint on the latent space has the beneficial effect of regularizing it. In this framework the encoder Q_θ and decoder P_ϕ are better seen as probabilistic distribution, and we will note: $Q_\theta(z|x_i)$ the distribution of variable z given input data x_i. $P_\phi(x_i|z)$ will be the distribution of the reconstructed vector $\hat{x}_i$ from its latent representation z by the decoder P. Training this network is then equivalent to adjust parameters θ and ϕ under the loss:

$$\mathcal{L}_{\lambda-VAE} = \underbrace{\mathcal{L}_{recon}}_{\text{reconstruction loss}} + \lambda \cdot \underbrace{\mathcal{L}_{KL}}_{\text{divergence loss}} \tag{1}$$

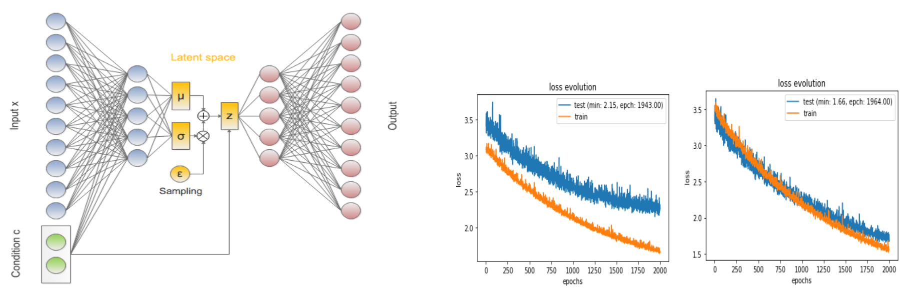
$$= -E[\log P_\phi(x_i|z)] + \lambda \cdot \mathcal{D}_{KL}(Q_\theta(z|x_i)||P(z)) \tag{2}$$

The reconstruction loss measure how well the reconstructed vector $\hat{x}_i$ is close to the original representation x_i, as in the vanilla Autoencoder. The divergence loss minimizes the KL-divergence between $Q_\theta(z|x_i)$, *i.e.* the distance between the latent distribution and its target distribution (the normal distribution usually).

CVAEs Lastly, we also want to learn adapted representations given existing knowledge. Conditional Variational Autoencoders (CVAE) [18] are an extension that enable to bypass the latent space with some conditional information, such as previously extracted features in our case, to be used for signal reconstruction, freeing the encoder from encoding such information while still achieving proper reconstruction. Figure 3 shows a schematic of this model. The adapted loss function used for the training is:

$$\mathcal{L}_{\lambda-cVAE} = -E[\log P_\phi(x_i|z_{|c}, c)] + \lambda \cdot \mathcal{D}_{KL}(Q_\theta(z_{|c}|x_i, c)||P(z_{|c})) \tag{3}$$

Note that in this case, the conditioning variables represented by vector c is given as input to both the encoder Q_θ and the decoder P_ϕ. It is an architecture that is used to here explicitly disentangle known factors c from other latent residual factors in $z_{|c}$. We will later assess if they factorize properly. In supplementary materials, we explain how we were eventually able to learn properly CVAEs, which are notoriously hard to train [25], [24], [23], more especially in our case where we consider training a full conditional network module over extracted features, and not solely inputting a conditional vector over simple target labels.



(a) CVAE architecture (b) Learving curve with and without BN

Fig. 3: Example of CVAE with 2 hidden layers: the architecture is similar to a VAE, except for the additional conditional embedding vector c. Adding a batch normalization (BN) layer unlock the proper learning of the network module for feature conditional embedding to avoid overfitting, as explained in annex.

4 Scores to recover and discover knowledge

We here define scores to assess quantitatively the nature of the representations.

Scores to assess feature importance in latent space z The goal of our method is to extract and further discover some knowledge embedded in the latent space representation we learnt. To assess knowledge recovery of known expert features in the latent space, we can assign them a prediction score to figure out if those features were implicitly selected when compressing information. All the scores are based on a local knn predictor model, either classifier for categorical and binary variables or regressor for continuous variables, using standard euclidean distance metric and a default 5 nearest neighbors. In more details, here is the list of scores used in our experiments for our features of interest apriori:

- F_W and F_M, the scores for day of the week W and months M which are simply the fraction of correctly predicted samples for balanced categories.
- F_{is_w}, the score for *is_weekday* variable is the fraction of properly predicted samples but reweighted, since this is an imbalanced variable (5 weekdays compare to 2 weekend days in a given week). This score is known as an F1 score in the literature, as also defined in sklearn library.
- F_H the score for holidays. Since it is an atypical event variable, we are mostly interested for the instances it occurs. Our score is hence the fraction of true positive holidays only.
- $F_{T_{mean}}$ the score for temperature. For temperature daily profile feature T, we first consider a proxy variable T_{mean} of the average temperature over a day. We then learn a knn regressor for it and use the R2 score over it.

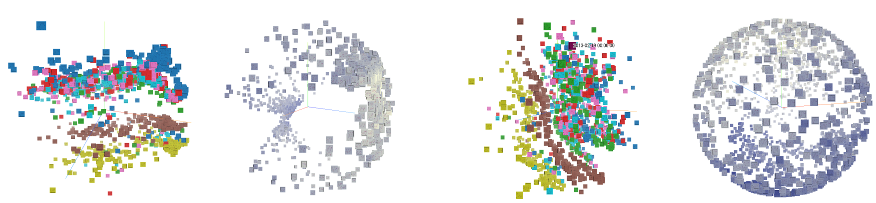
The highest possible value for each score is 1. Scores close to 1 are always a good indicator of a strong dependency of the latent space with the feature associated to the score, hence of its importance. On the contrary, scores of a random projection, give a lower bound for the score to be expected from a feature that is independent from the latent projection, hence insignificant in the encoding.

This feature importance approach could be further extended to be used in automatic feature selection [5]. The goal of our work however here is rather to assess how informative and interpretable a learnt conditional representation is for an expert, rather than defining a new automatic feature selection method.

Score to detect event as a local outlier in the latent space Once we decided which knowledge to extract and integrate as conditions in the CVAE model, we can train our network in order to obtain a disentangled latent representation of our data. Because of its independence towards the selected feature conditions, this representation also allows us to look for abnormal samples with respect to those conditional features, to guide our exploration in discovering new knowledge. For instance, when conditioning by the day of the week W and the holiday H variables, we expect our representation to be strongly guided by other unselected variables like temperature. What we want to do now is to look for events that are not well represented by these conditions: in other words, outliers or atypical conditions in our representation. To do so, similarly to [4], we decided to use for now a k_{th} nearest-neighbour based outlier score as defined in [28][29]. For the sake of simplicity we use here the 1-nn as our first goal is not to identify systematically outliers but rather study and interpret the representation. Once detected, we want to understand their context and eventually detect if ensemble of atypical events, explained by common factors, are close in the latent space.

5 Experiments to learn adapted representations

Our goal in the following experiments will be to demonstrate our ability to learn successively more specific representations between signals in our dataset with AEs that help reveal some new knowledge to experts. While recovering expert knowledge, we further want to integrate it to enable a more comprehensive exploration than previously possible of other latent features through a similarity space, especially atypical conditions. In our first experiment, we aim at recovering common expert features while highlighting the difficulty at first of learning relevant representations for known atypical signals during holidays. In a second experiment, we show that we can actually learn specific residual representations after conditioning over full sets of extracted features with CVAEs, one main technical contribution of our paper compared to previous CVAEs conditioning on simple target labels. In the third experiment, we focus on the representation most adapted to holidays and explore it to recover some knowledge about them and discover other unexpected similar atypical days. Finally, we explore in a fourth experiment unknown weather atypical conditions, after learning a new representation given previous knowledge over daily features, taking into account atypical conditions from the third experiment. Supplementary materials, code, data and interactive representations are available on GitHub [5]. We explain in the annex our choice of parameters, we never tuned, and hyperparameters, λ and number of training epochs, that we explored and selected.



(a) PCA 3D visualizations & features (b) VAE 3D visualizations & features

Fig. 4: Similarity-based representations with day of the week label on the left, and max daily temperature label on the right over a spherized projection in tensorboard. Data points are more homogeneously spread in VAE representation which makes it easier to navigate into similar or outliers examples.

[5] GitHub for paper, annex, data, code and interactive visualizations: https://github.com/marota/Autoencoder_Embedding_Expert_Caracteristion_

Experiment 1: embedding and expert features recovery In this first experiment, we study 2 models, a classic PCA and a simple VAE, and compare them to a random baseline to measure their respective abilities to learn similarity-based representations and recover common features and dependencies that are described in Table 1: day of the week, weekday or weekend, month of the year, temperature and holidays. First we figured out that most of the variance in the dataset is retrieved on the first 3 dimensions in the case of PCA. 3 dimensions are actually significant as well in the VAE latent space. When learning with more than 3 dimensions, the other dimensions are pruned out with L1-norm in reconstruction Loss. We eventually select a dimension of 4 for our VAE latent spaces in all our experiments to leave it some more freedom during training. Qualitatively, the VAE projection looks easier to study interactively for a human expert since it is more homogeneously spread, given the gaussian prior, as illustrated on figure 4 and better visualized on GitHub. We could expect a human expert to better navigate and study similarities through examples.

Figure 5 summarizes the results of the first experiment. Quantitatively, both models gives very similar scores. As F_{is_w} and F_{Tmean} scores are equally close to 1 for PCA and VAE, they are both able to recover the most salient features to represent daily profiles: temperature and *is_weekday*. Month of the year and day of the week variables also appear as significant features, but they seem less expressive as their scores F_M and F_W are close to 0.5. While weekday and weekend have quite different consumption patterns as illustrated by F_{is_w} score, weekdays have a lot in common which explains the score for day of the week F_W. In the same way, consumption patterns are rather dependent on seasons than individual month, and successive months hence share common consumption patterns decreasing its score.

Eventually, even if holiday score F_H is greater than random, which can suggest some dependency of the consumption over this feature, it is lower than any other features. When looking manually at the representation of holidays in the latent space, we still recover that holidays look more like weekends (in between Saturdays and Sundays) than weekdays. But no holiday appear similar to one another. It is hard yet to ascertain the existence of a condition. The representations created by the two models in Experience 1 hence do not seem appropriate to the analysis of the influence factors of consumption on holidays. To study them properly and independently of common influent factors, it is thus necessary to learn a new specific representation, as we present in Experiment 2.

Experiment 2: Conditioning over existing knowledge to learn new representations to be explored In this second experiment, we analyze how CVAE models can create more specific and selective representations when conditioning over features we do not want to be influenced by in our exploration. To first assess the quality of the learnt representation, we try in the first place to recover the knowledge described in Figure 2. In a second step, we will show how we are able to learn a representation more relevant to the analysis of holidays.

Figure 5 summarizes the results of learning different conditional models over known causal factors, either day of the week W, month M and/or temperature T. For all models, residual latent spaces effectively appear to be quite independent of conditional factors, as shown by scores highlighted in green, which are getting close to random. For instance in $CVAE_{|\{T\}}$ model, we condition over temperature T and the corresponding F_T score is near random, highlighting the latent representation independence over temperature. As another result, we see that day of the week condition does not affect the dependency over month M and temperature T in the latent space in model $CVAE_{|\{W\}}$ and conversely in $CVAE_{|\{M,T\}}$. We thus retrieve as a sanity check the natural independence of those factors like previously described in Figure 2. In addition, we see a simultaneous arising independence of the latent space from $is\_weekday$ variable and W in $CVAE_{|\{W\}}$. This highlights the obvious dependency of $is\_weekday$ over day of the week W. We observe the same but little obvious dependency between M and T. This suggests that beside temperature effects, there might not be strong shift in daily consumer habits from one month to another.

Finally, our known atypical feature, holidays, is eventually well-represented with a high F_H score in the latent spaces of $CVAE_{|\{W\}}$ and $CVAE_{|\{W,M,T\}}$, when conditioning over day of the week W, since they have competing dependencies in common. As a matter of fact, holiday happening on a weekday shifts usual weekday habits to non-working day habits most similar to weekend. Holidays hence appear to be very atypical from the expected consumption prototype on a given weekday. This gives us a new residual latent space in which to study holidays that are well-represented to eventually validate this atypical condition as a relevant feature.

Model Type	Conditions	Conditional layer size before embedding	Feature Scores				
			F_{is_w}	F_W	F_M	F_H	F_{Tmean}
Experiment 1							
random	{ }	[ ]	$0.5_{\pm 0,01}$	$0.14_{\pm 0,01}$	$0.08_{\pm 0,01}$	$0.04_{\pm 0,01}$	$-0,4_{\pm 0,03}$
PCA	{ }	[ ]	$0.96_{\pm 0,01}$	$0.51_{\pm 0,01}$	$0.56_{\pm 0,01}$	$0.23_{\pm 0,01}$	$0.91_{\pm 0,01}$
VAE	{ }	[ ]	$0.96_{\pm 0,02}$	$0.50_{\pm 0,02}$	$0.56_{\pm 0,02}$	$0.27_{\pm 0,03}$	$0.90_{\pm 0,02}$
Experiments 2 & 3							
CVAE	{T}	$Dims_T$=[48,12, 4]	$0.96_{\pm 0,02}$	$0.56_{\pm 0,02}$	$0.19_{\pm 0,03}$	$0.56_{\pm 0,03}$	$-0.09_{\pm 0,1}$
CVAE	{M}	$Dims_M$=[12,6,3]	$0.95_{\pm 0,02}$	$0.45_{\pm 0,04}$	$0.12_{\pm 0,02}$	$0.13_{\pm 0,03}$	$-0.10_{\pm 0,1}$
CVAE	{M, T}	[$Dims_M$, $Dims_T$]	$0.96_{\pm 0,02}$	$0.57_{\pm 0,02}$	$0.13_{\pm 0,02}$	$0.22_{\pm 0,03}$	$\mathbf{-0.31_{\pm 0,07}}$
CVAE	{W, M, T}	[$Dims_W$,$Dims_M$, $Dims_T$]	$0.57_{\pm 0,02}$	$0.18_{\pm 0,02}$	$0.12_{\pm 0,02}$	$\mathbf{0.79_{\pm 0,02}}$	$-0.23_{\pm 0,07}$
Experiment 4							
CVAE	{W}	$Dims_W$=[7,5,3]	$0.54_{\pm 0,02}$	$0.18_{\pm 0,02}$	$0.48_{\pm 0,03}$	$\mathbf{0.75_{\pm 0,02}}$	$0.89_{\pm 0,02}$
CVAE	{W, H}	[$Dims_W$, [1,2]]	$0.55_{\pm 0,02}$	$0.15_{\pm 0,02}$	$0.45_{\pm 0,03}$	$0.10_{\pm 0,02}$	$0.87_{\pm 0,02}$
CVAE	{W, H+}	[$Dims_W$, [1,2]]	$0.57_{\pm 0,02}$	$0.16_{\pm 0,02}$	$0.50_{\pm 0,03}$	$0.06_{\pm 0,01}$	$0.89_{\pm 0,02}$

Fig. 5: Table of feature scores given several models with different conditions. Experiment 2 highlights that conditioning is properly learnt (green). Holidays appear as a new important feature (blue) when conditioning over weekdays.

Experiment 3: Exploring similar holidays & discovering additional weekday events After demonstrating proper conditional learning over extracted features, we will now focus on $CVAE_{|\{W,M,T\}}$ to deepen our knowledge over holidays. Since our goal is to recover consumption behaviors specific to these peculiar days, we not only conditioned on weekdays, but also on month and temperature, to focus on unknown specific factors beyond existing knowledge.

In this conditional representation, we reach a feature importance score F_{is_w} of 0.79 for holidays, which is a lot higher than for previous unconditioned representations. Theses results indicate that this CVAE model is suitable to study the peculiarity of most holidays. However, some holidays are not well predicted and we first need to understand why. When analyzing them, most are actually occurring on weekends. This result is interpretable since weekends are already non-working days, hence not really shifting consumer behaviors and thus not atypical: this is a well-known fact we recover. Without integrating expert knowledge over weekdays, this fact could not be recovered in the first unconditioned VAE representation or in $CVAE_{|\{T\}}$, $CVAE_{|\{M\}}$ and $CVAE_{|\{M,T\}}$. The only exception to this statement is a Christmas day happening on a Saturday which actually appears as different from a typical Saturday. This is understandable for this very particular day with huge celebrations. For the remaining 2 holidays not well predicted and happening during weekdays (2015-05-08 and 2016-11-11), they are actually similar to non-working days surrounding holidays as we explain in the following paragraph. We here explained all the instances that were supposedly not well-predicted, highlighting the power of such a representation to study instances collectively in context rather than some model limitation.

Figure 6 shows the conditional latent representation, which confirms that holidays appear more similar to one another than to other days, as they are clustered in the latent space. We then used a manual semi-expert exploration step with Tensorboard Embedding Projector [30] to create new categories of days. In turquoise, we discover 27 days at once in a similar location of the representation which happened during Christmas weeks and during which the great majority of people take vacations: we quickly identified a shared underlying phenomenon as a new condition. As they are also non-working days, this makes them similar to holidays. In green, we discover and label 17 “bridge” days, which are days happening between a holiday and a weekend. Bridge days are often non-working days, with the opportunity for people to take a 4-day break. However, it is not always the case for everyone and every company, leading to a fuzzy mix of working and non-working day behaviors which are hardly measured otherwise. Finally, an exception is the 6th May of 2016, which was actually a bridge day and is not labeled yet. We will see later in the last experiment that it is due to a conjunction of conditions, and not an error in learning this representation. Defining all these new labels interactively and efficiently, 44 in totals related to holidays, demonstrate how informative this representation is for semi-experts to study the characteristics of atypical conditions like holidays, and of days with similar characteristics as well.

In a last step, we looked at the 10 most atypical days based on our outlier score. 4 days were already identified as non-working days previously. The 6 others can actually be interpreted as weather events: 2017-01-21, 2017-01-28, 2013-03-04, 2013-03-13, 2013-03-11 which were all important snowy days in France and 2013-04-14 which was equivalent to a punctual summer day with high temperature gradients from the day before and after. However, this representation is better suited to discover daily events similar to holidays than weather ones. To study further atypical weather conditions independent of daily behaviors, we will explore a last conditional representation over weekdays W and over those new labels related to holidays $H+$, deepening our knowledge by building on it.

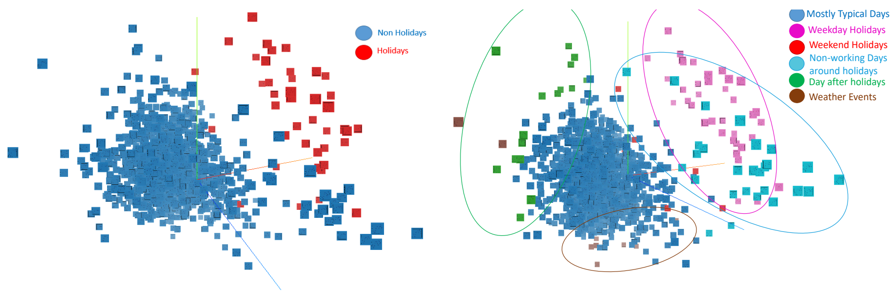


Fig. 6: Latent representation from $CVAE_{|\{W,M,T\}}$, before and after expert labelling. Holidays during weekdays are identified as similar and other non-working days are also discovered. In addition, first weather events are discovered.

Experiment 4: discovering new weather events In this last experiment, we want to explore how weather-related events can be represented in a more suited latent space. We first learnt the $CVAE_{|\{W\}}$ model to remove the weekday effect as illustrated in Figure 7 but this representation only highlighted holidays not yet integrated. As a result, we decided to condition, not only on weekdays W but also on the knowledge of the holidays H. In this representation, non-working days could still be predicted and we decided to include all the labels of Experiment 3 to even more properly condition our latent representation over daily effects. The only remaining working days still predicted were during the Christmas week of 2015, and it is understandable for a weather perspective since it was the hottest Christmas week in French history, hence sharing an additional atypical condition. The resulting weather representation is shown in Figure 7 (middle).

After creating a representation which qualitatively makes sense to explore the temperature factor, we tried to locate in it the previously detected weather-related events.From Experiment 3, we observed that some previous weather events were actually clustered in this new latent space. We discovered that 2017-01-21 and 2017-01-28 were part of a cold snap starting mid-January and lasting

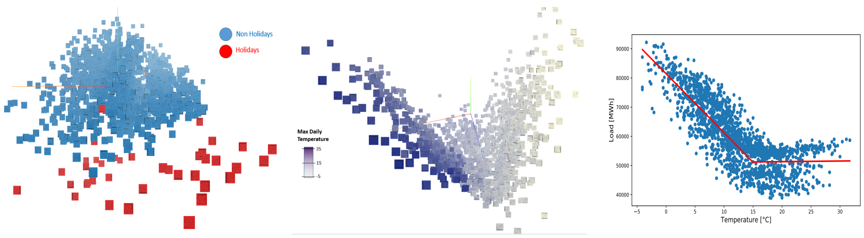


Fig. 7: Left, latent representation from $CVAE_{|\{W\}}$. Holidays are recovered as an important feature here as well and hide the temperature dependency. When conditioning additionally over non-working days with $CVAE_{|\{W,H+\}}$, a smooth 3D V-shape appears, similar to a scatter plot consumption vs temperature.

until the end of the month rather than just punctual snowy days: by reusing some previous discovery from a previous representation we here confirmed the underlying condition and strengthen our knowledge. As for the other snowy days mentioned in Experiment 3, they were all surrounded by other snowy days at different times, which let us identify new labels.

In order to discover other weather-related atypical events, we looked for the top-100 outliers. As we could expect, some of them had already been manually identified as snowy days. Bridge day 6th of May 2016 also appears as a strong outlier here, indeed associated with a rare dry wind event from the Sahara over France increasing consumption, explaining why it was not strongly detected as a non-working day before with a lower expected consumption. Finally we also discovered recurring hot periods in August accounting for almost a quarter of the top-100 outliers. In this CVAE model, almost all August days between the 7th and 28th of August (except for the 15th, a bank holiday) appeared as atypical. We believe it is due to another underlying feature of interest, not taken into account yet in our data: the significant proportion of employees taking a two-week vacation in August. A new representation, conditioned on temperature additionally, might be interesting to explore in the future to study remaining monthly characteristics.

Across all these experiments, we have explored how these conditional representations could help an expert improve its intuitions on the influent factors for consumption in an interpretable and iterative process, building and strengthening knowledge iteration after iteration. We have first recovered existing expert knowledge, which can be seen as a functionally-grounded evaluation in the taxonomy of [9]: a first level of interpretability. We further showed that we could explore specific representations to discover new events and interpret them as non-working days and weather characteristics. Such experiments can be used in the future for a human-grounded evaluation, the second level of interpretability.

6 Conclusion

We showed how CVAEs could actually be used to recover existing expert knowledge and further learn specific representations for atypical conditions discovery in electrical consumption. This helped study those peculiar situations collectively to eventually interpret quickly some latent additional conditions to augment expert knowledge. In particular, we recovered holidays and their characteristics and discovered similar non-working days as daily events. We eventually detected unknown influential weather events and interpreted them in the appropriate representation. New time scales could be explored and our method improved with more specific architectures such as temporal convolution or attention-based ones. New scores for atypical condition detection could also be used for an even deeper exploration. More generally, given the ability of neural nets to deal with many kinds of data, we believe our approach could be applied more generically to other systems. It could finally be integrated in new iterative and interactive tools for experts [26], to help them explore, interpret and label more exhaustively specific cases of interest within relevant representations of their data.

References

1. A. K. Singh, Ibraheem, S. Khatoon, M. Muazzam and D. K. Chaturvedi, "Load forecasting techniques and methodologies: A review," Power, Control and Embedded Systems Conference (ICPCES), 2012
2. Wang, Y., Chen, Q., Hong, T., & Kang, C. (n.d.). Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. Journal IEEE TRANS. SMART GRID, 2018
3. ao Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, R. J. Hyndman, Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond, International Journal of Forecasting, Volume 32, Issue 3, July–September 2016
4. R. Corizzo et al., Anomaly Detection and Repair for Accurate Predictions in Geo-distributed Big Data, Big Data Res. (2019)
5. Jundong Li, Kewei Cheng and al. 2017. Feature Selection: A Data Perspective ACM Comput. Surv. 9, 4, Article 39 (March 2010)
6. H. Kriegel, P. Kröger, A. Zimek, Outlier detection techniques, Tutorial 665 KDD10.
7. V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, ACM Comput. Surv. 41 (3) (2009)
8. H. Fanaee-T and J. Gama. Event labeling combining ensemble detectors and background knowledge. Progress in AI, 2(2-3):113–127, 2014.
9. Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. Retrieved from http://arxiv.org/abs/1702.08608
10. J. Pearl. Causality. Cambridge university press, 2009
11. I. Guyon B. Donnot, A. Marot and al. Introducing machine learning for power system operation support. IEEE IREP conference, 2017
12. Xie, J., Girshick, R., & Farhadi, A. (2015). Unsupervised Deep Embedding for Clustering Analysis. ICML 2016
13. I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. 2016. MIT Press.
14. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed Representations of Words and Phrases and their Compositionality. NIPS 2013.

15. M. Johnson, M. Schuster & al., Google's multilingual neural machine translation system: Enabling zero-shot translation. CoRR, abs/1611.04558, 2016
16. Tschannen M, Zurich E and al. Recent Advances in Autoencoder-Based Representation Learning, Workshop on Bayesian Deep Learning NeurIPS, 2018
17. G. Hinton, R. Salakhutdinov, Reducing the dimensionality of data with 675 neural networks, science 313 (5786) (2006)
18. D. P Kingma, S. Mohamed, D. J. Rezende, and M. Welling. Semi-supervised learning with deep generative models. NIPS 2014
19. Fan, C.; Xiao, F. and al. Analytical investigation of autoencoder-based methods for unsupervised anomaly detection, in building energy data. Appl. Energy 2018.
20. M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in iot," Sensors, vol. 17, no. 9, p.1967, 2017
21. Esposito F., Malerba D., Ripa V., Semeraro G., "Discovering Causal Rules in Relational Databases", Cybernetics and Systems'96, Studies, 1996.
22. Cheng-Yuan Liou, Jau-Chi Huang, Wen-Chie Yang. Modeling word perception using the Elman network. Journal Neurocomputing, 2008
23. Dieng, A. B., Kim, Y., Rush, A. M., & Blei, D. M. (2018). Avoiding Latent Variable Collapse with Generative Skip Models, AISTATS 2019
24. C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in beta-VAE, NIPS 2017
25. Yeung S., Kannan A., Dauphin Y., Fei-Fei Li, Tackling Over-pruning in Variational Autoencoders, ICML Workshop on Principled Approaches to Deep Learning, 2017
26. L. Boudjeloud-Assala, P. Pinheiro, A. Blansch, T. Tamisier, B. Otjacques. Interactive and iterative visual clustering. Information Visualization, 15(3):181–197, 2016.
27. Kingma, Diederik P and Welling, Max. Auto-encoding variational bayes. 2013.
28. S. Upadhyaya, and K. Singh. Nearest neighbour based outlier detection techniques. International Journal of Computer Trends and Technology 3.2 (2012): 299-303.
29. S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. ACM Sigmod Record. Vol. 29. No. 2. ACM, 2000.
30. D. Smilkov, et al.: Embedding projector: Interactive visualization and interpretation of embeddings., arXiv preprint arXiv:1611.05469 (2016).