A novel method for decomposing electricity feeder load into elementary profiles from customer information

Alexis Gerossier, Thibaut Barbier, Robin Girard

A novel method for decomposing electricity feeder load into elementary profiles from customer information

Alexis Gerossier* Thibaut Barbier* Robin Girard

{alexis.gerossier},{thibaut.barbier}@mines-paristech.fr
* Both authors equally contributed to this paper
MINES ParisTech, PSL Research University, Center for processes, renewable energy and systems (PERSEE), 1 rue Claude Daunesse, 06904 Sophia Antipolis, France

June 2017

Abstract

To plan a distribution grid involves making a long-term forecast of sub-hourly demand, which requires modeling the demand and its dynamics with aggregated measurement data. Distribution system operators (DSOs) have been recording electricity sub-hourly demand delivered by their medium-voltage feeders (around 1,000—10,000 customers) for several years. Demand profiles differ widely among the various considered feeders. This is partly due to the varying mix of customer categories from one feeder to another. To overcome this issue, elementary demand profiles are often associated with customer categories and then combined according to a mix description. This paper presents a novel method to estimate elementary profiles that only requires several feeder demand curves and a description of customers. The method relies on a statistical blind source model and a new estimation procedure based on the augmented Lagrangian method. The use of feeders to estimate elementary profiles means that measurements are fully representative and continuously updated. We illustrate the proposed method through a case study comprising around 1,000 feeder demand curves operated by the main French DSO Enedis. We propose an application o that uses the obtained profiles to evaluate the contribution of any set of new customers to a feeder peak load. We show that profiles enable a simulation of new unmeasured areas with errors of around 20%. We also show how our method can be used to evaluate the relevancy of different customer categorizations.

1 Introduction

1.1 Motivation

Electricity represented 18% of total final energy consumption in 2013 [3] and is expected to constitute a quarter of final energy consumption by 2040 [1]. 42% of global CO2 emissions in 2012, *i.e.* 13.8 gigatons of CO2, are due to electricity and heat production [2]. To reduce CO2 emissions due to electricity, many states are developing energy transition strategies. This kind of transition involves significant changes to electricity flows in the distribution network (with *e.g.* decentralized production, improved efficiency of buildings and appliances, new uses and demand

Nomenclature

a^f	Consumption trend of feeder f relative to temperature
b^f	Temperature threshold of feeder f
B	Matrix of demand profiles of customer categories
β	Column vector associated with B
c_k^f	Annual consumption of category k for feeder f
d^f	Demand of feeder f
d_k	Elementary profile of customer category k
ε^f	Residual term for modeling feeder demand f
F	Number of feeders
f	Feeder
K	Number of customer categories
k	Customer category
m_k	Average demand share of a given category k
p_k^f	Share of electricity used by category k for feeder f
σ_k^2	Empirical variance of $p_k^1, \ldots, p_k^F$
T	Number of instants
t	Instant
T^f	Outside temperature of feeder f
u	Vector $(1, \ldots, 1)^\intercal$ of length K
v	Vector $(T^{-1}, \ldots, T^{-1})^\intercal$ of length T
V_{inter}	Inter group variability
V_{tot}	Total variance
X	Matrix of feeder demands
x	Column vector associated with X
y	Year
$\otimes$	Kronecker product

response enabling energy consumption management [18]).

These changes impact the planning process of distribution system operators (DSOs). The current network planning process considers the two most extreme situations [16], *i.e.* maximum demand with minimum supply, and maximum supply with minimum demand. While planning with such a method does not require a deep modeling of the different dynamics and their correlations, it does not take into account the aggregation effect between supply and demand [15]. The above-mentioned changes make it necessary to model all of the aggregated demand dynamics.

1.2 Literature review

In this section we present two kinds of existing approach for modeling aggregated demand. The first is bottom-up, and uses individual customer profiles, which are summed to obtain aggregated demand. The second is a global approach in which the aggregated load curve is directly modeled using aggregated measurement data.

1.2.1 Bottom-up approaches

Measuring the electricity demand of individual electricity customers is a simple way to establish their load profiles and dynamics, and therefore a necessary step in bottom-up modeling. The current smart-meter roll-out in Europe will provide precise measurements of individual demand profiles. Around 80% of customers are scheduled to receive a smart-meter by 2020 [28]. However, this massive deployment is hindered by cost and privacy issues [21]. In 2014, only 23% of smart-meters in the European Union were installed in localized areas for private customers [13]. In some countries, this share is still insufficient to be representative, and the corresponding deployment is too recent to adequately cover long periods. To deal with the lack of individual measurements and characterize the behavior of electricity customers, researchers have attempted to classify them into different categories.

The classification of electricity demand profiles is a flourishing research topic (see reviews [19], [25]). Researchers use individual measurements from smart-meters as input and apply different clustering methods [31]. This reduces the dimension, which makes it easier to manipulate data [22]. With the resulting classification, each customer is associated with a cluster and its corresponding load profile [26]. The classification and the obtained load profiles can be used for a number of applications.

First, a fine classification can be made in order to help decision-makers design personalized policies for specific customers [7].

Secondly, the classification allows a DSO to plan its network and anticipate its investments [23, 27]. For example, the French DSO uses a model named "Bagheera" combining about 50 customer categories to plan its low-voltage network [16]. Classification is combined with the evolution of category distributions to forecast aggregated demand in prospective scenarios [5].

Last, classification and load profiles allow us to understand the contribution made by each category to aggregated demand [27].

Large measurement campaigns are necessary with these methods since a representative set of customers is required. This constraint makes continuous updating of the profiles difficult, which is an issue since it remains necessary to adapt the profiles to the changing consumption habits [4, 26].

1.2.2 Global approaches

In global approaches, models forecast aggregated electricity demand with past measurements and explanatory variables, such as expected temperature or sometimes economic progress [30].

In order to obtain past measurements, most DSOs have been recording the electric power delivered by their medium-voltage feeders (around 1,000—10,000 customers) for several years. These measurements are aggregated, but exhaustive, since all electricity customers' contributions are taken into account. This aggregated electricity demand data is considered as a "nonlinear, non-stationary series, and is often made up by a superposition of several distinct frequencies" [29] with daily to monthly periods in global models [8]. Additionally, the demand series can be divided into different parts (*e.g.* working time, holidays) [9, 17].

The global approach produces accurate forecasts. However, these are based on aggregated past measurements, which are not available when planning a new unmeasured zone. This type of planning is improved with specific information about customers, which DSOs possess thanks to the Customer Information System (CIS) [23]. The CIS stores information on all customers regarding their electric connection to the grid, annual energy consumption, type of contract, and contracted power.

In all of the reviewed global methods [29] for modeling demand dynamics, the explanatory variables used, such as expected temperature or sometimes economic changes [30], do not characterize the feeder-specific local features. In particular, none of them employs CIS general statistics.

Finally, the drawback of these methods when used for planning purposes is that they cannot adapt to a change in the mix of customer categories. For example, in the case of the development of a commercial area in a residential feeder, such methods fail to take into account the corresponding information. If the profile differences of the two sectors is not accounted for, this might result in an overestimation of the future peak and hence an over-sizing of the network.

1.3 Contributions

Our paper presents a novel method to estimate elementary profiles. The proposed method relies on a statistical model that takes into account the mix of customer categories. To do this, we assume that the demands aggregate different shares of elementary profiles associated with different customer categories. These profiles are optimally found by minimizing prediction errors in a new algorithm relying on the augmented Lagrangian method.

Unlike bottom-up methods, our method only requires several feeder demand curves and a description of customers. The advantages of aggregated measurements compared to a set of individual load curves are: the availability of long-term historical data, full representativeness, and continuous updates. We show that the method performs similarly or better than a bottom-up method in the literature when predicting new local areas.

We illustrate the proposed method through a case study comprising around 1,000 feeder demand curves operated by the main French DSO Enedis. The profiles obtained are essential to size the distribution network. This is illustrated by an application that evaluates the contribution of any set of new customers to a feeder peak load. We show that profiles enable a simulation of new unmeasured areas with errors of around 20%. We also show how our method can be used to evaluate the relevancy of different customer categorizations.

1.4 Description of the paper

In section 2, the methodology is described. A case study is presented in section 3 with the resulting profiles by category. Section 4 describes two applications that use the obtained profiles. One is employed to estimate the contribution of set of new customers to a feeder peak load. The other evaluates forecasting errors for unmeasured areas, by testing different categories and comparing performances with a similar framework case study in the literature. Finally, some conclusions are presented and discussed in section 5.

2 Methodology

2.1 The problem of recovering load profiles and the forecasting method

Our paper assumes that the sub-hourly demands $d^f(t)$ of a feeder f aggregate different profiles $d_1(t), \ldots, d_K(t)$ associated with K categories of customers with weights $p_1^f, \ldots, p_K^f$,

$$d^f(t) = \sum_{k=1}^{K} p_k^f d_k(t) + \varepsilon^f(t). \tag{1}$$

We take the elementary profiles $d_k(t)$ to be common to all feeders, while the weights vary from one feeder to another. The corresponding residual term $\varepsilon^f(t)$ is meant to be small. The time t can vary along any set. The aim is to recover unknown elementary electricity profiles $d_k(t)$. For each feeder $f \in \{1, \ldots, F\}$, $d^f(t)$ is observed and, thanks to the CIS, for each category $k \in \{1, \ldots, K\}$, we also have access to the weight p_k^f.

The process of obtaining proportions from the CIS and defining categories is the categorization step, and is described in subsection 2.3. Once the K profiles have been obtained on a set of feeders, it is possible to turn Equation (1) into a simulation algorithm. The process is described in Figure 1. In the signal processing community, the corresponding problem is called blind signal separation and is well-known (see *e.g.* [11]).

2.2 Optimization problem

The aim is to find the elementary profiles $d_k(t)$ from aggregated demand $d^f(t)$ according to Equation (1). We write and solve the following optimization problem.

To mathematically write this optimization problem, we define a matrix A of size (F, K) whose elements are proportions p_k^f for $k \in \{1, \dots, K\}$ and $f \in \{1, \dots, F\}$. Aggregated demands $d^f(t)$ for all feeders and instants $\{1, \dots, T\}$ are gathered in a matrix X of size (F, T). We are trying to compute demand profile $d_k(t)$ for all categories and instants: these unknown values can be put in a matrix B of size (K, T). It is useful to define β (resp. x), the column vector obtained by stacking rows of B (resp. X) on top of each other. Two constraints limit the values of matrix B:

1. Each component of β is an electricity demand. Since electricity producers are not considered in this paper, components should be positive.

2. For each class k, components should have an average unit, *i.e.* $\sum_t d_k(t) = T$, to have comparable profiles. To write this constraint in mathematical terms, we define the column of length K, $u = (1, \dots, 1)^\intercal$, and the column of length T, $v = (T^{-1}, \dots, T^{-1})^\intercal$ in order to write the average unit constraint, with a Kronecker product $\otimes$, as $(I_K \otimes v^\intercal)\beta = u$.

The optimization problem then writes

$$\begin{aligned} \min_{\beta} \quad & \|x - (A \otimes I_T)\beta\|^2 \\ \text{s.t.} \quad & \beta \geq 0 \\ & (I_K \otimes v^\intercal)\beta = u \end{aligned} \tag{2}$$

An alternating direction method of multipliers [10] is used to recursively solve problem (2):

1. minimize the function with the equality constraint by employing the augmented Lagrangian method,

2. retain only positive components to satisfy the positivity constraint,

3. adjust a penalty variable balancing positivity and the minimization.

The algorithm is implemented with the R language [24]. Special care is taken on the first step, since the minimization requires inverting a large matrix of size $K(T+1)$. With common Kronecker product rules, matrix to be invert is reduced to size K divided the number of flops by approximately T^3.

2.3 Categorization of electricity customers

The aggregated demand profile $d^f(t)$ of a feeder f aggregates a large group of customers (a few thousands). The CIS provides general features on these customers, *i.e.* annual consumption, type of contract, and contracted power, which can be used to cluster them into K different categories. Once the features are selected, the

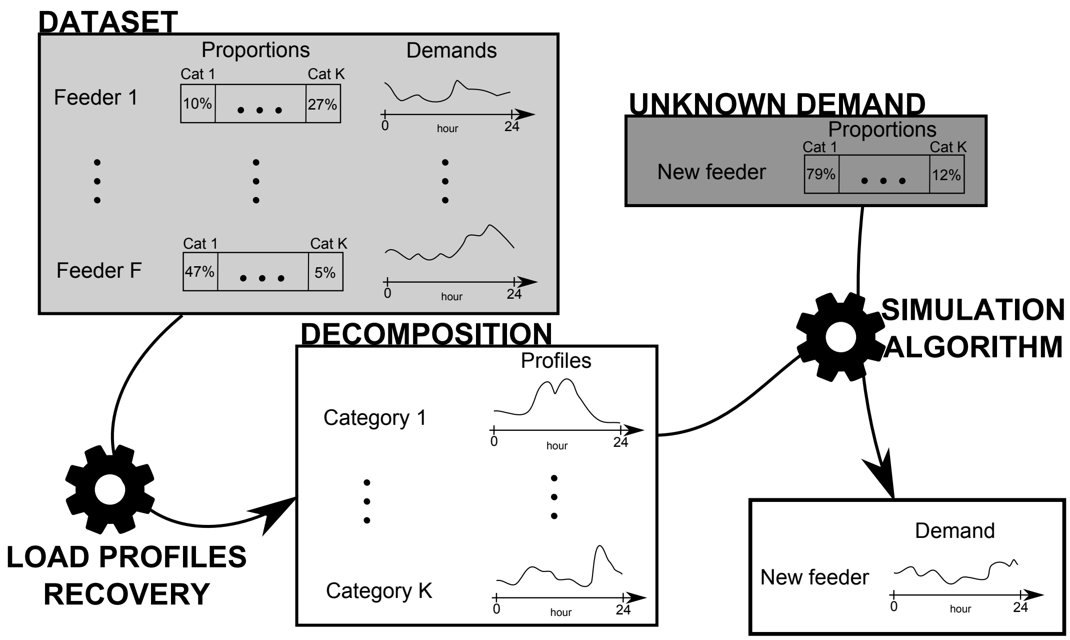

Figure 1: Diagram detailing the method. A dataset of F feeder measurements is used to find the K category profiles. Once the load profiles recovery is operated, a new feeder whose category distribution is known can be run through the simulation algorithm in order to obtain its expected demand.

total annual consumption c_k^f of a category $k \in \{1, \ldots, K\}$ in a feeder $f \in \{1, \ldots, F\}$ is computed from each annual individual consumption. The corresponding weight p_k^f is a normalized version of this consumption

$$p_k^f = \frac{c_k^f}{\sum_{k=1}^{K} c_k^f} \in [0, 1] \tag{3}$$

It is important that the size of the dataset F should be larger than the number of categories K. Empirically, it was observed that the condition $F > 5K$ is preferable in order to obtain a wide range in the set of category distributions, and thus a more precise result. Features should be general enough to keep a reasonably low K for three reasons: (i) to obtain a robust profile, (ii) to avoid an excessively long computing time, and (iii) to ensure that user privacy is not violated.

Figure 2 sets out four different categorizations, based on information from the CIS. The first categorization divides the total energy into two groups: residential and tertiary. The second splits the tertiary into 7 categories to make a total of 8 categories, *i.e.* residential, agriculture, commercial, public equipment, office and hospital, industry, restaurant and hotel, and medium-voltage (MV) customers (*e.g.* large buildings that have a specific contract with the operator). A 9-group division results from splitting the residential share into two groups: base tariff and special tariff[1]. Finally, an even more precise cat-

[1]Special tariff charges less during fixed off-peak peri-

egorization, *i.e.* 12 groups, is proposed. Commercial buildings are split into 2 categories reflecting low and high annual consumption. Similarly, MV customers are divided into 3 groups: low, medium and high.

On Figure 2, category heights for a category k represent the average demand shares for a given category $m_k = \frac{1}{F}\sum_{f=1}^{F} p_k^f$.

The share in category distribution is different for every feeder. For instance, there are more restaurants in a city center than in a rural area and so the two electricity shares are different. This share has to vary between feeders to efficiently compute the demand profiles. We computed the coefficients of variation

$$\frac{\sigma_k}{m_k} \tag{4}$$

where σ_k^2 is the empirical variance of $p_k^1, \ldots, p_k^F$. The coefficients are always higher than 40%, and thus the different categorizations are sufficiently spread from one feeder to another for our algorithm.

3 Case study

3.1 Data description

In this case study, we use electricity feeder demand measured every ten minutes in 3 geographical regions in France. Data come from the main French DSO, Enedis. The three regions encompass a large French city and the surrounding countryside. The three cities are Blois, Lyon and Rennes. Each region is divided into around 500 feeders, and each of these feeders provides electricity for about 1,000 customers. For each feeder, we know the demand measured for 4 years from 2010 to 2013. We discard some feeders because the measures are too scarce and their overall quality is not sufficient. This can result from database errors or from network reconfiguration or physical injuries on the grid [17]. Ultimately, between 200 and 400 feeders are selected for each region.

ods (*i.e.* during the night) but more during peak hours.

3.2 Temperature effect and normalization

Aggregated demand measurements cannot be directly compared since some feeders are connected to more customers than others, causing a large discrepancy in average consumption. In order to be used as inputs in the method, measurements therefore need to be pre-processed. The two steps of this pre-processing are: removal of the temperature effect, and normalization by weekly consumption.

Electricity demand is mostly influenced by outdoor air temperature, as residents turn on electric devices to adjust their indoor temperature (heating and air conditioning). In France, the air conditioning effect is low and not considered in this paper, but the heating effect is high during cold weather. French electric demand represents 40% of the European thermal sensitivity [14]. Indeed, since most French heating devices are electric, demand strongly increases when temperature decreases. However, this effect is well understood and can be removed and treated separately with a method used by the French TSO [20, pp 11—12]: one linear regression for each hour of week. Therefore, for each feeder f, we can determine a temperature threshold b^f and a trend $a^f > 0$ such as for each degree colder than threshold b^f, demand increases by a^f. A new demand series is defined from the

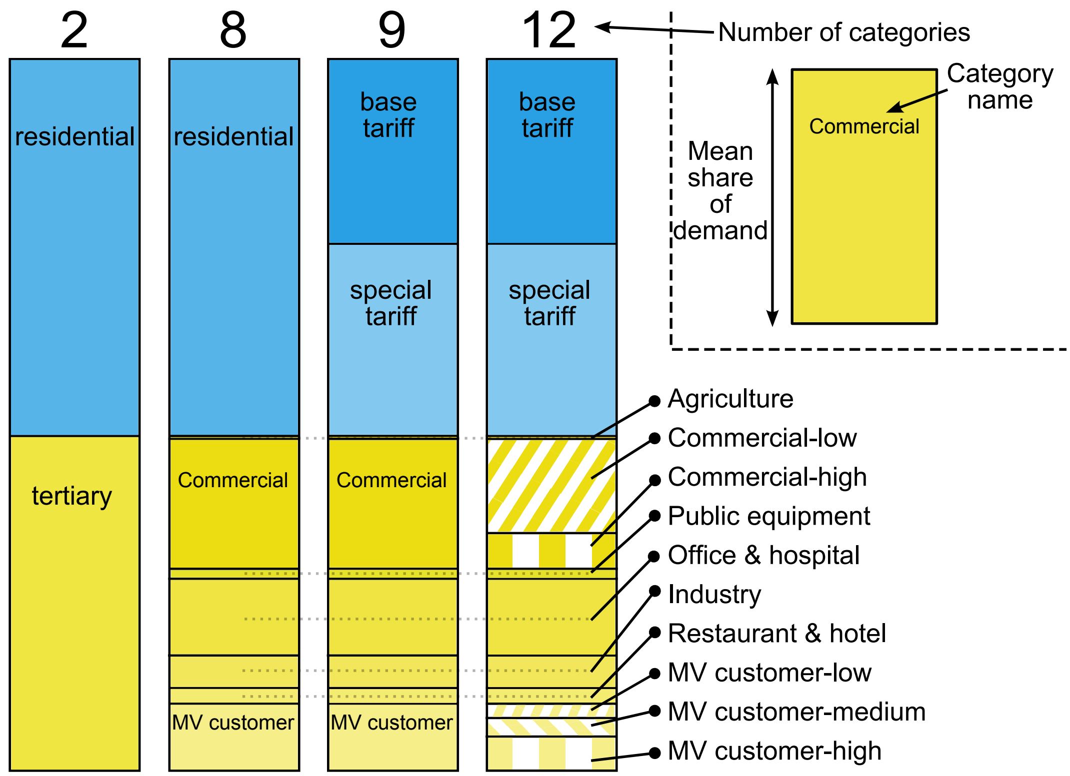

Figure 2: Example of different categorizations (in 2, 8, 9 or 12 groups) for the region near Lyon. There are $F = 320$ feeders in this dataset. The height of a division shows the mean share of the category in all feeders in the region.

initial $d_0^f(t)$

$$d_1^f(t) = \begin{cases} d_0^f(t) & \text{if } T^f(t) > b^f \\ d_0^f(t) - a^f\left(b^f - T^f(t)\right) & \text{otherwise.} \end{cases} \quad (5)$$

where $T^f(t)$ is the outside temperature of feeder f at instant t. In fact, trends a^f and threshold b^f are calculated for each hour of the day but the hour index is omitted for clearer notation. The new series is thus supposed to be independent from the temperature, and demand dynamics are supposed to be similar during cold and warm periods.

To obtain comparable measurements between feeders, demand is normalized. Each measurement within a given week is divided by the energy it consumed during that week. This total energy can be predicted using different models, such as that employed in [6], and is thereafter supposed to be known. After the normalization, data values fluctuate around a dimensionless value equal to 1.

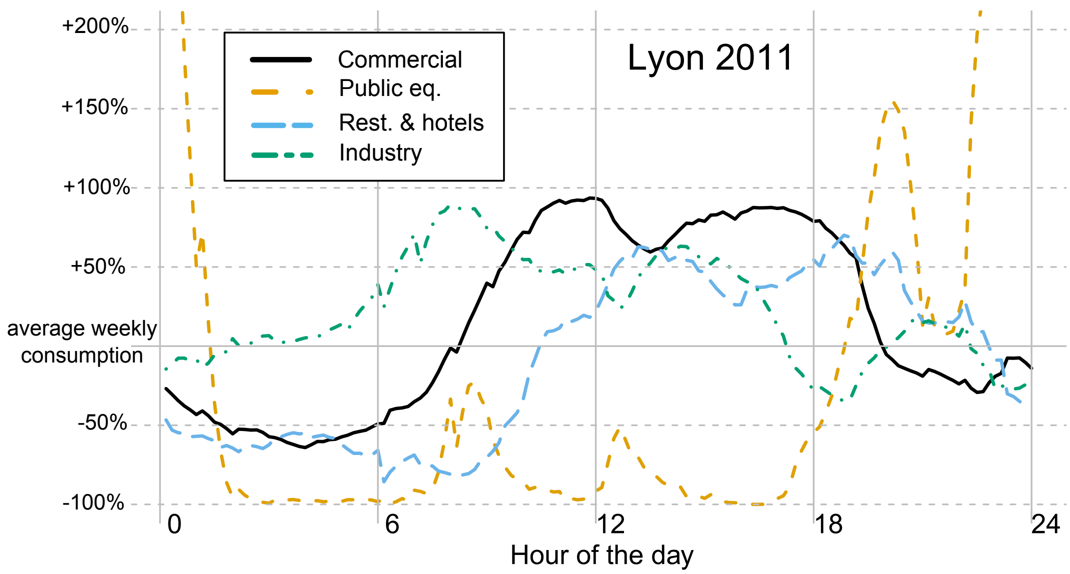

Figure 3: Weekday profiles of 4 different categories computed with the algorithm (9 overall categories) using aggregated consumption data relating to Lyon in 2011. Plots represent the variations around the average weekly consumption and not absolute consumptions.

3.3 Profiles

As previously described (see Figure 1), we disaggregated the electricity demand in order to recover a load profile $d_k(t)$ for each category $k \in \{1, \ldots, K\}$. The number of overall categories depends on the customer categorization: 2, 8, 9 and 12 categories were tried out (see Figure 2). A total of 12 datasets is formed (for each region: Blois, Lyon and Rennes; and for each year: from 2010 to 2013) and separately used as input into matrix X in problem (2).

Figure 3 presents the profiles obtained for $K = 9$ with only 4 categories shown: commercial, public equipment, restaurant and hotel, industry. Profiles are computed with the demand dataset of Lyon in 2011. Profiles are presented for a typical weekday (144 values, once every 10 minute). Since we have normalized the data, the variations around the average weekly consumption are displayed. Different effects are noteworthy, *e.g.* the electricity consumption of commercial buildings increases by around 75% during working hours, and decreases by 50% during the night. Conversely, the consumption of public equipment (mainly public lighting and lifts) greatly increases at night. These profiles are a pertinent way to understand electricity demand patterns. Profiles can be plotted for other datasets (another region or another year) in order to analyze specific characteristics.

4 Applications of the method

4.1 Estimation of the contribution of new customer sets to a feeder peak load

To plan the expansion of a new area, the DSO has to estimate the evolution of peak demand. The profiles obtained enable it to quantify and forecast the contribution of the new set of customers in the peak load demand. Indeed, for a feeder f at year y_0 with proportions $p^f_{1,y_0}, \ldots, p^f_{K,y_0}$ we can determine the residuals $\varepsilon^f_{y_0}(t)$ in Equation 1 and for new proportions $p^f_{1,y_1}, \ldots, p^f_{K,y_1}$ in a future year y_1 the forecast demand is obtained by

$$d^f_{y_1}(t) = \sum_{k=1}^{K} p^f_{k,y_1} d_k(t) + \varepsilon^f_{y_0}(t). \tag{6}$$

Figure 4 depicts the peak change obtained with this formula in the case of different evolutions for both offices and special-tariff residential consumers. In this case study, the considered feeder is from the Lyon region and has the following distribution of customers: 30% commercial, 15% offices, 30% basic residential and 20% special special-tariff residential. The initial peak occurs at 12:10 and is 650 kW. The profiles used are taken from the 9-category breakdown. We quantify the influence on the peak value (black lines with value added to the initial peak value, per 50 kW) by adding an office category load (Y axis) and a special-tariff residential load (X axis). We also depict the evolution of the peak hour (black dashed line). Adding offices contributes to increasing the 12:10 peak, whereas the residential load increases the 23:00 peak, which corresponds to the start of the special-tariff period.

This is an illustration of an application of the method that can for example help decision-makers to choose between two projects (offices or a new residential area) and quantify the impact on the existing feeder demand.

4.2 Evaluation, comparison of the method and category relevancy

4.2.1 Simulation evaluation

Thanks to the computed profiles, the aggregated demand of a feeder can be simulated. Each category profile is multiplied by the consumption share of the category. The category distribution is the only information required for the simulation; there is no need for historical demand recordings. We show a simulation example on Figure 5. Demand is simulated with only two categories: residential (green area) and tertiary (orange area). We sum the two profiles multiplied by their respective share (here 75% residential and 25% tertiary consumption). The measured consumption of a feeder with a 75/25 proportion is superimposed in black. The respective contribution of the two categories at each time step is clearly observable on the aggregated demand.

To assess the quality of the model, we use the Root-Mean-Square Error (RMSE) index. For each region and for each year, we compute profiles for 2, 8, 9 and 12 overall categories and use them to simulate new feeders. We then compare the simulation with actual demand with a leave-k-out approach ($k = 50$). This means that a subset of k feeders (that are not used in the training stage) is simulated. An RMSE for each of these feeder subsets is obtained and the average value is computed. This process is repeated 100 times to remove the volatility effect caused by the random subset of a 50-feeder selection. Computation takes roughly 16 hours for every region and every year on a 3.50 GHz machine.

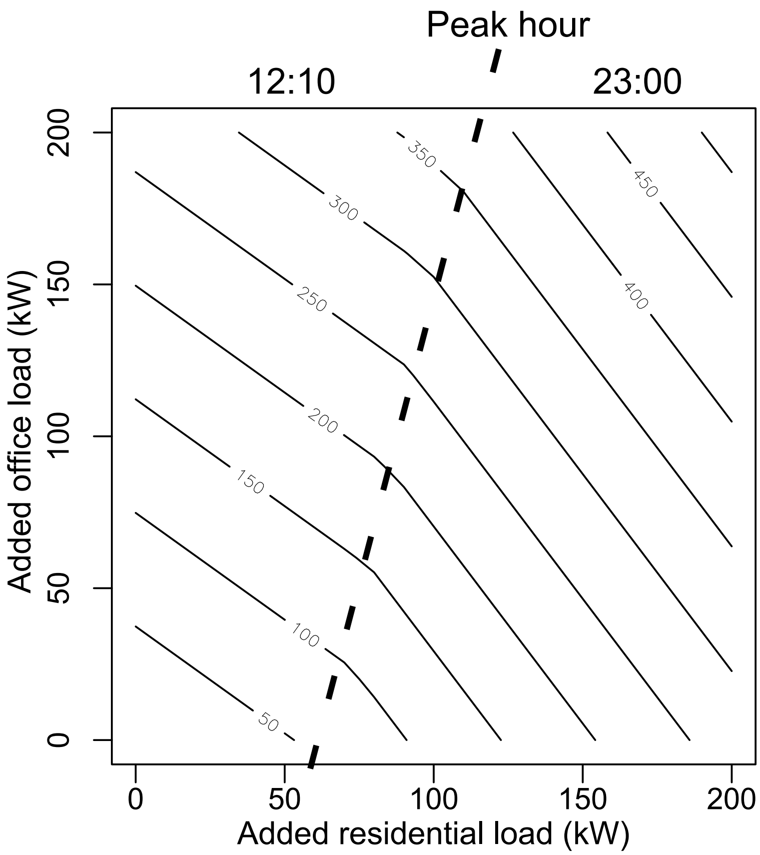

Figure 4: Peak change with a new load in a given feeder.

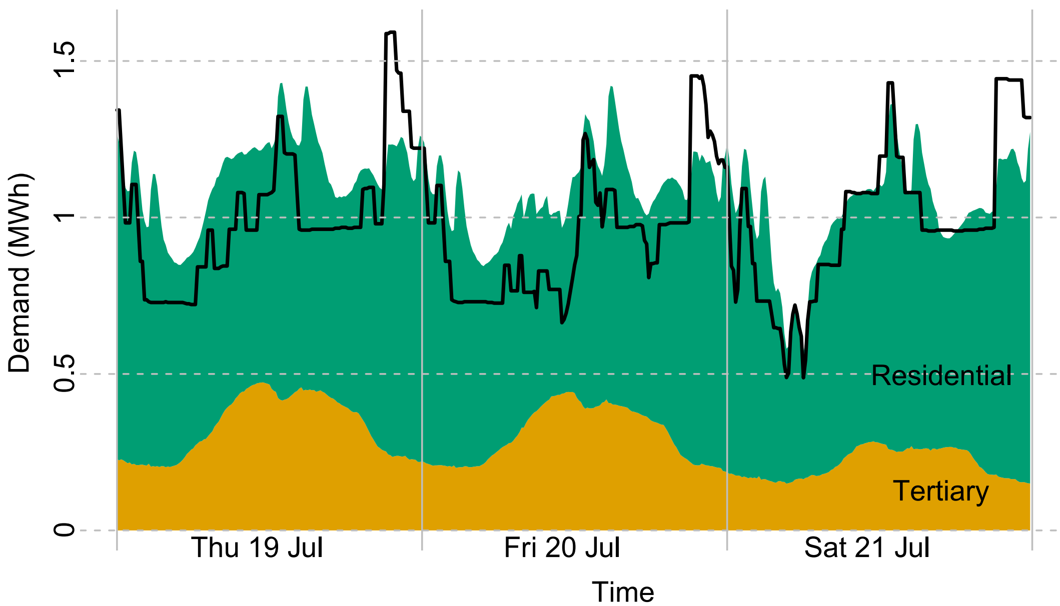

Figure 5: Simulation for one feeder. The profiles were obtained using demand data from Blois for 2012. The black line represents the actual consumption of the unknown feeder (not used in the training dataset). Our algorithm obtained two profiles: the orange part represents the tertiary demand and the green part the residential demand.

Table 1 reports the average RMSE and its deviation for the Blois, Lyon and Rennes during the 4 years for different numbers of categories. As a reminder, with consumption normalization, average consumption is dimensionless and equal to 1 (see Section 3.2). Hence, the RMSE reported is also dimensionless, and can be expressed as a percentage.

4.2.2 Category relevancy

Average RMSE is 22.59% for Blois, 18.16% for Lyon and 22.42% for Rennes with 9 categories. The errors are highly dependent on the regions, meaning that some regions are less predictable than others. Increasing the number of categories improves the overall model quality. The 8 category scheme almost always outperforms the 2 category one (by 2.5%). The 9 category scheme slightly improves results compared to the 8 version (by 1%), and so dividing customers into basic and special tariffs is meaningful. However, splitting small categories into even smaller categories is not recommended, as can be seen by the poor results of the 12 category scheme. A first reason may come from the use of CIS for classification: previous works have stated that using directly the CIS classification does not necessarily lead to the best profiles [12].

Another reason can come from the inter-group variability. As in any blind source separation task, a class is easy to recover and predict if it is distinctly separated from the other classes, and if it is observed in many different configurations. The statistics literature proposes many different separation metrics, but the simplest is a ratio between an inter-group variability measure and a total variability. In this context, since the variable of interest if a vector or even a curve it is not obvious to define the variability. We propose to define an inter-group variability measure with the weighted distance between d_k and d^f

$$V_{inter} = \sum_{f,k} p_k^f \|d_k - d^f\|_2^2,$$

and a total variance by $V_{tot} = \sum_f \|d^f\|_2^2$, where $\|x\|_2^2$ is the sum of the square of a vector x. The ratio between inter-groups and the total variance should be as high as possible. Measuring the diversity of configurations in which the final signal is observed can be related to the variance σ_k^2 and mean m_k of p_k^f among the feeders, the larger this variance and mean the more accurate the estimation will be. These separations and variability measures can be used to evaluate the value of adding categories. The inter-variance requires the computation of the d_k but σ_k^2 and m_k can be computed before any estimation.

4.2.3 Comparison to other models

Errors are higher than for middle-term forecasting methods, which can be around 7 to 10% of RMSE (see *e.g.* [8], [17]). However, our problem is different, and the relationship between the demand for a feeder f_1 for a given year y_0 and the demand for a feeder f_1 for the next year y_1 is much stronger than the relationship between the consumption of a feeder f_1 and the consumption of feeder f_2 for the same year y_0.

Framework of Andersen *et al.* is more similar to ours [5]. This presents "a model calculating local consumption by categories of customer with specific consumption profiles and different weights in local areas". Unlike us, their profiles are obtained by clustering representative smart-meter measurements, *i.e.* a bottom-up method. Their results from simulating local areas without using past measurements are expressed with R^2

value and are between 0.95 and 0.56 (their mean R^2 is 0.84). In their case study, the mean consumption of areas is 55.3 MW while in our case, for a given feeder it is between 0.5 and 7 MW. In order to compare our method with their method, we aggregated our areas to obtain similar average power levels and computed the R^2 between prediction and measurements. The results are shown in Table 2.

The performances of our method are a little higher than Andersen *et al.*'s method in the Lyon and Rennes case studies, and similar in the Blois study.

Area	Avg. demand 2010 (MW)	R^2
Blois	31.5	0.82
Lyon	46.2	0.88
Rennes	37.4	0.87

Table 2: Coefficient of determination R^2 for different areas showing the predictive performance of our method with a 9-category breakdown. The prediction of a group of 20 feeders is compared to the measured demand of the 20 feeders. We also report the average demands, which are comparable to the areas described by Andersen *et al.* [5] with similar R^2 values : on average they found R^2 of 0.84 for predicting different areas with an average demand of 55.3 MW.

5 Conclusions

Our paper has proposed a novel method to estimate elementary profiles. The main assumption of the method relies on feeder demands that aggregate various shares of elementary profiles associated with different customer categories. The profiles are optimally found by minimizing prediction errors in a new algorithm relying on the augmented Lagrangian method.

Unlike bottom-up methods that require individual load curves, our method only requires several feeder demand curves and a description of customers. One of the advantages of using aggregated measurements on a set of individual load curves is that they can be updated regularly and are fully representative. In the meantime, we have shown that our method performs similarly or better than a bottom-up method in the literature to predict a new local area.

The method has been applied in a case study comprising three zones in France, with around 300 available feeder measurements over 4 years per zone. The result is a load profile for each customer category. We have shown that each load profile gathers intrinsic features of the given category.

A first application using the resulting profiles was presented for planning the expansion of a new area at DSO level. The resulting profiles allow for different quantification and forecasting of the contribution made by the new set of customers to peak load demand. This was illustrated by a case study on a specific feeder where the evolution of peak demand in the case of adding two share categories was discussed. A second application of the profiles is to simulate the electricity demand of the new unmeasured areas. This can be used to test the relevancy of various types of categorization (2, 8, 9 or 12 groups were tested). By analyzing forecasting errors, we observe that using more categories does not necessarily lead to more efficient models, several causes are discussed.

Further research could investigate the creation of an automatic way to create categories, *e.g.* by maximizing entropy information, to create the best profiles and minimize prediction errors.

Socio-demographic statistics might be efficient to accurately describe categories. Information such as mean household area and building age are very meaningful in electricity demand forecasting, and are thus areas for further research.

6 Acknowledgments

The authors would like to thank Enedis for supplying data to make this work possible, and particularly Nicolas Kong from the Direction Technique, Politiques et Stratégie group, for his precious expertise on measurements, CIS data, and the Bagheera planning model.

References

[1] I. E. Agency, *World Energy Outlook*, IEA Publishing, Paris, 2015.

[2] ——, *Co2 emissions from fuel combustion highlights 2016*, IEA Publishing, Paris, 2016, ch. Key trend in CO2 emission from fuel combustion, p. 12.

[3] ——, *Key World Energy Statistics*, IEA Publishing, Paris, 2016.

[4] F. Andersen, H. Larsen, and T. Boomsma, *Long-term forecasting of hourly electricity load: Identification of consumption profiles and segmentation of customers*, Energy Conversion and Management, 68 (2013), pp. 244 – 252.

[5] F. Andersen, H. Larsen, and R. Gaardestrup, *Long term forecasting of hourly electricity consumption in local areas in Denmark*, Applied Energy, 110 (2013), pp. 147 – 162.

[6] F. Andersen, H. Larsen, N. Juul, and R. Gaardestrup, *Differentiated long term projections of the hourly electricity consumption in local areas. the case of Denmark West*, Applied Energy, 135 (2014), pp. 523 – 538.

[7] N. Bassamzadeh and R. Ghanem, *Multiscale stochastic prediction of electricity demand in smart grids using bayesian networks*, Applied Energy, 193 (2017), pp. 369 – 380.

[8] K. G. Boroojeni, M. H. Amini, S. Bahrami, S. Iyengar, A. I. Sarwat, and O. Karabasoglu, *A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon*, Electric Power Systems Research, 142 (2017), pp. 58 – 73.

[9] K. G. Boroojeni, S. Mokhtari, M. H. Amini, and S. S. Iyengar, *Optimal two-tier forecasting power generation model in smart grids*, CoRR, abs/1502.00530 (2015).

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends® in Machine Learning, 3 (2011), pp. 1–122.

[11] J.-F. Cardoso, *Blind signal separation: statistical principles*, Proceedings of the IEEE, 86 (1998), pp. 2009–2025.

[12] G. Chicco, R. Napoli, and F. Piglione, *Comparisons among clustering techniques for electricity customer classification*, IEEE Transactions on Power Systems, 21 (2006), pp. 933–940.

[13] E. Commission, *Benchmarking smart metering deployment in the EU-27 with a focus on electricity*, European Commission, Brussels, June 2014.

[14] R. de transport d'électricité, *Bilan prévisionnel de l'équilibre offre-demande d'électricité en france*, 2016, ch. Consommation d'électricité en France, p. 40.

[15] J. Dickert and P. Schegner, *Residential load models for network planning purposes*, in 2010 Modern Electric Power Systems, Sept 2010, pp. 1–6.

[16] N. Ding, *Load models for operation and planning of electricity distribution networks with metering data*, theses, Université de Grenoble, Nov. 2012.

[17] Y. Goude, R. Nedellec, and N. Kong, *Local short and middle term electricity load forecasting with semi-parametric additive models*, IEEE transactions on smart grid, 5 (2014), pp. 440–446.

[18] M. Jin, W. Feng, P. Liu, C. Marnay, and C. Spanos, *Mod-dr: Microgrid optimal dispatch with demand response*, Applied Energy, 187 (2017), pp. 758 – 776.

[19] K. le Zhou, S. lin Yang, and C. Shen, *A review of electric load classification in smart grid environment*, Renewable and Sustainable Energy Reviews, 24 (2013), pp. 103 – 110.

[20] V. Lefieux, *Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d'électricité.*, PhD thesis, Université Rennes 2, 2007.

[21] E. McKenna, I. Richardson, and M. Thomson, *Smart meter data: Balancing consumer privacy concerns with legitimate applications*, Energy Policy, 41 (2012), pp. 807–814.

[22] F. McLoughlin, A. Duffy, and M. Conlon, *A clustering approach to domestic electricity load profile characterisation using smart metering data*, Applied Energy, 141 (2015), pp. 190 – 199.

[23] A. Mutanen, M. Ruska, S. Repo, and P. Jarventausta, *Customer classification and load profiling method for distribution systems*, IEEE Transactions on Power Delivery, 26 (2011), pp. 1755–1763.

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015.

[25] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, and M. E. Webber, *Clustering analysis of residential electricity demand profiles*, Applied Energy, 135 (2014), pp. 461 – 471.

[26] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, *Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data*, Applied Energy, 87 (2010), pp. 3538 – 3545.

[27] A. Seppälä, *Load research and load estimation in electricity distribution*, theses, Technical research center of Finland, VTT Publications, Jan. 1996.

[28] E. P. R. Service, *Smart electricity grids and meters in the EU Member States*, European Parliament, Brussels, September 2015.

[29] Z. Shao, F. Chao, S.-L. Yang, and K.-L. Zhou, *A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting*, Renewable and Sustainable Energy Reviews, (2016), pp. –.

[30] Z. Shao, F. Gao, Q. Zhang, and S.-L. Yang, *Multivariate statistical and similarity measure based semiparametric modeling of the probability distribution: A novel approach to the case study of mid-long term electricity consumption forecasting in China*, Applied Energy, 156 (2015), pp. 502 – 518.

[31] J. L. Viegas, S. M. Vieira, R. Melício, V. Mendes, and J. M. Sousa, *Classification of new electricity customers based on surveys and smart metering data*, Energy, 107 (2016), pp. 804 – 817.

Region	Year	2 categories	8 categories	9 categories	12 categories
Blois	2010	24.36 (2.32)	**23.90 (2.75)**	24.04 (2.99)	26.78 (2.91)
	2011	23.87 (1.62)	**22.79 (1.42)**	22.91 (1.16)	24.78 (2.09)
	2012	22.84 (1.26)	22.54 (1.17)	**22.09 (1.24)**	24.09 (2.17)
	2013	22.34 (2.06)	22.31 (1.98)	**21.32 (1.96)**	23.34 (2.04)
	Average	23.35 (1.86)	22.89 (1.93)	**22.59 (1.98)**	24.75 (2.33)
Lyon	2010	19.05 (2.39)	19.42 (2.71)	**18.29 (2.24)**	19.23 (1.94)
	2011	19.28 (1.24)	**18.06 (1.55)**	18.56 (1.20)	18.46 (1.42)
	2012	19.07 (1.35)	**18.21 (1.72)**	18.30 (1.34)	19.00 (1.86)
	2013	18.06 (1.03)	17.92 (2.03)	**17.49 (1.12)**	18.68 (1.91)
	Average	18.87 (1.59)	18.40 (2.05)	**18.16 (1.58)**	18.84 (1.79)
Rennes	2010	22.57 (0.96)	21.67 (1.23)	**21.59 (1.04)**	22.70 (1.57)
	2011	22.62 (1.22)	**21.54 (1.48)**	21.57 (1.06)	22.10 (1.08)
	2012	22.75 (1.11)	22.96 (0.99)	**22.39 (0.98)**	22.61 (0.84)
	2013	24.94 (1.03)	**23.99 (1.37)**	24.14 (1.26)	24.08 (1.37)
	Average	23.22 (1.08)	22.54 (1.28)	**22.42 (1.09)**	22.87 (1.25)

Table 1: RMSE (in %) of the models for the 3 different zones over the 4 years with a different number of categories. The simulation is run 100 times. We reported the average RMSE and its standard deviation between parentheses. The best results over the 4 numbers of categories are written in bold.