Electricity Load Forecasting for Residential Customers: Exploiting Aggregation and Correlation between Households

Samuel Humeau, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer
School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
Email: {samuel.humeau, tri-kurniawan.wijaya, matteo.vasirani, karl.aberer}@epfl.ch

Abstract—The recent development of smart meters has allowed the analysis of household electricity consumption in real time. Predicting electricity consumption at such very low scales should help to increase the efficiency of distribution networks and energy pricing. However, this is by no means a trivial task since household-level consumption is much more irregular than at the transmission or distribution levels. In this work, we address the problem of improving consumption forecasting by using the statistical relations between consumption series. This is done both at the household and district scales (hundreds of houses), using various machine learning techniques, such as support vector machine for regression (SVR) and multilayer perceptron (MLP). First, we determine which algorithm is best adapted to each scale, then, we try to find leaders among the time series, to help short-term forecasting. We also improve the forecasting for district consumption by clustering houses according to their consumption profiles.

I. Introduction

Energy management has become an important matter all around the world Most of the electricity distribution networks currently in place were not designed to handle today's huge demands, let alone tomorrow's. Congestion and atypical power flows threaten to overwhelm the system, while there is ever more demand for greater reliability, and better security and protection [1]. Renewable energy sources add a further challenge (as governments plan to reduce dependency on fossil fuels), first because of the intermittent nature of their production, and second because they are widely dispersed (e.g., rooftop solar panels) compared to traditional, centralized electricity production. This introduces a new paradigm to demand side management, i.e., attempting to match the demand side with the available supply [2], [3].

In this context, anticipation and load forecasting (or prediction) become increasingly crucial. Long-term load forecasting (1-10 years ahead) is important for planning both transmission and distribution networks [4]. On the other hand, short-term load forecasting (hours to weeks ahead) is important for the online scheduling and security functions of an energy management system [5]. This paper focuses on short-term load forecasting, one hour and 24 hours ahead, which is especially important for planning generation and demand side management.

Furthermore, the necessary geographical scale of accurate forecasting is becoming ever finer: from a country in the 1990s, down to a district in recent years. Techniques for forecasting have been inspired by research on machine learning, and have passed from linear regression and autoregressive moving average models [6], to neural networks [7] and boosting approaches [8], and finally to the support vector machine for regression (SVR) that is a state of the art forecasting method [9], [10]. These techniques have been used successfully for consumption forecasting at country scale.

The rest of the paper is organized as follows. After having presented related work in Section II and our dataset in Section III, we apply some of these techniques at the household level in Section IV. More broadly, we look for the best method at each scale. In Section V, we try to use relationships between consumption time series to forecast one and 24 hours ahead. In Section VI we group similar households into clusters in order to improve the forecast for overall consumption. In Section VII we present the optimization of our predictors. Finally, we conclude in Section VIII.

II. Related work

The SVR for electricity load forecasting was introduced by Chen et al., who used it to win the EUNITE competition in 2001 [9]. It has remained a popular tool for load or electricity price forecasting; Oldewurtel et al. recently used it to regulate a building's consumption in a grid-friendly way [11]. Predicting the consumption of residential houses and small companies is not a recent idea. Standard consumer baselines have been established for a long time and PJM has elaborated and compared them [12].

Our work to find leaders in time series in order to improve load forecasting was greatly inspired by the work of Wu et al. who established a method to use these leaders for predicting exchange rates [13]. A time series is considered to be a leader if its rise or fall has an impact on the behavior of several other time series. We have applied Wu et al.'s techniques to consumption series to help short-term forecasting. In Section VI, we split the set of households into several clusters in order to adopt a bottom-up approach. This type of hierarchical view is often used to predict consumption on a large scale. For example, the goal of the 2012 Kaggle Global Energy Forecasting Competition was to forecast the consumption over 20 small areas in order to obtain good overall accuracy [8].

III. Preliminaries

A. Dataset

The CER Irish Smart Metering Trial took place during 2009 and 2010, with over 5,000 Irish homes and businesses participating [14]. The data set consists of a measure of each house's consumption (in kWh) every half an hour. This study only considered 782 residential houses whose owners were

not aware of the applied pricing policy and kept consuming as they usually had. We aggregated the measurements into hourly timeslots. For all the results presented, we have considered the first 12 months as the training set (8,460 hours), and the final 6 months as the test set (4,317 hours).

B. Evaluation

In the literature, the two well-known methods for evaluating forecasted values are the mean of average percentage error (MAPE) and the root mean square error (RMSE). In this paper, we sometimes make forecasting at the household level. For many households, there are periods when the consumption is 0, or extremely low. In these cases, the MAPE is infinite, making its evaluation inconvenient. RMSE does not have this problem. The RMSE for a time series $S = (s_1, ..., s_N)$ of its estimation $\tilde{S} = (\tilde{s}_1, ..., \tilde{s}_N)$ is defined by:

$$RMSE(S, \tilde{S}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (s_i - \tilde{s}_i)^2}$$

We observe that the RMSE is strongly influenced by the order of magnitude of the data. Our data set contains numerous households whose hourly consumption ranges between 0.05 kWh and 3.83 kWh. In order to reduce their influence, we use the following normalization for the RMSE:

$$NRMSE(S, \tilde{S}) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (s_i - \tilde{s}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} s_i^2}}$$

That is, we normalize RMSE by the L_2 norm of the series. NRMSE refers to the normalized RMSE.

C. Implementation details

All the programmes developed for this study were written using JAVA and Matlab; the code is publicly available.[1] We also used the Weka framework developed by the University of Waikato [15] to compute the predictions.

Our dataset offers measures for 782 households over a time span of 18 months. Since MLP and SVR computations are extremely time consuming, the evaluation of individual household consumption forecasting was not carried out for each house, but for each new experiment a panel of 25 households were selected randomly. Thus, we provide standard deviation on the mean in our results. The correct term with which to refer to an algorithm that makes predictions, would be a *predictor*, however, since the techniques we use are extensions of classification methods, the term *classifier* will also be used.

IV. FORECASTING WITHOUT CORRELATING LOADS BETWEEN HOUSEHOLDS

In this section, we define the features used to make predictions. We compute predictions for a single household and for the whole district, both one hour and 24 hours ahead. This is all done without exploiting relationships between households. Data at the level of a single household is not structured enough to efficiently use algorithms such as SVR. We work towards finding the appropriate scale at which SVR starts to bring improvements.
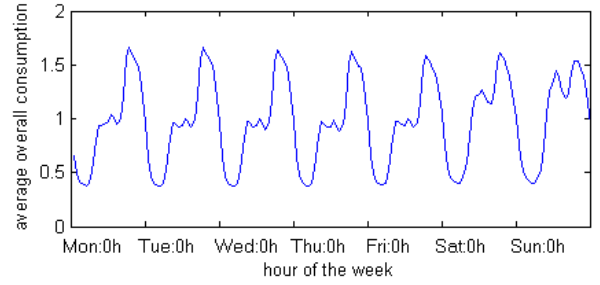

Fig. 1. Average energy consumption (in kWh) over time and households by the hour of the week; starts Monday 00:00. Note, we can clearly distinguish each day, and patterns for the weekend differ significantly.
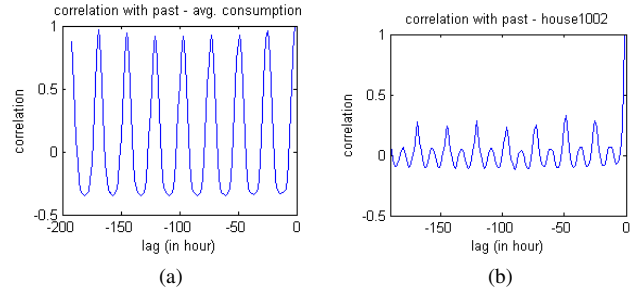

(a) (b)

Fig. 2. (a) Auto-correlation of aggregated consumption over 782 households. (b) Auto-correlation of the consumption of individual households (id 1002).

A. Features

1) Calendar features and historical load: Figure 1 shows the average consumption of one residential household over one week (averaged over time and houses). We can clearly distinguish each day (5 similar days, then Saturday and Sunday that follow a different pattern). Moreover, each hour of the day seems to follow a different pattern. This is very intuitive: consumption is very low at night and reaches a maximum at around 8 pm. This leads us to add the hour of the day and the day of the week as nominal features.

It is also very intuitive that consumption is related to historic loads, in particular with the consumption 24 hours before. However, due to the high variability in the consumption of any one household, the accuracy of this intuitive rule is low at this scale. Figure 2 illustrates this fact, by showing the auto-correlation of consumption aggregated over all the houses. It clearly shows that the consumption at any one time is highly related to consumption at the same hour both on the day before and seven days before. It also shows that this auto-correlation is lower for individual households. Consumption is highly correlated to consumption up to three hours previously, which means it is important to consider the last three hours in the case of one hour ahead prediction.

2) Temperatures: Figure 3 shows that there is a relationship between the temperature and consumption at 7 pm. The lower the temperature, the higher consumption. This relationship is mostly due to the fact that the country considered – Ireland – is relatively cold. So the increase in consumption linked to (low) temperature can be attributed to the use of heaters. We used the Wunderground database to obtain the temperatures in Dublin for the period of the of the data set. For simplicity, real temperatures have been used, and not those that could

---

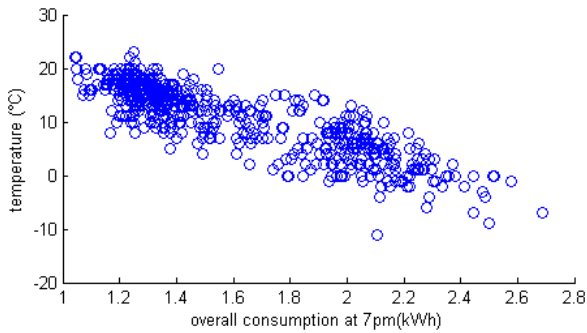[1]https://github.com/samuelhumeau/residential_load_forecasting.git

Fig. 3. Relationship between the average consumption at 7pm, and the temperature (averaged over all houses)

have been predicted.

3) Derivatives: The features above, added to our model, are absolute: they do not indicate a variation. In order to give the machine learning algorithm an indication of the evolution of consumption, we add consumption's first derivative $(s_i - s_{i-1})$, and second derivatives $(s_i - 2s_{i-1} + s_{i-2})$ as features.

B. Algorithms used

A lot of machine learning based algorithms have been used in related work for load forecasting. Given that recent literature has described SVR as one of the most effective at predicting future consumption [9], [10], we have used it too. It uses a radial basis function (RBF) kernel and its optimization is described in Section VII. In order to compare our results, we also consider 2 other predictors:

1) a simple linear regression on all the features considered, and
2) an MLP, composed of one hidden layer with sigmoid activation function and one output. The setting of this classifier is also described in Section VII.

Moreover, in the case of the prediction for a single household, we compared our results with two predictors that simply calculate consumption one hour and 24 hours ago.

C. Forecasting for a district

TABLE I. FEATURES USED FOR FORECASTING IN SECTION IV-A

One hour ahead	24 hours ahead
$cons.t - 1h$	$cons.t - 1 \times 24h$
$cons.t - 2h$	$cons.t - 2 \times 24h$
$cons.t - 3h$	$cons.t - 3 \times 24h$
$cons.t - 1 \times 24h$	$cons.t - 4 \times 24h$
$cons.t - 2 \times 24h$	$cons.t - 5 \times 24h$
$cons.t - 3 \times 24h$	$cons.t - 6 \times 24h$
$cons.t - 7 \times 24h$	$cons.t - 7 \times 24h$
hour of the day	hour of the day
day of the week	day of the week
derivatives	-

Once our features are defined (they are summarized in Table I), we apply the 3 predictors to the aggregated consumption of 782 residential households in our dataset. As for all predictors in this paper, we use the first 12 months as a training set and the last 6 months as a test set to evaluate solutions. The results are given in Table II in terms of MAPE, RMSE, and NRMSE. A few things are noticeable. First, given that the district's average consumption per hour is 782 kWh, the error in the prediction for each method is quite low. As expected, predictions one hour ahead are far more precise than those 24
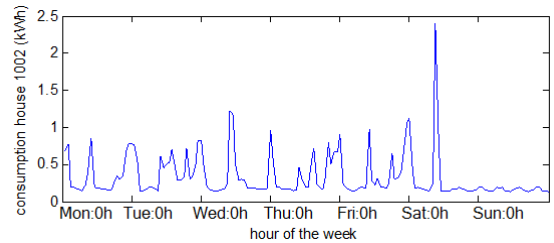
(a)
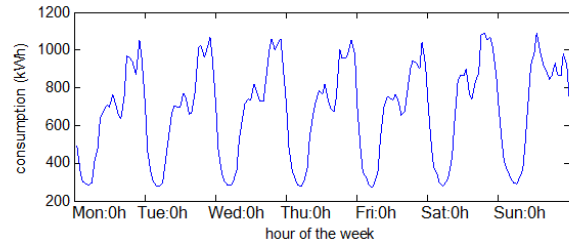
(b)

Fig. 4. (a) Household consumption (house id 1002) for the week from 2009-09-07 to 2009-09-13, (b) Aggregated consumption for the whole district during the same period

hours ahead. Finally, at the district scale, SVR outperforms the two other methods by far.

TABLE II. PERFORMANCE OF THREE PREDICTORS TRAINED ON AGGREGATED DATA (ONE HOUR AND 24 HOURS AHEAD FORECASTING). .

	One hour ahead			24 hours ahead		
	Lin. Reg.	MLP	SVR	Lin. Reg.	MLP	SVR
MAPE	5.2%	6.0%	3.4%	6.1%	6.9%	4.3%
RMSE	58.3	51.1	37.4	77.0	68.5	54.4
NRMSE	0.069	0.061	0.045	0.092	0.081	0.064

D. Forecasting at the single household level

In this section, we study the possibility of predicting the consumption of a single household. Data for such a case is more chaotic. Figure 4 shows the difference between the consumption of a single household over a week and the district's aggregate consumption over the same period. While we can recognize patterns in the latter, it is hard to find them in the former. The features and algorithms used for prediction are the same as in Table I, and results are shown in Table III. One noticeable result is the fact that linear regression outperforms SVR at this scale, for predictions both one hour and 24 hours ahead. This behavior is not really a surprise since SVR relies on the idea that data possesses an internal structure, which is not the case here.

TABLE III. NRMSE OF DIFFERENT PREDICTORS ON THE TEST SET FOR (A) ONE HOUR, AND (B) 24 HOURS AHEAD FORECASTING USING FEATURES IN TABLE I.

		h-24	h-1	Lin. Reg.	SVR	MLP
A.	Average NRMSE	0.80	0.71	0.56	0.57	0.59
	σ on the average	0.030	0.038	0.023	0.025	0.027

		h-24	Lin.Reg	SVR	MLP
B.	Average NRMSE	0.80	0.61	0.64	0.70
	σ on the average	0.030	0.019	0.023	0.038

We also perform forecasting using temperature as an additional feature. However, this does not improve the result shown in Table III, i.e., temperature does not add any predictive power to the features in Table I.
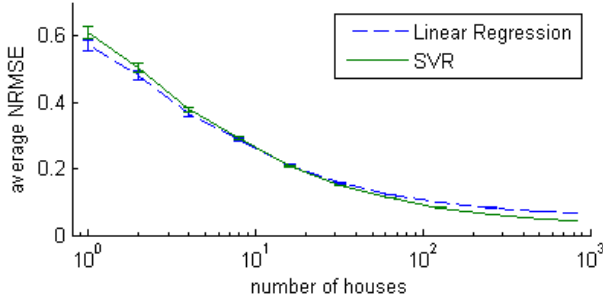
Fig. 5. Performance of linear regression and SVR for one hour ahead forecasting expressed using NRMSE (described in Section III-B) and its standard deviation. We run the prediction 36 times. SVR outperforms linear regression if we consider more than 16 residential houses.

E. Influence of scale

Both the previous section and Section IV-C highlighted two paradoxical results. To predict the consumption of a large number of houses, SVR is better, but for the prediction for a single household, linear regression is the preferred method. The next significant question is up to what scales should we use the different models? To determine this, we consider groups of $1, 2, 2^2, ..., 2^9$ and 782 houses (the whole set). Let K be the number of houses considered. We randomly take K houses from the dataset, and aggregate their consumption. Then, using the features in Section IV-C, we train a linear regression and a SVR regression (using parameters described in Section VII). We repeated this it 36 times, in order to obtain a representative average. The result is shown Figure 5, in terms of NRMSE. This leads us to conclude that the transition happens when $K \in [8,32]$ for the CER Irish dataset. This means that for a group of more than 32 households, using SVR rather than linear regression increases the quality of predictions.

V. FORECASTING BY CORRELATING LOADS BETWEEN HOUSEHOLDS

The CER Irish dataset gives us the possibility of establishing relationships between households, and exploiting them. In this section, we aim to find the households that lead others' consumption, in order to improve one hour ahead prediction. Thus, we focus on the improvement of one hour ahead forecasting for a single household.

A. Finding leaders in time series

The starting point for this section is the hypothesis that some households play the role of leaders in terms of consumption. This approach was exploited by Wu et al. [13] to forecast financial trends. If we are able to find some households that seem to lead consumption, then we could exploit them for the predictions.

1) Finding leaders using cross correlation: The task is to determine which households lead other household's consumption. We use the cross correlation by interpreting households' consumptions as signals. The cross-correlation between two signals $S^{(1)} = (s_1^{(1)}, ..., s_N^{(1)})$ and $S^{(2)} = (s_1^{(2)}, ..., s_N^{(2)})$ for a certain lag τ is given by:

$$CC(S^{(1)}, S^{(2)}, \tau) = \frac{\mathbf{E}((s_i^{(1)} - \mathbf{E}(s^{(1)})) \times (s_{i-\tau}^{(2)} - \mathbf{E}(s^{(2)})))}{\sigma(s^{(1)})\sigma(s^{(2)})}$$

Where $\mathbf{E}$ is the average and σ is the standard deviation. Since the goal is a short-term prediction, we can keep the

delay τ small, but not 0. In our case, for each pair of households (h_1, h_2), we compute the cross correlation between their consumptions for τ comprised between -4 and 4, and different from 0. We then determine the lag $\tau_{h_1 h_2}$ with the highest correlation between the two signals. A positive $\tau_{h_1 h_2}$ means that household h_2 leads the consumption of household h_1. We store the result obtained in two symmetric matrices, namely the maximum cross correlation (MCC) matrix and the leader of cross correlation (LCC) matrix.

$MCC(h_1, h_2)$ shows the maximum cross correlation between household h_1 and h_2 while $LCC(h_1, h_2)$ shows $\tau_{h_1 h_2}$. On the diagonal, we observe each household's correlation with itself (lagged). In most cases, a household's consumption is more closely correlated with its own past consumption than with that of any other household. We can now extract the top-k leaders for each household.

Fig. 6. Illustration of the MCC and LCC matrices. House id 1027 seems to "lead" the house id 6817. This because the correlation between the consumption of the house id 1027 at time $t - 1$ and the consumption of house id 6817 is 0.36, i.e., close to its self correlation with lag 1 (0.45).

2) Leaders determined using mutual information: Determining leaders using cross correlations means that we are looking for households that have very similar consumption, but with a delay. However, with the objective of making predictions using sophisticated machine learning techniques, we are interested in any type of relationship between two households' consumption series. In this case, mutual information is more appropriate. This concept comes from Information Theory and expresses the quantity of information that links two random variables. Let X and Y be two discrete random variables. The mutual information between X and Y is expressed as:

$$I(X, Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$$

Where $H(X)$ is the entropy of variable X, and $H(X, Y)$ the joint entropy between X and Y. We then discretize each consumption series (using 8 bits uniform discretization) and consider them as random variables evaluated each hour. For other alternatives for discretizing smart meter data, we refer the interested reader to the work by Wijaya et al [16]. Thus, in the same way as with cross correlation, by looking for the

largest amount of mutual information in the consumption series we can extract the top-k leaders for each household.

B. Taking advantage of leaders for one hour ahead forecasting

In order to evaluate whether the introduction of leaders can really improve the very short term forecasting (one hour ahead); we applied our predictors on two different sets of features:

- The first set of features, $Set_{without}$ is the same as that used in Section IV-D.
- The second, $Set_{with} = Set_{without} + Set_{leaders}$ where $Set_{leaders}$ is composed of the consumption of the 5 best leaders with the corresponding lag.

The experiment is repeated for both kinds of leaders and the results are shown in Table IV. They are directly comparable with the results from Section IV-D. Unfortunately, the information added by the leaders' consumption does not bring more accuracy to the forecast.[2] This result is true for both types of leaders, i.e., using cross-correlation and mutual entropy. This leads us to conclude that, providing the historic loads and calendar features as in Section IV-D, leaders' information does not offer any additional predictive power.

TABLE IV.   NRMSE OF 4 CLASSIFIERS, ON THE SET OF FEATURES, INCLUDING THE LEADERS DETERMINED USING A.CROSS CORRELATION, B. MUTUAL INFORMATION.

	Cross Correlation	Linear Regression	SVR	MLP
A.	Average NRMSE	0.55	0.57	0.58
	σ on the average	0.023	0.024	0.024

	Mutual Information	Linear Regression	SVR	MLP
B.	Average NRMSE	0.55	0.57	0.58
	σ on the average	0.023	0.024	0.027

VI.   SUBSCALES

Since we have access to the consumption of every residential household in our dataset, there are two ways of forecasting a whole district's consumption. The first is described in Section IV-C and consists of using the aggregated consumption as features (large-scale prediction). Another method would be to calculate a prediction for each household separately and aggregate those predictions (small-scale prediction). Our experiments shows that the former method provides better results, with an RMSE of 37.4 kWh for the district (one hour forward prediction) against 60.1 kWh for the latter method. This section studies the possibility of a medium scale prediction that could outperform both. In other terms, can we group households into clusters, predict the consumption for each cluster, aggregate these predictions, and obtain a better load forecast for the district.

A. Clustering

The consequent idea is to regroup houses that have similar consumption into clusters. To carry out the clustering, we consider the average consumption of each household in each hour. Each household is thus defined by a 24-dimensional vector, which is used as an interpretation of that household's consumption profile. In order to find clusters among our 782-household dataset, we apply a k-means algorithm. We repeat the operations for 1, 2, 4, 8, 16, 32, 48 and 64 clusters. We run k-means 1,000 times using different random seeds and choose

---

[2]There are small improvements, however they are not statistically significant.
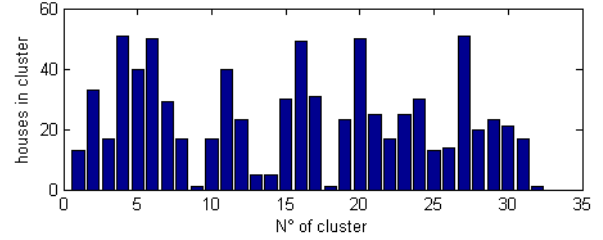

Fig. 7.   Number of households in each cluster when splitting the houses into 32 clusters using k-means.
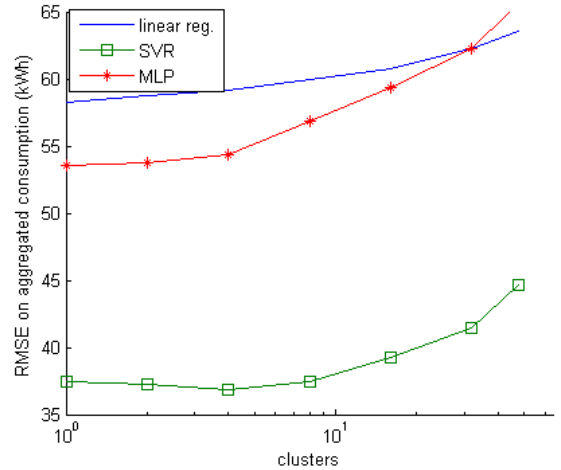

Fig. 8.   Evolution of RMSE on aggregated consumption, for different numbers of clusters, with different classifiers

the configuration with the lowest sum of squared errors within clusters. Figure 7 illustrates the distribution of households among (32) clusters.

B. Results

We forecast one hour and 24 hours ahead for each cluster, and aggregate the consumption. Figure 8 shows the evolution of the error on the overall consumption for different numbers of clusters, and different classifiers. Notice that results for MLP are the average over 25 iterations. Since linear regression and SVR both give a global minimum of error, there was no need to use redundancies for those algorithms.

It appears that for MLP and linear regression, error increases with the number of clusters. However, for SVR, we found a better result for 4 clusters, i.e., the global error decreases from 37.4 kWh to 36.8 kWh. A similar result is obtained for the 24 hours ahead prediction, whose error decreases from 54.4 kWh to 53.2 kWh. Results are shown in Figure 8 (for one hour hour ahead prediction). This is only a slight improvement of course, but it shows that in some cases, the use of a subscale can lead to better results. It is a start, and future work could consider the best classifier for each cluster, or use different criteria or features to create clusters.

VII.   CLASSIFIER OPTIMIZATION

Apart from linear regression, this work mainly used two classifiers: MLP and SVR.

A. Configuration of the MLP

The implemented MLP features only one sigmoidal hidden layer. Therefore, the output is $Y = W_2 \times \sigma(W_1 \cdot X + B)$

where X is the input, Y is the output, W_1, W_2, and B are matrices optimized by the MLP and σ is the sigmoid operator. The input is also normalized (removing the mean and dividing by the standard deviation). To avoid overfitting, a validation set is considered; its size is 30% (chosen randomly) of the training set. The gradient descent (learning rate=0.3) is stopped whenever the error on the validation set (calculated each epochs) has increased 10 times in a row.

B. Configuration of the SVR

SVR is a regression method based on support vector machine (SVM) that developed in 1996 by Vapnik. It is fully described in Smola and Schlkopf's tutorial [17]. During this work, we used the implementation of SVR by the LIBSVM library developed by Chang and Lin [18].

In SVR, there are parameters to decide. These include the type of kernel, and the cost of the SVM error, usually noted as C. Like Chen et al. [9], our experiments used an RBF kernel: $k(x,y) = e^{-\gamma\|x-y\|^2}$, which introduces a parameter γ. The following section deals with the optimization of the SVR classifier used in terms of C and γ.

1) The aggregated consumption case: In order to find the suitable C and γ, we apply the following methodology. We split the training set into two parts: a sub-training set and a validation set. We use the features described in Section IV-C. We tested $C = 1, 10, 10^2, 10^3, 10^4, 10^5$, and $\gamma = 0, 0.01, 0.1, 1$. In this case, changing the settings of C and γ gives significant differences in NRMSE. We found that the settings $C = 1000$ and $\gamma = 1$ outperformed the others, and those values are used in Section IV-C.

2) The individual household prediction case: We also used the training and validation sets in this case. We use the 25 households mentioned in Section III-C. The performance is measured in terms of average NRMSE over the households. We found that, changing the settings of C and γ does not give significant NRMSE differences. However, they strongly affect the computation time, which explodes when $C > 1000$ and $\gamma > 0.1$. Thus, for individual household prediction, we use $C = 100$ and $\gamma = 0.01$.

VIII. Conclusion

In the literature, impressive results for electricity load forecasting are obtained using SVR. We verified that when applied to a small district (782 houses), SVR does indeed seem to be the best method (of those tested) for load forecasting. However, with regards to making forecasting at the level of individual households, we showed that the method that produces the lowest error is the simple linear regression; it outperforms both MLP and SVR. We also established that there is an aggregation size (in our case, an aggregation of 32 households) at which SVR starts to outperform linear regression.

We also studied the idea of taking into account leadership relationships between households in order to improve short-term load forecasting. However, our experiments could not provide any relevant results. The accuracy of the forecasts, both with and without leadership information, was the same.

With regard to the problem of predicting the district's overall consumption, deriving the aggregate prediction from each individual household independently led to a poor forecast of overall consumption. However, SVR predictions could be improved by grouping houses with similar consumption into clusters.

Further work could consist in selecting the best classifier for each cluster, in order to improve the overall quality of the prediction. We are also working on using this method with measurements from sources other than residential households alone. An interesting line of research would be to investigate whether clustering can improve prediction in cases involving very different consumption profiles. Finally, when (big) data size, privacy risks, and execution times are all considered to be important constraints, using symbolic representation of smart meter data could be a valuable alternative solution [16].

References

[1] S. Amin and B. Wollenberg, "Toward a smart grid: power delivery for the 21st century," *IEEE Power and Energy Magazine*, vol. 3, no. 5, 2005.

[2] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings, "Putting the 'smarts' into the smart grid: a grand challenge for artificial intelligence," *Commun. ACM*, vol. 55, no. 4, pp. 86–97, Apr. 2012.

[3] T. K. Wijaya, K. Larson, and K. Aberer, "Matching demand with supply in the smart grid using agent-based multiunit auction," in *The Fifth International Conference on Communication Systems and Networks (COMSNETS)*, 2013, pp. 1–6.

[4] H. Daneshi, M. Shahidehpour, and A. Choobbari, "Long-term load forecasting in electricity market," in *IEEE International Conference on Electro/Information Technology, 2008. EIT 2008.*, 2008, pp. 395–400.

[5] G. Gross and F. Galiana, "Short-term load forecasting," *Proceedings of the IEEE*, vol. 75, no. 12, pp. 1558–1573, 1987.

[6] J. W. Taylor, "Triple seasonal methods for short-term electricity demand forecasting," *European Journal of Operational Research*, July 2010.

[7] H. Hippert, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, February 2001.

[8] S. B. Taieb and R. J. Hyndman, "A gradient boosting approach to the Kaggle load forecasting competition," *International Journal of Forecasting*, 2013.

[9] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on EUNITE competition 2001," *IEEE Transactions on Power Systems*, vol. 19, no. 4, 2004.

[10] N. Sapankevych and R. Sankar, "Time series prediction using support vector machines: A survey," *IEEE Computational Intelligence Magazine*, May 2009.

[11] F. Oldewurtel, A. Ulbig, A. Parisio, G. Andersson, and M. Morari, "Reducing peak electricity demand in building climate control using real-time pricing and model predictive control," in *The 49th IEEE Conference on Decision and Control (CDC),*, 2010, pp. 1927–1932.

[12] "PJM Empirical Analysis of Demand Response Baseline Methods," White Paper, KEMA, Inc., Apr. 2011.

[13] D. Wu, Y. Ke, J. X. Yu, P. Yu, and L. Chen, "Detecting leaders from correlated time series," *Database Systems for Advanced Applications*, 2010.

[14] "Smart Metering Trial Data Publication," Irish CER, 2012.

[15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGBDD Explorations*, vol. 11, no. 1, 2009.

[16] T. K. Wijaya, J. Eberle, and K. Aberer, "Symbolic representation of smart meter data," in *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, ser. EDBT '13. New York, NY, USA: ACM, 2013, pp. 242–248.

[17] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[18] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.