Electricity load forecasting at single household level using disaggregated data

Katarzyna M. Zawadzka
Instituto Superior Técnico, University of Lisbon
Lisbon, Portugal
kas.z@onet.eu
October, 2019

Abstract **- The main purpose of the following research is to further contribute to the study of demand forecasts at the individual household level. In order to do this, nine models were built to predict one-hour-ahead energy consumption. Three different scenarios were tested in terms of distribution and amount of input data: forecast model based on aggregated data; forecast model based on disaggregated data collected during direct metering (with one-hour resolution) and forecast model based on data disaggregated by NILM algorithm. In addition, three different machine learning algorithms were used, namely: random forest (RF), multi-layer perceptron (MLP) and long short term memory (LSTM). The models were built for one house and additionally tested for four others to check their scope of application. The results obtained in the analysis showed that models using NILM data provided similar results as models using data with direct measurements of energy consumption by individual devices. Moreover, in almost every case considered data containing only aggregated data provided the most accurate results to perform aggregated consumption forecast obtaining MAPE at the level of 27.6% in the best case scenario. Of the three algorithms tested, RF provided the best results, and was also the fastest and easiest to implement.**

Keywords: load forecasting, disaggregated data, NILM, machine learning, individual household

I. INTRODUCTION

According to the data provided by European Commission [1], buildings are one of the largest energy consumers in Europe. In 2017 [2], the residential sector contributed to consumption of 27.2% of final energy in the European Union, which made it the second most energy-intensive sector, right after transport (34.2%). Moreover, numerous factors indicate the rising trend in energy consumption, such as growth of number of electrical appliances in homes, increased demand of comfort as well as overall rise in the number of households [3]. The above presented numbers and trends suggest that there is a vast potential for energy efficiency gains in the residential sector.

In addition, the use of renewable sources such as solar or wind power is characterized by high volatility in the production of electricity, therefore, one of the biggest challenges to increase the integration of energy coming from renewable sources into transmission networks is to accurately control the system. To effectively balance real-time demand and supply, avoid the congestion of the grid, and be able to retain secure, competitive and sustainable energy systems, the development of methodologies that allow as accurate as possible to forecast production as well as consumption of the electricity may prove to be crucial.

Moreover, there has been a significant drop of prices of solar panels and battery-based storage systems. Together with a variety of incentives offered by the governments, this created a great demand and interest in installing these systems in the residential sector resulting in an increased number of residential energy self-producers, prosumers and local energy communities. All of these factors, connected with the general trend of switching from energy sales market to services based market, created a new business sector, where the production and sale of energy between a typical producer - consumer model is not the dominant model anymore. The prosumer model has disturbed the current market and opened the way for a new system where energy management will become one of the most important elements of the chain.

Further control over electricity supply and demand can be achieved by continuous investments and improvements in technologies such as Advanced Metering Infrastructure (AMI) [4]. In recent years, projects related to smart networks in Europe have grown significantly - currently, the database contains more than 950 R&D and demonstration projects, which totalled over 5 billion euro investment.

With the smart meter deployment, the sector dealing with Non-Intrusive Load Monitoring (NILM) has significantly developed over the last years, as it allows to extract the energy consumption profiles of individual devices from the total energy profile obtained through smart meters.

All this indicates an increasing demand for accurate control of electricity production and consumption, and this requires accurate forecasting of these two dimensions. Smart sensors and increasingly better machine learning algorithms create ideal conditions for the development of this field. And indeed, from year to year, more and more articles on forecasting electricity consumption can be found. And although the majority of them are still focusing on aggregated data on the system demand level, the topic of forecasts with the use of disaggregated data at the level of individual houses is discussed more and more frequently.

A. Research goal

The main purpose of the following research is to further contribute to the study of demand forecasts at the household level, striving to improve forecasting by adding consumption data to the models at a disaggregated level, and thus checking whether direct measurements of energy consumption by individual devices or NILM results improve the forecasting of total energy consumption in individual households. In order to obtain the most accurate forecast model, different machine learning methods, several configurations of those and various data representation will be implemented and compared. For that purpose, five different houses will be compared in three following scenarios:

1. Power consumption forecast model based on aggregated data.
2. Power consumption forecast model based on disaggregated data collected during direct metering.
3. Power consumption forecast model based on data disaggregated by NILM algorithm provided by Watt-IS company.

All models will be used to forecast household demand one hour in advance.

Furthermore, answers to the following questions will be sought:

- *Is the same model sufficient to forecast power consumption of different houses?* - A model will be built and optimized specially for on single household. After that, the same model's performance will be tested on the other households with different power consumption patterns. The level of forecast accuracy will be compared.
- *Is there an advantage in using disaggregated data over aggregated data?* - Models built with disaggregated data, thanks to capacity to process more detailed patterns, are expected to perform better than models based on aggregated data. However, this group of models requires more complex collection of the data and results in higher complexity of the model and increased computation time. It will be discussed whether the possible increase in accuracy rewards the additional model costs mentioned above.
- *Is NILM algorithm necessary to significantly improve the forecast?* - Despite the fact that models based on aggregated and NILM data require the same data (total consumption), the latter approach results in higher complexity because of the additional algorithm used for data disaggregation. Again, the level of forecast accuracy of two models will be compared.

II. LITERATURE REVIEW

Despite the growing interest in forecasting power consumption, still relatively few scientific papers focus on models based on disaggregated data. For the overall view, various methodologies and scenarios have been analyzed before proceeding to the target model development.

Gajowniczek and Ząbkowski [4] focus on 24 hour ahead electricity consumption forecast for individual household. In their paper they confirm that forecast for a single unit generates a larger error than forecasting for the cluster of houses. To increase the accuracy of their model, they create matrix with the probabilities of "appliance turn ON" events in each hour. Testing models like RF, ARIMA, ANN and SVR they were able to obtain MAPE in the range 26.8-56.3%.

Rodrigues, Cardeira and Calado [5] focused on 93 households in Lisbon, Portugal. Their main objective was to prove that it is possible to obtain satisfactory results without weather data, which is usually used in load forecast models. Using FFNN model and data such as historical consumption, apartment area and the number of occupants, they were able to obtain a mean absolute percentage error at the level of 10 % for daily consumption forecast.

Oprea, Pîrjan et al. [6] used sensor recorded and disaggregated data to forecast electricity consumption of 8 residential houses with renewable electricity generation system and smart metering system. With mixed ANN model using both non-linear autoregressive with exogenous input (NARX) ANNs and function fitting neural networks (FITNETs) they test three different training algorithms (LM, BR, SCG) and apply gap-filling technique based on the interpolation. In their paper two case studies were investigated: the first one for the total electricity consumption from the grid and the second one for the total electricity consumption of all the individual appliances. Both approaches (NARX and FITNETs) offered a very satisfying predictions with correlation coefficient ranging between 0.961 and 0.998 depending on the scenario and training algorithm.

Dinesh, Makonin and Bajic [7] first decompose the aggregate power signal into individual appliance signals. Those signals are used to make a forecast of the power consumption of each device separately. Finally, the overall forecast is created by summing up individual projections. Their original model is based on identifying ON and OFF states of the appliances - using graph construction, spectral representation and clustering, they predict the ON-set of devices at time t. The proposed model was compared with four other methodologies (ANN, ARIMA, SPLF) and in each of the considered scenarios it obtained the best results.

Sereno and Alves [8] used data on exergy consumption (disaggregated on different types of exergy) collected over nearly 50 years in Portugal to build an ANN model applying the LM algorithm. In data pre-processing, all the values were normalized to obtain null average and unitary standard deviation. After that, for optimal features selection, correlations analysis and principal components analysis were implemented. The accuracy of the model has been compared with models based on multiple and multivariate linear regression and a smaller forecast error was recorded while using the ANN method.

III. METHODOLOGY

The methodology adopted in this work strictly adheres to the standard workflow in a machine learning project [9]. The first step was to conduct exploratory data analysis in data sets, then it was necessary to process the data to prepare the application of the machine learning algorithm. In the next steps, the models that best fit the problem were selected, and the parameters describing these models were defined. Finally, the models were implemented and evaluated to obtain the final model.

A. Exploratory Data Analysis

One of the key sub-steps during the machine learning project is to perform an exploratory data analysis (EDA). During this process, a preliminary analysis of various attributes of a given data set is carried out. It is important to become familiar with their properties, characteristic points and dependencies. It is also a step into which data visualization technique should be performed in order to better understand the data specificity, as well as to be able to discover previously invisible patterns [10].

Available data comes from 100 residential houses in France. The collected data includes the timestamp, total consumption data, appliances consumption data and weather data. The dataset was collected on a similar time span for all houses (slightly over 1 year) with hourly resolution, so the number of timestamps is in range from 8955 to 9136. From the whole dataset, five houses that significantly differ in the level of electricity use were chosen to develop and test the methodology. The average daily power usage profile of them are presented below (Figure 1).

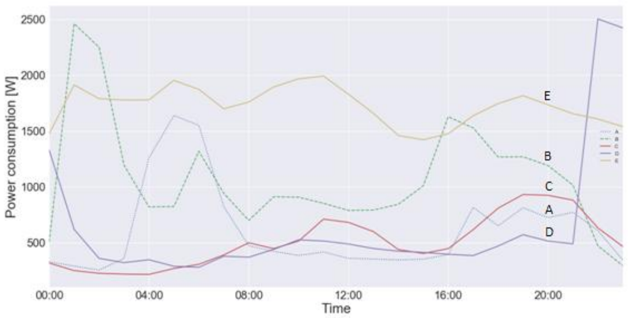

Figure 1. The average power usage during the day in each selected household (A-E).

In case of households A, B and D only 65% to 70% of the total electricity consumption is described by the direct measurements of the appliances. The remaining part, 29.6% in household B, 34.3% in household D and 34.8% in household A, respectively, is the consumption of unspecified electric appliances. Extreme cases represent the other analyzed households - when it comes to C, as much as 62.3% of energy consumed came from unregistered devices. In the case of E, the opposite is true, namely only 6% of the total energy was consumed by unspecified devices, and as much as 94% by directly measured devices.

B. Data processing

At the very beginning, data is often unstructured and contains missing and noisy data, which can significantly confuse the algorithms in the training process. Therefore, data processing, i.e. the process of transforming the input data into the most appropriate form for ML algorithms, has been carried out. These included the following operations:

1. *Identifying missing values and removing them*: zero values, that in any of the used datasets did not exceed 1%, were completely removed. The decision was made on the basis o the fact that the number of samples is just a small fraction of the rest of the data and assigning the estimated values to the missing records could introduce errors to the data due to the highly dynamic changing patterns of power use in a single household.
2. *Identifying outliers and assigning new values to them*: outliers present in the data were identified using interquartile range (IQR), which is commonly used as a measure of how spread-out the values are. The 1.5 x IQR rule implies [11] that all values that are outside the range [Q1-1.5xIQR; Q3+1.5xIQR], where Q1 and Q3 represent the values of the first and third quartiles, and IQR equals the difference of these quartiles, can be considered an outlier.
3. *Feature extraction and engineering*: the following steps were carried out:
   - extraction of 4 new features - number of the month (1-12), day (1-31), hour (0-23) and day of the week (0 - Monday, 6 - Sunday) - from the time stamp available in the raw data;
   - addition of the season designation - value 0 for winter months (1,2,12), 1 for spring and autumn months (3-5, 9-11) and 2 for summer months (6-8);
   - adding a column ('Holiday') containing information on non-working days (value 1 for non-working days, value 0 for working days);
   - addition of a column containing the level of power used by unmeasured devices (overall consumption minus the sum of the load of measured devices);
   - combining 'Weekday' and 'Holiday' features into one (by multiplying their value);
   - creation of columns containing power consumption in the previous hour of each device;
   - creation of column containing total power consumption at the corresponding day and hour from previous week.
4. *Feature scaling*: in the case discussed in this paper, the power consumption in the previous hour may range from 20 W to 4500 W, and temperature fluctuations are on a scale from -9,6°C to +31,6°C. Here, the first feature with high magnitude will weigh a lot more in distance calculations then the other, smaller in magnitude, feature. To overcome this effect, all

features should be brought to the same level of magnitudes. The three most popular methods have been tested (mean normalization, data scaling, data standardization), of which data standardization was ultimately chosen for the best prediction results.
5. *Feature selection*: 4 subsets have been prepared: 1 set with all the available features and then 3 subsets that result from the application of 3 feature selection algorithms: Extra Trees Classifier (ETC), Recursive Feature Elimination (RFE), "K-Best" feature selection. The final models were built using all available features due to the best predictive results.

C. Model Selection

Looking at the current solutions in the subject of forecasting electricity consumption, three models were selected for analysis:

1. Random Forests (RF) - an example of ensemble method where the main idea is to combine (ensemble) weak learners to build a strong learner - a robust model, which has a better generalization error and is less prone to overfitting [12]. Therefore, RF action is based on training many decision-tree models on subsets of the input data and input features. Each decision tree includes a random subset of features and only reaches to a random set of the training data points. These individual trees then contribute to the total outcome. RF models are quite robust models resistant to interferences appearing in the data. They also do not require fine tuning of many hyperparameters, but in return represent a strong generalization.
2. Multi-Layer Perceptron (MLP) - based on combining multiple Perceptrons and composed of three main elements: one input layer, one (or more) layers of LTUs, called hidden layers, and one final layer of LTUs called the output layer. This architecture is used in wide range of applications from different areas and time series forecasting is among the most important ones. The beginning of vast applicability of those networks started with presentation of the learning algorithm called backpropagation which is now used for MLP's training. MLP networks are the examples of feedforward architecture, which means that the flow within the network occurs only in one direction, namely from the input to the output layer.
3. Long-Short Time Memory (LSTM) - an example of a Recurrent Neural Network (RNN). One of the problems of typical RNN networks is that if the sequence fed into the network is extremely long, then there may be a problem with the transfer of information from earlier time steps to later ones. In this way, information from the beginning of the learning process may be lost. LSTM is one of the solutions of the short-term memory mentioned above. Thanks to the built-in internal elements called gates, it can regulate the flow of relevant information in the learning process. These gates can decide which data in the sequence is important to keep or discard. In this way, it can convey relevant information along a long chain of sequences to make predictions.

D. Parameters optimization

To provide the best results, it is rarely possible to use the basic set of algorithms available in various ML dedicated libraries. These algorithms have various specific parameters whose values must be properly tuned to the different data, project requirements, and expected results. Bearing in mind the bias-variance trade-off, a so-called good fit is sought to avoid model underfitting (the algorithm does not learn the hidden data structure) or model overfitting (the algorithm is too adapted to the training data set). After a preliminary analysis determining which of the parameters has the greatest impact on selected models, 3 parameters in the RF model were selected for optimization (the number of trees in the forest, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node), 5 in the MLP model (activation function, parameter alpha, the number of neurons in each hidden layer, the number of hidden layers, maximum number of iterations) and 2 in the LSTM model (the number of neurons, the number of epochs).

The optimization process consisted of three phases in which another set of parameters was tested. For RF and MLP models, optimal parameters were found using two additional algorithms: Random Search and Grid Search. Following the practices of the authors of various studies (like [4], [5], [8], [13]) the data for both models was divided as follows: 80% for training and 20% for testing. In addition, 5-cross validation was used (dividing the training set into 5 subsets and using one of them for validation). In the case of the LSTM model, due to the source code, the search for final parameters was carried out manually, which was considered to be sufficient for prototyping. In this case, the data has been divided into three parts in advance: 70% for the training set, 20% for the testing set and 10% for the validating set.

Hyperparameter optimization was carried out for household A, and then the best configurations were used for the next datasets.

E. Model evaluation

After building the model, it is important to subject it to an assessment process based on specific criteria for assessing its accuracy and performance. This is most often done using pre-defined functions that allow you to get a numerical value that helps to decide on the effectiveness of a given model and allows to compare models with each other. In order to accomplish that, the following parameters were measured and used:

1. Mean Absolute Percentage Error (MAPE)
2. Mean Squared Error (MSE)
3. Mean Absolute Error (MAE)

4. Coefficient Of Determination (R2)

IV. RESULTS

The following results were obtained for forecasting household A one-hour-ahead energy consumption. In addition to the measured evaluation values, the time required for the learning process of each model is given.

From the data presented below (Table 1), it can be seen that the best results for household A were obtained using the RF model for aggregated data. The results are similar for the model using data from direct measurements, and in this case, the values do not differ significantly from those obtained with the MLP and LSTM models. For NILM and aggregated data, both MPL and LSTM present significant higher errors that RF. In addition, RF is the model that required one of the shortest training time. In conclusion, it can be said that RF seems to be the best model for the case considered and that for aggregated consumption forecast, the use of direct consumption of appliances or NILM information does not improve the results.

Table 1. Evaluation of household A.

		MAPE	MSE [kWh^2]	MAE [kWh]	R^2	Time [s]
Direct meter.	RF	31.9 %	90.59	0.15	0.74	14.4
	MLP	41.6 %	98.97	0.18	0.71	15.8
	LSTM	37.0 %	110.58	0.19	0.74	1492.3
NILM	RF	36.4 %	101.77	0.18	0.75	14.7
	MLP	52.2 %	141.44	0.22	0.66	15.5
	LSTM	39.2%	125.55	0.21	0.75	2539.2
Aggreg.	RF	**30.9 %**	**86.84**	**0.16**	**0.64**	**14.7**
	MLP	40.6 %	105.72	0.19	0.56	57.1
	LSTM	59.1 %	137.45	0.24	0.43	107.7

The results of the assessment of the best scenarios in the other four households (RF built on aggregated data for houses B, C, D and RF built on data from direct measurements for house E) are presented below (Table 2). In all cases considered again, the model that provided the most accurate results was RF with MAPE = 27.6% in the best scenario and an MAE between 0.1 kWh and 0.39 kWh, depending on the type of data and the household. The MLP model worked slightly worse, but in some cases the results of both models are comparable. Also the training times of both models are similar (from 9 to 19 seconds with one exception). The LSTM model proved to be the least accurate and additionally requires a much longer training time (over 16 minutes on average). The reason for this behaviour of the LSTM model may be that relatively few data points (less than 8000 per household) have been used, and these models require more data for proper training than models such as RF.

Table 2. Evaluation of the best scenario in the houses B-E.

	MAPE	MSE [kWh2]	MAE [kWh]	R2	Time [s]
B	63.9 %	346.05	0.33	0.77	9.8
C	31.4 %	49.38	0.13	0.52	9.9
D	27.6%	28.53	0.10	0.40	9.2
E	38.6%	255.21	0.29	0.88	17.2

Considering only the best performing model (RF), the values of individual evaluation parameters are visualized below (Figure 2). With the exception of house E and the MAE values of house A and B, again, the lowest MAPE, MSE and MAE values were obtained when forecasting using data set only with aggregated data, therefore excluding the power consumption of individual devices.

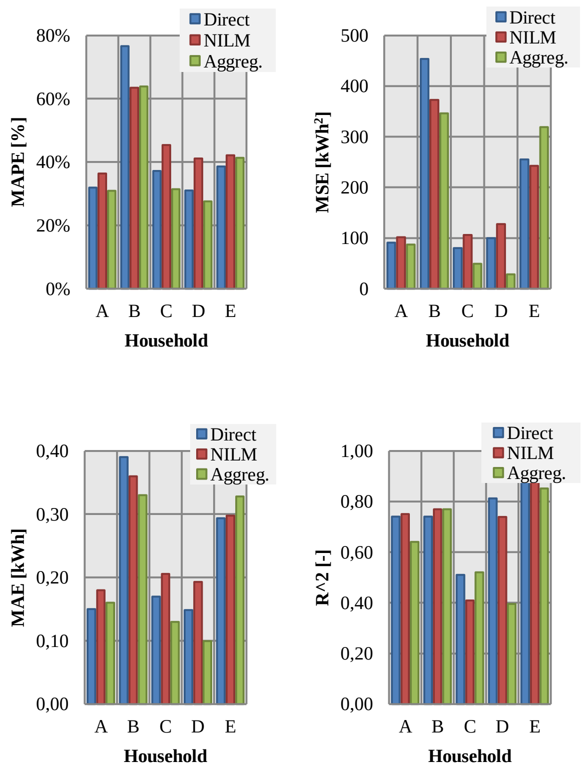

Figure 2. Evaluation parameters visualization : MAPE (top left), MSE (top right), MAE (bottom left) and R2 (bottom right).

Therefore, in all cases, except household E, the use of aggregated data is better or at least as good as the use of direct data. Only in household B, the use of NILM information is comparable to the aggregated forecast. This means that for aggregated forecast of energy consumption, providing detailed information of the appliances consumption (through direct measurements or NILM estimations) does not improve the forecasting results.

V. DISCUSSION

In the following section, based on the results presented in the previous chapter, an attempt will be made to answer the research questions posed in section one.

Is there an advantage in using disaggregated data over aggregated data?

It was expected that models based on disaggregated data, due to the possibility of more detailed pattern recognition, could work better than models based on aggregated data. Looking at the results in Figure 2, it turned out that this did not happen in general. In almost every case considered (A-E homes; RF, MLP, LSTM algorithms) data containing only aggregated data provided the most accurate results to perform aggregated consumption forecast. Moreover, it was easier and faster to process, both during data preparation and model training. Less accurate results obtained on disaggregated data may be due to insufficient amount of data, which did not allow individual models to notice and extract more intricate patterns of device behaviour. This caused that these additional features only unnecessarily confuse the model and do not bring the expected benefits. Therefore, the number of devices considered should be limited. The more immediate focus should be only on the most energy-intensive devices (e.g. heating, air conditioning), and those that are switched on sporadically, consume an extremely small amount of energy or are used in the same way throughout the year (e.g. TV) should be discarded.

Also, this encourages to investigate different approach in the future: focusing more on the probabilities of using a given device in a given hour of the day than on the power it consumes. For this it is worth using the measured values at shorter intervals (5-30 min), because a large number of appliances in the household has a duty cycle shorter than one hour. This approach will, probably, allow to better observe different behavioural patterns and obtain more accurate forecasting results.

Is NILM algorithm necessary to significantly improve the forecast?

The results obtained in the analysis showed that models using NILM data provided similar results as models using data with direct measurements of energy consumption by individual devices. Still, the first option is much easier and cheaper to obtain than the second, which requires expensive installation and supervision of a large number of sensors. However, compared to aggregated data, results obtained using NILM data are less accurate. Moreover, the latter requires an additional step, which is the separation of total consumption into individual devices or device subgroups.

Worse results may be due to the fact that the algorithm for obtaining disaggregated data was not tuned enough to individual houses. In addition, different cases should be re-examined using the approach suggested above - using data from only the most energetically intensive devices with less than one-hour time steps.

Is the same model sufficient to forecast power consumption of different houses?

The created model was built and optimized specifically for a single household (A), and then the performance of the same model was tested in other households with different energy consumption patterns. Due to the multitude of models, it was decided to focus on the results provided by the most accurate model (Table 3).

Table 3. Evaluation of RF model based on aggregated data of different households (A-E).

	MAPE	MSE [kWh^2]	MAE [kWh]	R^2	Time [s]	Daily avr. [kWh]
A	30.9 %	86.84	0.16	0.64	14.7	14.7
B	63.9 %	346.05	0.33	0.77	9.8	25.7
C	31.4 %	49.38	0.13	0.52	9.9	12.1
D	27.6%	28.53	0.10	0.40	9.2	15.2
E	41.3%	318.88	0.33	0.85	16.5	41.7

The highest values of MAPE, MSE and MAE were obtained for houses B and E, but it should be taken into account that they were also the most energy-intensive cases. On the other hand, they are distinguished by the highest R^2 result (higher than for house A) - various energy peaks appearing in their test data sets have been very well mapped in the model. House D presents a completely opposite situation - although it has a similar average daily energy consumption to the reference house A, it obtained the best results of MAPE, MSE and IEA, and the worst when it comes to R^2.

The results obtained indicate that similar results should not be expected even if the format and amount of data fed with the algorithm is very similar. Individual houses, which in the vast majority of cases, are characterized by exceptional irregularity of energy use and differentiation of the energy level, are extremely difficult to reproduce in one universal model. The next step would be to compare the accuracy of individual models on different subgroups of houses - clustered by similar average energy consumption, a similar pattern of energy use during the day and others.

VI. CONCLUSIONS

The main goal of the presented research was to examine whether direct measurements of energy consumption by individual devices improve the forecasting of total energy consumption in an individual household. In addition, it was checked whether a single model is able to provide results at a similar level of accuracy for different houses, where the fed data for the algorithm are of similar structure and size. During the search for the most accurate forecast model, various machine learning methods, several configurations and different data representations were implemented and compared. Five different households were used to verify the results, given by three different datasets: aggregated power consumption, disaggregated data collected during direct measurement of power consumption of different devices and disaggregated data according to the NILM algorithm provided by Watt-IS. The results obtained did not show the superiority of disaggregated data over aggregated data. Nevertheless, it is recommended to carry out further research with the following improvements proposed below:

- check larger parameter ranges (more accurate model optimization);
- analyze other machine learning algorithms (e.g. SVR);
- create separate models for forecasting the energy consumption of individual devices and then add them together to obtain the total electricity consumption;
- use the measurements of individual devices to create a map of the probability of switching on particular devices at a given hour during the week and using it as an additional feature fed in the algorithm;
- use data with less than one-hour time step (e.g. 10 minutes or even 1 minute);
- use data collected over a longer time period than one year to avoid under- and overfitting of the algorithms;
- carry out research on a more recent data set (different patterns and energy levels may have changed significantly over the last 20 years);
- analyze models with a larger forecasting horizon (e.g. 3 hours / 24 hours ahead).

The use of energy at higher levels, when many consumers are aggregated, forms a smooth pattern with clearly marked and very often repeatable consumption peaks. Accurate estimation of energy consumption at the level of a single household provides exceptional difficulties due to the extremely variable pattern of energy consumption. Still, due to the increasing penetration of renewable energy in the network, increasing efforts to reduce energy consumption, the emergence of individual power generators (e.g. through residential solar panels), as well as the need to ensure the most secure supply of energy, increased efforts to energy management improvements must be undertaken. This in turn may mean a compulsory step towards forecasting energy consumption at the level of a single household and NILM disaggregation may be a key to high accuracy of those forecasts.

REFERENCES

[1] European Commission, "Clean energy for all Europeans," Publications Office of the European Union, Luxembourg , 2019.

[2] Eurostat, "Final energy consumption by sector," 23 04 2019. [Online]. Available: https://ec.europa.eu/eurostat/web/products-datasets/-/ten00124. [Accessed 13 10 2019].

[3] Eurostat, "Manual for statistics on energy consumption in households," Publications Office of the European Union, Luxembourg, 2013.

[4] K. Gajowniczek i T. Zabkowski, „Electricity forecasting on the individual household level enhanced based on activity patterns," *PLoS ONE,* tom 12, nr 4, pp. 1-26, 2017.

[5] F. Rodrigues, C. Cardeira and J. M. Calado, "The daily and hourly energy consumption and load forecasting using artificial neural network method: A case study using a set of 93 households in Portugal," *Energy Procedia,* vol. 62, pp. 220-229, 2014.

[6] S. V. Oprea, A. Pîrjan, G. Cărutasu, D. M. Petrosanu, A. Bâra, J. L. Stănică and C. Coculescu, "Developing a mixed neural network approach to forecast the residential electricity consumption based on sensor recorded data," *Sensors (Switzerland),* vol. 18, no. 5, 2018.

[7] C. Dinesh, S. Makonin and I. V. Bajic, "Residential Power Forecasting Using Load Identification and Graph Spectral Clustering," *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 7747, 2019.

[8] R. Sereno and V. Alves, "Application of Neural Networks' Models to Predict Energy Consumption," Instituto Superior Técnico, Universidade de Lisboa, Lisbon, 2016.

[9] D. Sarkar, R. Bali i T. Ghosh, „Machine Learning Fundamentals," w *Hands-On Transfer Learning with Python : Implement Advanced Deep Learning and Neural Network Models Using TensorFlow and Keras,* Packt Publishing, 2018, pp. 9-29.

[10] P. Ankur A., Hands-On Unsupervised Learning Using Python: How to Build Applied Machine Learning Solutions From Unlabeled Data, Sebastopol: O'Reilly Media, Inc., 2019.

[11] T. A. Chawsheen and I. S. Latif, "Detection and treatment of outliers in data sets," *Iraqi Journal of Statistical Science,* vol. 9, pp. 58-74, 2006.

[12] S. Raschka, „A Tour of Machine Learning Classifiers Using Scikit-learn," w *Python Machine Learning,* Birmingham , Packt Publishing Ltd., 2015, pp. 90-92.

[13] J. Moon, J. Park, E. Hwang and S. Jun, "Forecasting power consumption for higher educational institutions based on machine learning," *Journal of Supercomputing,* vol. 74, no. 8, pp. 3778-3800, 2018.