

Short-Term Forecasting of Hourly Electricity Power Demand

Regression and Cluster Methods for Short-Term Prognosis

Simona Filipova-Petrakieva

Department of Theory of Electrical Engineering
Faculty of Automation
Technical University of Sofia
Sofia, Bulgaria
petrakievas-te@tu-sofia.bg

Vencislav Dochev

Department of Computer Systems
Faculty of Computer Systems and Technologies
Technical University of Sofia
Sofia, Bulgaria
ventsidochev@abv.bg

Received: 28 January 2022 | Accepted: 10 February 2022

Abstract—The optimal use of electric power consumption is a fundamental indicator of the normal use of energy resources. Its quantity depends on the loads connected to the electric power grid, which are measured on an hourly basis. This paper examines forecasting methods for hourly electrical power demands for 7 days. Data for the period of 1 January 2015 and 24 December 2020 were processed, while the models' forecasts were tested on actual power load data between 25 and 31 December 2020, obtained from the Energy System Operator of the Republic of Bulgaria. Two groups of methods were used for the prognosis: classical regression methods and clustering algorithms. The first group included "moving window" and ARIMA, while the second examined K-Means, Time Series K-Means, Mini Batch K-Means, Agglomerative clustering, and OPTICS. The results showed high accuracy of the forecasts for the prognosis period.

Keywords—short-term prognosis; hourly electricity power demand; regression analysis; clustering methods

I. INTRODUCTION

Electric power consumption is a key to the normal function of any economy. In recent years, the consumption of electrical power has increased along with the human population and the development of technology. Proper use of the available energy resources is a necessity since solid fuels are finite, they will run out in the future, and, their extraction and processing pollute the environment. For these reasons, various types of renewable energy sources have emerged. Unfortunately, for some of them, the generation of electricity depends on the season, geographical location, type of energy extraction technology, certain economic and political factors, etc. Also, the storage of this type of energy is difficult and costly. On the other hand, the production of electric power should neither exceed nor fall short of the amount of electric power required by the end-users. For this reason, algorithms have been developed for power generation in hydraulic power plants, guaranteeing the optimal parameters of the produced energy under the dynamic change of requirements from the electric supply companies and the market [1, 2]. Despite the physical nature of the predicted subject, the forecasts are long-term, medium-term, and short-term. The requirements that any reliable forecast must satisfy

were given in [3]. There are various models for determining forecast assessments using different mathematical approaches to obtain the prognoses. For the operational control of power plants, it would be desirable to obtain precise short-term load forecasting assessments to guarantee power supply and load dispatch. The Empirical Mode Decomposition (EMD) method and the Particle Swarm Optimization (PSO) algorithm were successfully hybridized with the Support Vector Regression (SVR) to produce a satisfactory forecast performance in [4]. Decomposed Intrinsic Mode Functions (IMFs) are defined as three items: 1) containing the random and the middle term, 2) containing the middle and the trend (residual) term, and 3) containing the middle terms only, where the random term represents the high-frequency part of the electric power load data, the middle term represents the multiple-frequency part, and the trend term represents the low-frequency part. Based on these assumptions, the SVR-PSO was created, and the forecast results were calculated as $(1) + (2) - (3)$. The suggested model improved the forecast accuracy, while the data for the model synthesis were taken from the Australian electricity market.

Short-term power demand forecasts contribute significantly to the synthesis of the smart grid. In [5], a deep model based on Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and Discrete Wavelet Transform (DWT) was proposed. This model was divided into two parts. The first part corresponded to the time-domain feature extraction, while DWT was responsible for the frequency domain. The model extracted both time and frequency-domain features separately using the neural network. Subsequently, they merged as time-frequency features. The latter was fed into LSTM to mine the features, which had a long-time dependency. In [6], a daily power consumption forecast was created for the high-temperature period. The model was based on a portrait-based multivariate regression model. The portrait of each substation area was derived using the clustering method based on the label system. Then, the regression model was applied to forecast each cluster. The synthesized model was validated using electricity consumption data from Shanghai. A hybrid deep learning neural network framework, which combined CNN

with LSTM for making power consumption forecasts, was proposed in [7]. The original short-term forecasting strategy was extended to a multi-step forecasting strategy to introduce more response time for electricity market bidding. The following forecasting approaches were built: Auto-Regressive Integrated Moving Average (ARIMA), persistent model, SVR, and LSTM. In addition, a k-step power consumption forecasting strategy was used to promote the proposed framework for use in real-world applications.

Power production of PhotoVoltaic (PV) plants is an important way of producing green energy, and different models exist to predict their self-consumption. Two hybrid models were suggested in [8] to forecast power consumption (CNN-LSTM and ConvLSTM) were suggested in [8]. These models were found to be more accurate than the standard LSTM. In [9], a deep model named TL-MCLSTM was proposed as a multiple-output strategy to predict multi-step short-term power consumption. The model contained three channels: power consumption, time location, and customer behavior. The first channel reflects the change and the general trend of use. The second channel reflects the hidden pattern of customer habits, recording information about time, day of the week, and holidays. The third channel combines a convolution autoencoder and K-means to identify customer behavior. The first two channels were individually trained through the LSTM, as it had an excellent memory function. The extracted features from the LSTM in these channels were combined with customer behavior as comprehensive features to forecast. Other mathematical approaches for creating short-term forecasting models are Imperialist Competitive Algorithms, Support Vector Machines (SVM), and Hierarchical Cluster Analysis SVM. These approaches were used in [10] for the prognosis of the hourly electricity load. In [11], a new modeling approach was proposed for medium-term probabilistic power consumption using neural networks and incorporating trend, seasonality, and weather conditions as explicative variables in a neural network with an autoregressive feature. In [12], fuzzy logic was incorporated into a neural network. In this study, a model was divided into two subsystems: a network with back-propagation making the forecast and using data from previous months, and an autocorrelation module using temperature and production load differences for air conditioning and consumption between the previous and the forecast months. This model is appropriate for two types of industrial consumers: consumers for climate control and consumers for production activities.

A mid-long term load structure forecasting model was developed in [13], based on grey theory, where the system state equations and the grey dynamic model group on various types of electrical load were established. The model provided a mid-long-term forecast, in terms of the system dominant and associated factors determined by the grey correlative degree analysis method, using the GM(1, N, x ((0))) model derived from the GM(1, N). The power consumption of the considered grid was predicted in the medium and long term in a case study utilizing the proposed model.

Long-term forecasting of electricity consumption is quite difficult as its reliability depends on many different factors.

During daytime, there may be several peaks in consumption which lead to a complete discharge of the battery to one of the peaks. As a result, the total peak power consumption does not decrease. To optimize the operation of storage devices, a day-ahead forecast is often used, which allows determining the total number of peaks. In this sense, a long-term forecast of power consumption based on the use of exogenous parameters in the decision tree model was used in [14]. This forecast was based on the idea of determining the optimal storage capacity for a specific consumer, which optimizes the costs of leveling the load schedule. In [15], a detailed overview of current forecasting methods for power consumption was presented. Some forecast approaches were summarized as grey and artificial neural network theory, among which the forecasting principle of the Grey Model GM(1,1) was developed. The econometric model method was also discussed. In addition, comprehensive analysis, analysis prediction, and other methods, which could be applied in short-term, medium-term, and long-term electricity demand forecasting were also discussed. The accuracy of the methods mentioned above improved by applying newly developed machine learning smart algorithms, such as deep learning, Q-learning, extreme learning, etc. The processed data described the power consumption of Guangzhou city between 2000-2008, and the forecasts were made for 2009.

This study examines a short-term 7-day forecast of the hourly demand for electric power in the Republic of Bulgaria. These prognoses can be used to analyze the energy system models used for planning the Bulgarian energy market, as described in [16]. A similar prognosis was presented in 2002 [17], but these models are obsolete today and do not guarantee high forecast accuracy. The increasing price of natural gas during the recent months has led to the rising prices of the entire production, supplies, and services. The sharp increase in the price of European natural gas prices [18] leads to lifetime evaluation and the urgent need to use the available gas resources economically and appropriately, by finding ways to make the production of other types of energy resources cheaper (nuclear plants, photovoltaics, wind turbines, etc.) as well as to forecast the quantities needed as accurately as possible.

II. THE ESSENCE OF FORECASTING METHODS

The models created by all forecasting approaches for electric power demand predict the consumption for a given period, based on data for power consumption in previous periods. The forecast value is either a point or an interval, where the actual electrical load is most likely wrong. Various methods use a variety of mathematical frameworks to make such forecasts, such as regression analysis [19-24], regression analysis, neural networks, and least square SVMs [25], regression analysis, decision tree, and neural networks [26], cluster analysis [27], fractional Brownian motion [28], etc.

A. Classical Regression Analysis Methods

Regression analysis is a time series approximation of points by a specific function, usually in the form of a polynomial. The accuracy of the approximation depends on the amount of input data and the order of the approximating function. Therefore, a compromise must always be searched between the amount of

input data and the order of the approximating polynomial. This study used the following two methods to perform short-term hourly electrical power demand forecasts based on regression models using the moving window and Autoregressive Integrated Moving Average (ARIMA) methods.

- Moving window [22]: This method takes samples from the input data and applies a formula to calculate the forecast value. The moving window method is commonly used for solving practical problems. When working with averaged data, the resulting value is used as the predicted value. The accuracy of the method increases as the processed data increase. But this fact, on the other hand, increases the time to process the data until the final forecast is obtained.
- ARIMA [23, 24]: This method consists of introducing a distance between the data in the form of an error. This distance can be computed using the Euclidean or another type of norm. The predicted quantity at a given moment is computed as a function of that quantity at previous moments and/or error values.

B. Cluster Analysis Methods

For each studied time interval, the measured data of the hourly electrical power demand are obtained and form an acceptable range of variation. The accuracy of the synthesized model is tested by checking whether the actual electrical power load during the hour is between the predicted interval. If so, the method determines to which part of it falls into.

- K-Means [29, 30]: This is a very simple method, for simple applications, which converges relatively quickly. Its main weakness is that it is not stable to statistical errors. On the other hand, there is a different result for each execution. Some improvements of classic K-Means were suggested in [31], by defining a hypercube of constraints for each centroid that acquires weights for each attribute of each class to use a weighted Euclidean distance as a similarity criterion in the clustering procedure and thus reduce the limitation effects.
- Time Series (TS) K-Means [32]: This method uses Dynamic Time Distortion and Soft Dynamic Time Distortion algorithms for distance calculation between the data, instead of the Euclidean formula. This ensures higher forecast accuracy of the TS K-Means algorithm but increases significantly the data processing time.
- Mini Batch K-Means [33]: This method was designed to process large data partitioned into clusters. Its main advantage over K-Means is the reduced time to partition the data into clusters, which is proportional to the size of the clusters formed. This is very important when processing the data from hourly electrical power demand measurements for a day. Thus, 24 new values are processed each day.
- Agglomerative Clustering [34, 35]: This method belongs to methods that use a hierarchical structure and is based on the so-called bottom-up approach. The algorithm supports an "active set" of clusters and decides which two clusters to merge at each stage. When two clusters are merged, the latter disappears from the set while the newly formed is

added. This is repeated until all clusters are finally merged into a single. During the execution of the algorithm, a binary tree that accounts for the union of pairs of clusters is formed and called a dendrogram. In this method, the problem of defining the number of clusters is solved by design. There is no need to change the algorithm as new data come in, as in Time Series K-Means and Mini Batch K-Means. A basic problem in K-Means forecasting is the problem of defining the number of clusters. This is not a problem when working with fixed data ranges, but when working with time series, updating the value for the number of clusters is required, and this is not an automatic process. Agglomerative clustering solves this problem at the core of its design.

- Ordering Points To Identify the Clustering Structure (OPTICS) [36, 37]: Like Mini Batch K-Means, this method is designed to process large data which can be fast changed and temporally ordered. This method is based on the density principle. The implemented algorithm (DBSCAN) searches for the point with the highest local density and forms a cluster around it. When the radius of a neighborhood change, the algorithm keeps the hierarchy of the clusters formed so far. OPTICS has high classification accuracy and smaller error than the above-mentioned clustering methods. Its main disadvantage is the slow performance due to large data processing.

III. ANALYSIS OF THE FORECASTS' RESULTS

This study examines the forecast models using data for electricity power demand (in MWh) on the territory of the Republic of Bulgaria. Data were acquired from the Bulgarian Electricity System Operator (ESO) [38]. The analyzed period was from 1 January 2015 to 24 December 2020. The load was averaged for each hour of the investigated day. The forecast assessments were also performed hourly for each day in the period between 25 and 31 December 2020. The forecasts were compared with the actual hourly electrical power demand for this period. For simplicity, without losing generality, only the forecast results for the 21st (8-9 p.m.) hour of the prognosis days are provided. This interval was chosen as the electrical power load is at its highest, and its prediction is the most important. The application for making forecasts was written in Python 3. Two regression forecasting methods were used: regression with moving window and ARIMA. The accuracy of the ARIMA model is determined by the extended Dickey-Fuller rooting test [39] implemented using the statsmodels.tsa.stattools function [40]. This study used five cluster forecasting methods: K-Means, Time Series K-Means, Mini Batch K-Means, Agglomerative clustering, and OPTICS. The quality of the clusters obtained by each method was evaluated by applying the Silhouette method [41]. This method calculates the distances between individual points in a cluster, as well as the distances to points belonging to other clusters. This gives an idea of the density of the cluster and the fitting of each point to the others in the cluster. The forecasted electrical power demand results for the 21st hour of the day (8-9 p.m.) for the period 25th - 31st December 2020 from each method used is shown in Figures 2-8 in blue, while the actual electrical power load is marked in red.

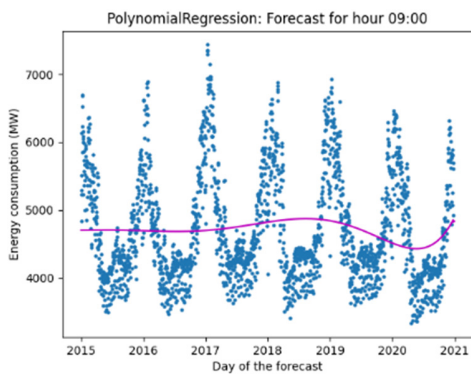


Fig. 1. Forecasts from moving window compared with real power load.

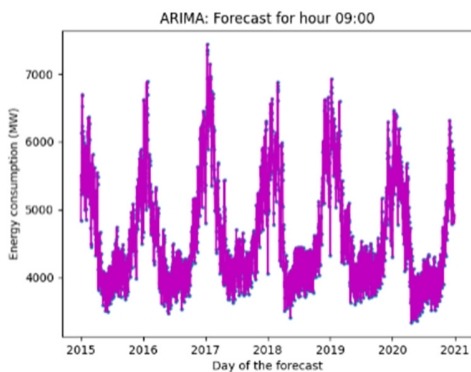


Fig. 2. Forecasts from ARIMA compared with real power load.

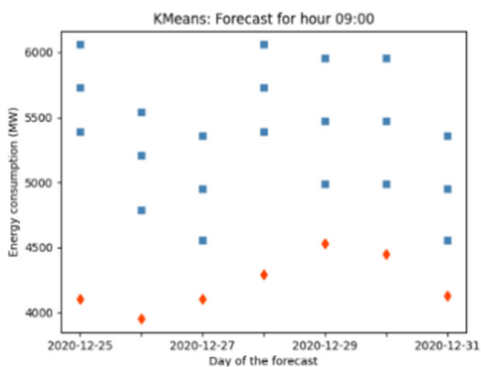


Fig. 3. Forecasts from K-Means compared with real power load.

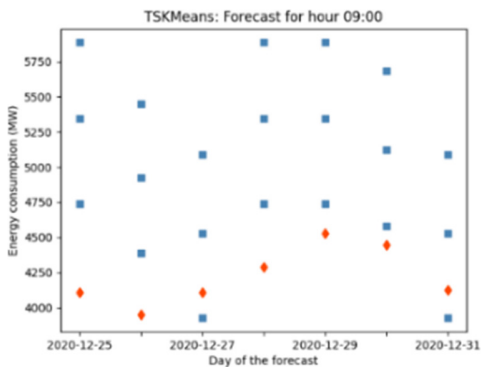


Fig. 4. Forecasts from TS K-Means compared with real power load.

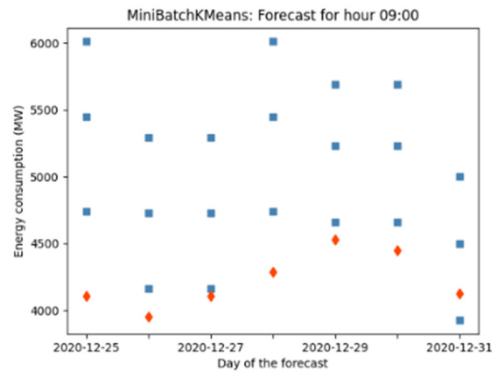


Fig. 5. Forecasts from Mini Batch K-Means compared with real power load.

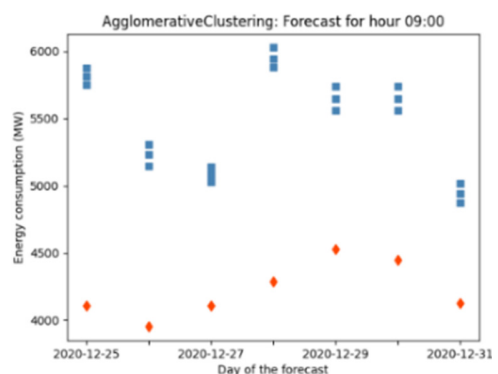


Fig. 6. Forecasts from Agglomerative clustering compared with real power load.

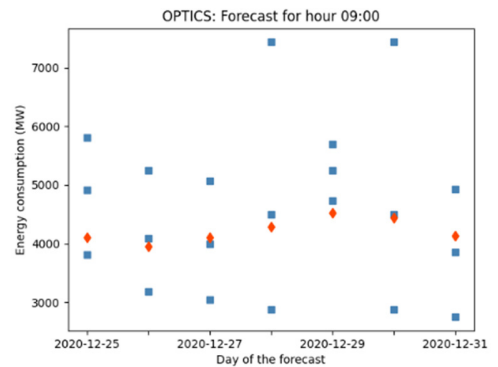


Fig. 7. Forecasts from OPTICS compared with real power load.

In both regression methods, moving window and ARIMA, the actual electrical power consumption data are dispersed around the predicted ones determined by the corresponding approximating polynomial. Therefore, the classical regression algorithms are quite inaccurate and do not follow the trend of the actual values as the cluster methods. This is because they provide point estimates of the forecasted values, while cluster algorithms provide ranges, defined by minimum, average, and maximum values. For the cluster methods, except for OPTICS, the actual hourly electrical power demands are always less than the lower bound of the respective forecast variation ranges of the load. On the one hand, this is preferable because it introduces conservatism to the resulting estimates. On the other

hand, it shows that the estimates given by the OPTICS cluster method are the most accurate. Mini Batch K-Means ranks right behind OPTICS. Its minimum forecasting assessments are the closest, mostly at the top and sometimes at the bottom, to the actual hourly power load. The TS K-Means graphs are similar to the Mini Batch K-Means, i.e. the actual hourly electrical power demands are quite close to the lower bound of the calculated range. These differences are slightly larger compared to those obtained with Mini Batch K-Means. Otherwise, they are all smaller than the corresponding differences estimated with the other clustering methods. The forecast ranges, determined by Agglomerative clustering, are the most compact (short) in comparison to the other clustering methods. However, the actual hourly power load for the study period is much smaller than the minimum forecast value for each time zone considered. Perhaps, with a larger input sample, the actual hourly electrical power demand would fall in those areas that are almost point-wise. Then the forecast assessments in the Agglomerative clustering method will be more accurate. The graphs in Figures 2-8 show that the most inaccurate estimates were obtained by K-Means and Agglomerative clustering. They contained the greatest degree of conservatism on actual electrical power loads. This is not dangerous, as it leads to a surplus of produced electric power and the users' needs will always be met.

The accuracy of each method, shown in Figures 9-15, was calculated for each hour of the day based on the resulting 7-day forecast period. The most accurate forecasts were obtained by the OPTICS method, closely followed by the Mini Batch K-Means. K-Means and Agglomerative clustering methods gave the most inaccurate prediction results. The largest error was seen in the estimate given by each method for the 7th hour of the day. The two methods using regression analysis gave similar accuracy results. For both methods, the most inaccurate estimate was for the 7th hour (0.22%), followed by about 0.5% for the 17th and 18th hours. The regression estimates were more accurate than K-Means and Agglomerative clustering, and they are comparable to TS K-Means, and Mini Batch K-Means, but contain a larger degree of dispersion around the average value of 0.2%. Last but not least, they are much less accurate than OPTICS.

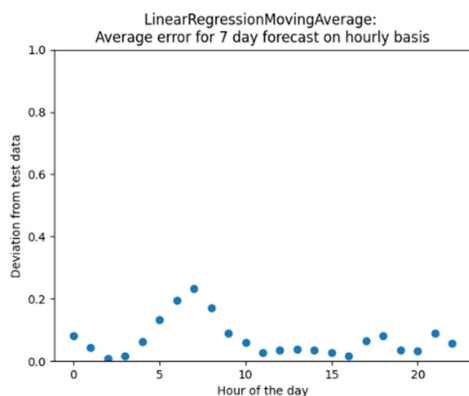


Fig. 8. Accuracy of the moving window method.

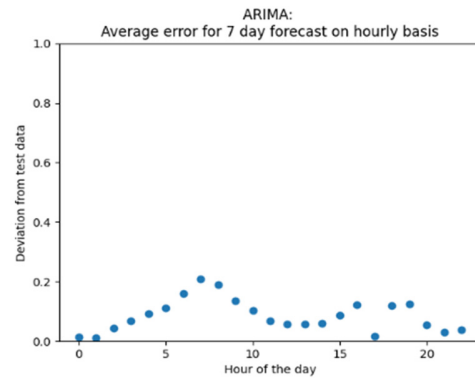


Fig. 9. Accuracy of the assessments obtained by the ARIMA method.

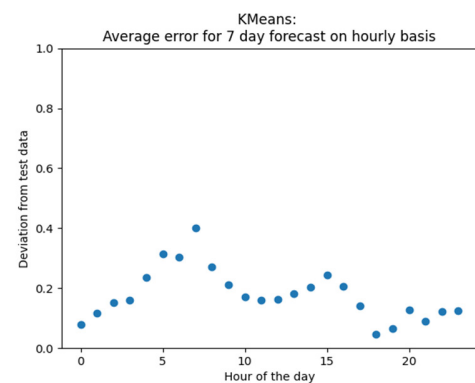


Fig. 10. Accuracy of the assessments obtained by the K-Means method.

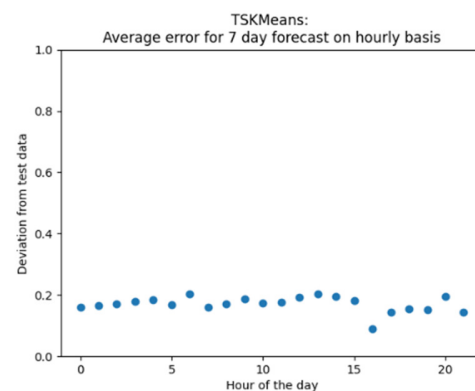


Fig. 11. Accuracy of the assessments obtained by the TS K-Means method.

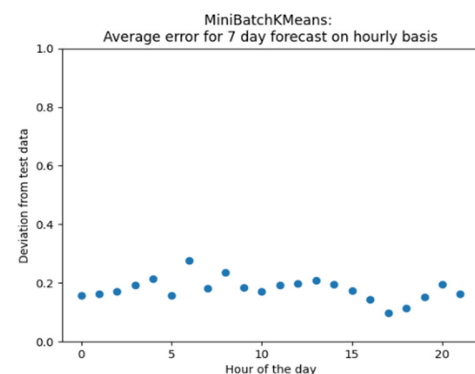


Fig. 12. Accuracy of the Mini Batch K-Means method.

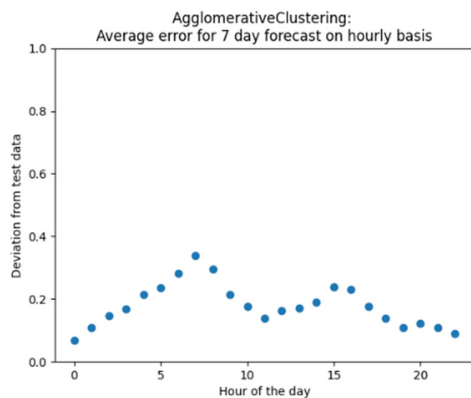


Fig. 13. Accuracy of the assessments obtained by Agglomerative clustering.

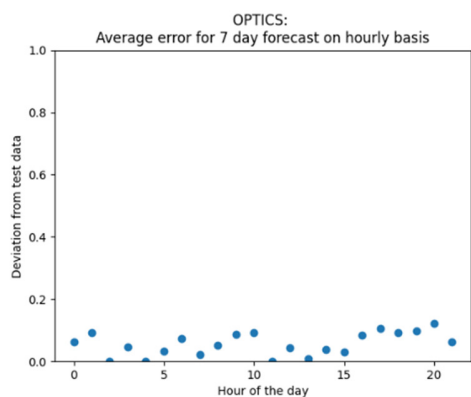


Fig. 14. Accuracy of the assessments obtained by the OPTICS method.

The most accurate and compact was the OPTICS method, varying in the interval $[0\%, 0.02\%]$ coincident with the mean of the exact interval 0.01% . It was followed by the TS K-Means, whose range of variation of estimates was very narrow $[0.18\%, 0.2\%]$ and the dominant error values were 0.19% . Mini Batch K-Means was third in terms of compactness of the forecasting error. It also provided estimates with small variance (0.18%) around the average value, and the range of variation of the estimates was $[0.9\%, 1.3\%]$. The error graphs of K-Means and Agglomerative clustering were quite similar. In both cases, the largest error was observed in the prediction for the 7th (0.4% and 0.38%) and the 15th hour of the day (0.3% and 0.28%) for K-Means and Agglomerative clustering respectively. The execution time of TS K-Means was worse by far in comparison to the other algorithms, although its performance is good enough for algorithms to be implemented and executed daily without compromising the process of forecasting. Regression methods were less accurate than the cluster ones, except K-Means and Agglomerative clustering. Another disadvantage is that they require a long time to train the models and a large amount of data to guarantee an adequate forecast, which is not typical for clustering methods. K-Means and Agglomerative clustering are in the last place in terms of forecasting accuracy, but their largest error is 0.38% and occurs only for one hour of the day, which is not fatal for short-term forecasts. In principle, the accuracy of the forecast results can be increased by increasing the amount of input data used to

train each algorithm. However, data older than 10 years should not be utilized, as the influence of many economic, political, geographic, and other factors change the trend of the forecasts and cannot be reliable.

IV. CONCLUSION AND FUTURE RESEARCH

Although the delta between the minimum and maximum forecast cluster values is smaller for insufficient data input, the results showed a decrease in accuracy with the quantity of input data. As a greater number of input points was fed to the algorithms, the predicted outcome was not as precise but much more accurate. This study trained each method with data for 4 years, 1 year, and 4 months. One conclusion is that there is no real impact from feeding the algorithms with data larger than 1 year since the accuracy remains roughly the same, but the precision is harmed.

Using additional functions and methods to adjust the needs to the strengths of the algorithms may be a great advantage, as the clustering algorithms are not very well suited to forecast time-series data. However, using the right toolset could advantage the system to be applied to an environment that does not provide enough input data, where other algorithms may suffer from data scarcity. Cluster methods' forecast results are much more accurate than those obtained by regression analysis methods. Moreover, they require fewer time resources to process the input data to train the models and obtain the final predictions. The data used in the models synthesized in this paper to perform short-term forecasts of hourly electricity power demand on the territory of the Republic of Bulgaria are universal. They can be used to perform any kind of short-term hourly forecasts of arbitrary quantities. They were applied to this case because the authors had this kind of data to process. And they are only illustrative of their applicability.

One prospect for further research could be to examine additional clustering methods, while the considered input data period could be enlarged to 10 years

ACKNOWLEDGMENT

The authors would like to thank the Research and Development Sector at the Technical University of Sofia for the financial support.

REFERENCES

- [1] A. Tsolov and B. Marinova, "Optimal Power Factor for the Reactive Load of Small Hydro Power Plants," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2755–2757, Apr. 2018, <https://doi.org/10.48084/etasr.1909>.
- [2] A. Tsolov, "Precise Generators Synchronization a Small HPP with an Excitation System," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2839–2846, Apr. 2018, <https://doi.org/10.48084/etasr.1978>.
- [3] S. Filipova-Petrakieva and V. Dochev, "Short-Term Forecasts of the Electrical Energy Consumption in Republic of Bulgaria," in *2021 13th Electrical Engineering Faculty Conference (BulEF)*, Varna, Bulgaria, Sep. 2021, pp. 1–6, <https://doi.org/10.1109/BulEF53491.2021.9690782>.
- [4] W. C. Hong and G. F. Fan, "Hybrid Empirical Mode Decomposition with Support Vector Regression Model for Short Term Load Forecasting," *Energies*, vol. 12, no. 6, Jan. 2019, Art. no. 1093, <https://doi.org/10.3390/en12061093>.
- [5] X. Shao, C. Pu, Y. Zhang, and C. S. Kim, "Domain Fusion CNN-LSTM for Short-Term Power Consumption Forecasting," *IEEE Access*, vol. 8,

- pp. 188352–188362, 2020, <https://doi.org/10.1109/ACCESS.2020.3031958>.
- [6] R. Jin, Y. Lu, Y. Wang, and J. Song, "The Short-Term Power Consumption Forecasting Based on the Portrait of Substation Areas," in *2020 IEEE International Conference on Knowledge Graph (ICKG)*, Nanjing, China, Dec. 2020, pp. 649–653, <https://doi.org/10.1109/ICKG50248.2020.00097>.
- [7] K. Yan, X. Wang, Y. Du, N. Jin, H. Huang, and H. Zhou, "Multi-Step Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy," *Energies*, vol. 11, no. 11, Nov. 2018, Art. no. 3089, <https://doi.org/10.3390/en11113089>.
- [8] A. Agga, A. Abbou, M. Labbadi, and Y. El Houm, "Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models," *Renewable Energy*, vol. 177, pp. 101–112, Aug. 2021, <https://doi.org/10.1016/j.renene.2021.05.095>.
- [9] X. Shao and C. S. Kim, "Multi-Step Short-Term Power Consumption Forecasting Using Multi-Channel LSTM With Time Location Considering Customer Behavior," *IEEE Access*, vol. 8, pp. 125263–125273, 2020, <https://doi.org/10.1109/ACCESS.2020.3007163>.
- [10] M. Khoobiyan, A. Pooya, A. Tavakkoli, and F. Rahimnia, "Taxonomy of Manufacturing Flexibility at Manufacturing Companies Using Imperialist Competitive Algorithms, Support Vector Machines and Hierarchical Cluster Analysis," *Engineering, Technology & Applied Science Research*, vol. 7, no. 2, pp. 1559–1566, Apr. 2017, <https://doi.org/10.48084/etasr.1022>.
- [11] R. Baviera and M. Azzone, "Neural Network Middle-Term Probabilistic Forecasting of Daily Power Consumption," *Journal of Energy Markets*, vol. 14, no. 1, May 2021.
- [12] L. Davlea and B. Teodorescu, "A neuro-fuzzy algorithm for middle-term load forecasting," in *2016 International Conference and Exposition on Electrical and Power Engineering (EPE)*, Iasi, Romania, Jul. 2016, <https://doi.org/10.1109/ICEPE.2016.7781292>.
- [13] W. Yichun, C. Zhenying, and L. Miao, "Med-long term system structure forecasting of power consumption based on grey derived model," in *Proceedings of 2013 IEEE International Conference on Grey Systems and Intelligent Services (GSIS)*, Macao, China, Aug. 2013, pp. 142–146, <https://doi.org/10.1109/GSIS.2013.6714759>.
- [14] N. D. Senchilo and D. A. Ustinov, "Method for Determining the Optimal Capacity of Energy Storage Systems with a Long-Term Forecast of Power Consumption," *Energies*, vol. 14, no. 21, Jan. 2021, Art. no. 7098, <https://doi.org/10.3390/en14217098>.
- [15] Faxuan Ma, "Exploration and discussion on electricity consumption demand forecasting theories and methods," in *Advances in Energy and Environment Research: Proceedings of the International Conference on Advances in Energy and Environment Research (ICAER2016)*, Guangzhou, China, Aug. 2016, pp. 7–16.
- [16] K.-K. Savov, K. Hadzhyska, D. Stoilov, T. Babinkov, and N. Nikolov, "Models for energy systems development planning," in *2020 12th Electrical Engineering Faculty Conference (BulEF)*, Varna, Bulgaria, Sep. 2020, pp. 1–4, <https://doi.org/10.1109/BulEF51036.2020.9326044>.
- [17] D. Stoilov and K. Ianev, "Generation Planning in the Bulgarian Power System under Current Market Restructuring – Method and Results," presented at the International Conference on Power Generation, Transmission, Distribution and Energy Conversion, Athens, Greece, Aug. 2002.
- [18] S. Stapczynski, "Europe's Energy Crisis Is Coming for the Rest of the World, Too," *Bloomberg.com*, Sep. 27, 2021.
- [19] G. Papageorgiou, A. Efsthadiades, M. Poullou, and A. N. Ness, "Managing household electricity consumption: a correlational, regression analysis," *International Journal of Sustainable Energy*, vol. 39, no. 5, pp. 486–496, Feb. 2020, <https://doi.org/10.1080/14786451.2020.1718675>.
- [20] V. Bianco, O. Manca, and S. Nardini, "Linear Regression Models to Forecast Electricity Consumption in Italy," *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 8, no. 1, pp. 86–93, Jan. 2013, <https://doi.org/10.1080/15567240903289549>.
- [21] I. Kostakis, "Socio-demographic determinants of household electricity consumption: evidence from Greece using quantile regression analysis," *Current Research in Environmental Sustainability*, vol. 1, pp. 23–30, Jan. 2020, <https://doi.org/10.1016/j.crsust.2020.04.001>.
- [22] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand," *Applied Energy*, vol. 238, pp. 1312–1326, Nov. 2019, <https://doi.org/10.1016/j.apenergy.2019.01.113>.
- [23] S. Amasaki and C. Lokan, "Evaluation of Moving Window Policies with CART," in *2016 7th International Workshop on Empirical Software Engineering in Practice (IWESEP)*, Osaka, Japan, Mar. 2016, pp. 24–29, <https://doi.org/10.1109/IWESEP.2016.10>.
- [24] M. H. Alsharif, M. K. Younes, and J. Kim, "Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea," *Symmetry*, vol. 11, no. 2, Feb. 2019, Art. no. 240, <https://doi.org/10.3390/sym11020240>.
- [25] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 431–438, Feb. 2015, <https://doi.org/10.1016/j.ijepes.2014.12.036>.
- [26] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761–1768, Jun. 2007, <https://doi.org/10.1016/j.energy.2006.11.010>.
- [27] Esmat Erguner Ozkoc, "Clustering of Time-Series Data," in *Data Mining: Methods, Applications and Systems*, London, UK: InTechOpen, 2021, pp. 87–105.
- [28] V. Bondarenko, S. Filipova-Petrakieva, I. Taralova, and D. Andreev, "Forecasting time series for power consumption data in different buildings using the fractional Brownian motion," *International Journal of Circuits, Systems and Signal Processing*, vol. 12, pp. 646–652, 2018.
- [29] D. J. Bora and D. A. K. Gupta, "Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 2, pp. 2501–2506, 2014.
- [30] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, Jul. 2002, <https://doi.org/10.1109/TPAMI.2002.1017616>.
- [31] P. N. Smyrlis, D. C. Tsouros, and M. G. Tsipouras, "Constrained K-Means Classification," *Engineering, Technology & Applied Science Research*, vol. 8, no. 4, pp. 3203–3208, Aug. 2018, <https://doi.org/10.48084/etasr.2149>.
- [32] X. Huang, Y. Ye, L. Xiong, R. Y. K. Lau, N. Jiang, and S. Wang, "Time series k-means: A new k-means type smooth subspace clustering for time series data," *Information Sciences*, vol. 367–368, pp. 1–13, Aug. 2016, <https://doi.org/10.1016/j.ins.2016.05.040>.
- [33] J. Béjar Alonso, "K-means vs Mini Batch K-means: a comparison," External Research Report, May 2013. Accessed: Feb. 24, 2022. [Online]. Available: <https://upcommons.upc.edu/handle/2117/23414>.
- [34] Ryan P. Adams, "Hierarchical Clustering," Princeton University.
- [35] F. Murtagh and P. Legendre, "Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm," *Journal of Classification*, vol. 31, no. 3, pp. 274–295, Oct. 2014, <https://doi.org/10.1007/s00357-014-9161-z>.
- [36] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: ordering points to identify the clustering structure," *ACM SIGMOD Record*, vol. 28, no. 2, pp. 49–60, Mar. 1999, <https://doi.org/10.1145/304181.304187>.
- [37] M. Shukla, Y. P. Kosta, and M. Jayswal, "A Modified Approach of OPTICS Algorithm for Data Streams," *Engineering, Technology & Applied Science Research*, vol. 7, no. 2, pp. 1478–1481, Apr. 2017, <https://doi.org/10.48084/etasr.963>.
- [38] "ESO.BG - Електроенергиен Системен Оператор." <http://www.eso.bg/?did=124> (accessed Feb. 24, 2022).
- [39] W. A. Fuller, *Introduction to Statistical Time Series*. New York, NY, USA: John Wiley & Sons, Ltd, 1996.

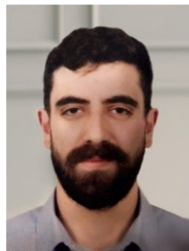
- [40] Statsmodels library API Reference: statsmodels.tsa.stattools <https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.adfuller.html>
- [41] SciKit-Learn API Reference: Silhouette, *scikit-learn*. https://scikit-learn/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html (accessed Feb. 24, 2022).

AUTHORS PROFILE



Simona Filipova-Petrakieva was born in 1971 in Sofia, Bulgaria. She finished the Secondary School of Mathematics in Sofia, Bulgaria (1989). She received B.S. and M.S. in Electronics and Automatics from the Technical University of Sofia, Bulgaria (1994), and acquired her Ph.D. degree in Electrical Engineering, Electronics and Automatics, Theory of Electrical Engineering in 2005.

Since 2009 she teaches as an Associate Professor in the Department of Theory of Electrical Engineering, Faculty of Automation, TU - Sofia, Bulgaria. Her research interests include Theory of Electrical Engineering, Cluster Analysis, Graph Theory, Interval Methods for Analysis and Synthesis of Linear Circuits and Systems, Discrete Event Systems, Discrete Structures in Mathematics, Electrical Power Consumption, Electric and Hybrid Vehicles. Her teaching activities include courses of the Theory of Electrical Engineering and Discrete Structures in Mathematics. She is a regular member of IEEE and in December 2021 she was invited to become a senior member.



Ventsislav Dochev was born in Pleven, Bulgaria, (1999). He finished the Vasil Levski High School in Troyan, Bulgaria (2017). He received a B.S. in Computer Science from the Technical University of Sofia, Bulgaria, (2021). He is currently studying for a M.S. degree in Information Retrieval and Knowledge Discovery at Kliment Ohridski Sofia University. Since 2018 he is working as a Quality Assurance Team Leader in a company developing camera software. His

research interests include Artificial Intelligence, Cluster Analysis, Knowledge Extraction, Applied Mathematics and Statistics.