Proceedings of the 26th International Conference on Automation & Computing, University of Portsmouth, Portsmouth, UK, 2-4 September 2021

A Comparative Study of Methods to Forecast Domestic Energy Consumption Aggregated with Photovoltaic and Heat Pumps System

Arqam Rafi, Timothy Lee, Wenyan Wu
School of Engineering and the Built Environment
Birmingham City University
Birmingham, United Kingdom
arqam.rafi@mail.bcu.ac.uk

Abstract— **Rise in the usage of photovoltaic (PV) system at the residential sector brought challenges for the distribution network operator (DNO) and with a high demand of the Heat Pump (HP) system to fulfil the target of low carbon emission potentially brought far greater trials to predict energy at the domestic network. Prediction is very crucial for electrical distribution companies since their business largely relays on how to make the most out of their energy generation without making it go to waste. This study compares different methods (from machine learning to deep learning) to forecast domestic energy consumption aggregated with HP and PV system. The prediction tool proudly uses large residential energy measured data at a minute frequency for a year combines synthetically with the real measured data of HP and PV system. The forecasting methods is different for various data type, this study allows to compare which one would be more efficient in which type of data set and which one to predicts the finest.**

Keywords- Residential Energy; Predictive Analytics; Machine Learning; Large Data; Low carbon technologies; Heat pumps; Photovoltaic system

I. INTRODUCTION

With the continued development of new technologies in order to reduce carbon emissions and energy dissipation, people have worked hard to achieve technologies like solar PV system, heat pump and electric vehicles. They are increasingly replacing the conventional energy consumption and generation technologies. Despite their high initial cost of installation, they are still considered to be the preferred technologies people would consider in their houses especially when the government is giving different incentives [1].

But as the new technologies are being fitted more and more into the residential sector [2], this already unpredictable sector will be more capricious for power distribution companies. The importance of forecasting energy consumption can be evaluated by looking at the number of project proposals submitted at the Ofgem Electricity Network Innovation Competition (NIC), which is an annual opportunity for electricity network companies to compete for funding by showing their most innovative projects proposals [3].

The residential network has seen a tremendous rise in the usage of photovoltaic systems over recent years. This is due to a number of reasons especially the incentive by the government introduced Feed-in Tariff (FiT) scheme for financial incentive mechanism in April 2010 which was then closed in 2019 April, this is because of the installed capacity went up from 23 GWh of the original 2010 deployment to 1420GWh in 2017 [4]. This demonstrates the UK is on a fast track to reach the target to reduce carbon emission at the domestic sector. With the large number of PVs generating energy during the day is causing great concerns for DNO to accurately predict domestic energy and supply only what's required [5].

The promotion to use heat pumps (HPs) is also introduced by a Domestic Renewable Heat Incentive (Domestic RHI) scheme. This would help to reduce carbon emissions and meet renewable energy targets stated in [6]. Moving from fossil fuel heating to heat pumps will make a high amount of reduction in greenhouse gas emission, although the high level of HPs penetration poses a new challenge for the low-voltage networks. Since HPs typically use induction motors the effect could be on power quality which includes voltage variation (dips and surges), harmonics and frequency variation [7].

Despite different methods used by distribution networks to calculate the type of consumption or generation by HP and PV respectively would take place at a housing unit, they are still largely dependent on historical, weather and active occupancy data. A large historical data helps to understand what type of consumption will take place at what time of the year or month or a day or in a minute. With the short-term forecasting, many other factors such as the weather and the active occupancy also matter. Active occupancy could be achieved from an updated data which shows how people spend their time such as the one in the Time Use Survey 2000 [8].

The methods used in this paper are mainly the most common methods used for forecasting [9]–[12] but the historical data selection for residential energy consumption, PVs, HPs, and its combination makes this paper unique. The data for residential energy consumption has taken from 2008-2009 years, making sure no PV or HP were present. Whereas the dataset for PV and HP has been taken from 2013-2014 housing unit. All three datasets are from different locations but have similar features such as the type of housing, age of property, and number of bedrooms etc. and it has been assumed that the weather, irradiance, and the active occupancy remained the same. This paper contributes toward the area of predicting energy with the uptake of HP and PV at the residential sector.

II. Residential Energy

Residential energy is the energy which is consumed by a housing unit, this includes all the electrical consumption by various consumer products. Utilizing historical residential electricity energy use data to predict the energy consumption has always been one of the key elements [13]. It depends a lot on the type of energy consumption that needs prediction but usually the shortest frequency data with at least a yearlong trend is considered to be most critical in forecasting [14]. Short frequency data at a minute resolution helps the DNO in making necessary arrangements in order to deliver the energy efficiently, improve decision-making and planning activities[15]. Many prediction problems involves in some time-based sections which makes time series forecasting important. The beauty of historical data patterns which is the base of the projection into the future is broken down into many different parts such as the season, trend and the cycles[11], [16], [17].

A. Data source preprocessing Analysis

Historical data used is from a part of CREST Energy Demand Model [14], it contains measured electricity used at a one minute resolution from 22 dwellings in the East Midlands, UK, and it is for two complete years 2008 and 2009. Although the more data used can often be advantageous as it allows better chance to detect patterns but due to the size of the data, only one dwellings reading used which offered continues reading with a very small missing valuesfor a slightly more than a year at a frequency of a minute. This selection is mainly based on the type of the housing and the continuity of the dataset.

B. Data Processing

The quality of the data is an importance factor in forecasting process therefore some level of data cleansing is always required. After carefully analyzing the dataset, the findings show some level of data which was missing in the data set and in order to fix this some adjustment had to be made. The missing values in the data could be filled or eliminated if not making any significant difference, since the prediction is at a minute frequency every bit of data was important, hence firstly the dataset missing values were filled with using Random values in Min Max range illustrated in Fig.1 that shows the complete year. Due to the results not being random, a deep learning artificial neural network method was used which is considered to be good with time series as used by [18]. The dataset were filled with Long Short Term Memory (LSTM), and only considered the data set from "21-August-2008" to "19-November-2009" due to its reputational trend. The result of this dataset is presented in Fig.2 where the LSTM predicted values shown in orange and the blue is the actual measured data. LSTM did some prediction showing the energy consumption in negative which was later removed.

III. Photovoltaic Energy

PV energy is one of the most widely used low-carbon technologies in the UK[19]. The importance of aggregating this LCT is because of its type of energy generation behavior. PV system typically generates energy under sunlight and its generation varies largely with how the weather is, especially the clouds and the irradiance levels. Consumption of residence energy during the day is typically low due to more people being out either for work, study or other social activities[8]. This create a case where a dwelling is producing more energy than it is using, hence the surplus goes back to the grid. It is important to analyze and forecast this energy supply as it can greatly benefit not only the customers but also the power generation companies only if it is managed properly.

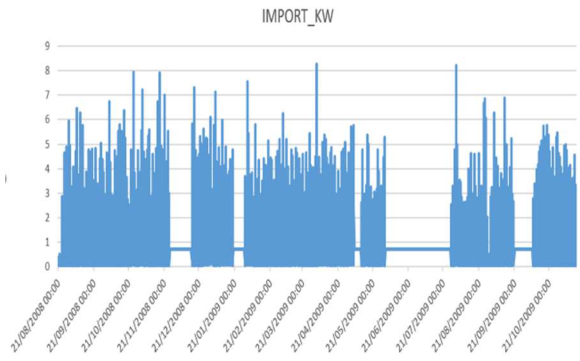

Figure 1. Repaired dataset using random values

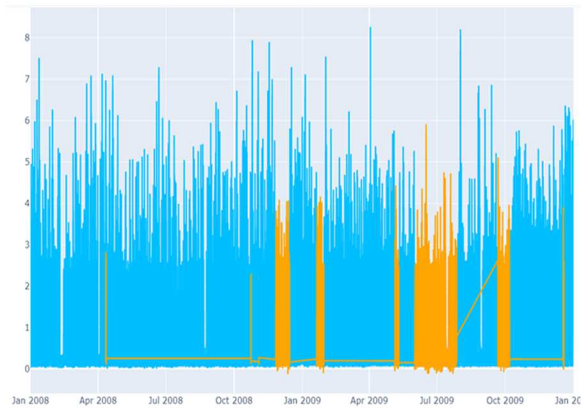

Figure 2. Repaired dataset using LSTM

A. Data source of PV.

In 2013, UK Power Network Company deployed monitoring equipment at various locations on their network to measure real-life data, 6 of those were placed directly at residential customers with PV installations [20]. These measurements captured every generation at an hourly frequency for a year from 2013 to 2014 out of which three months of high-resolution (one-minute) measurements was also taken over summer 2014.

B. Data Preprocessing

After analyzing the data, and keeping into account to synthetically aggregate it with the residential energy data set. The data was taken from "21-August-2013" and ends at "19-November-2014". The data frequency was hourly, and there were missing values in the data set. In order to aggregate the two datasets from different years a conversion of year 2013 to 2008 and 2014 to 2009 was performed and assumption is being established that the weather and the irradiance is same for those years. Also conversion of the frequency from hourly to minutely be performed using upsampling as assumed that the weather and irradiance remained constant for an hour. Missing values were repaired or filled with the help of LSTM. The up sampled and the repaired dataset can be illustrated in the Fig 3 and Fig 4 respectively.

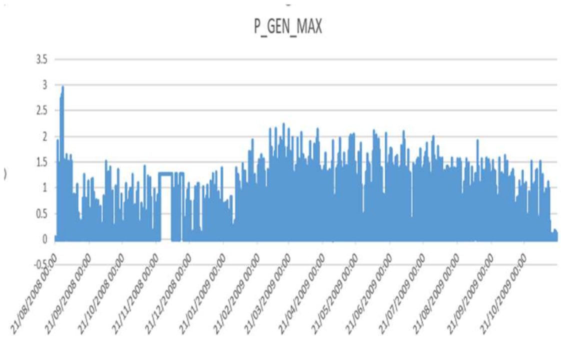

Figure 3. Up sampled dataset of photovoltaic system

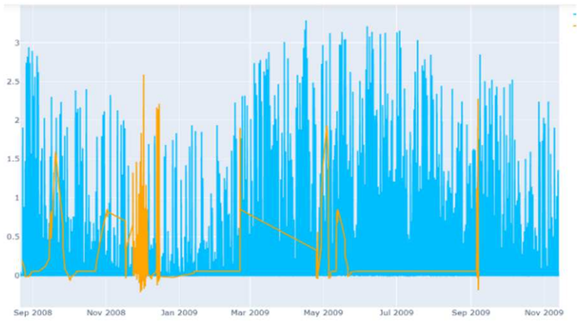

Figure 4. Repaired dataset of photovoltaic system using LSTM

IV. Heat Pump Energy

Heat Pumps are not as common as some low-carbon technologies but with an introduction to Air Source Heat Pump (ASHP) which has less installation cost compared to Ground Source Heat Pump (GSHP), the chances of having more HP at the residential sector will be common in the future. With its limitation to be more applicable for houses rather than apartments due to a number of constrains such as noise of its operation, piping installation, installation permission, and heat exchange which is mainly air in this case. There are some other factor such as the property type, age of the property, number of bedrooms, emitter type (Radiator, floor heating, or both), and if the HP will be supplying hot water also make a deep impact on its efficiency which is proportionally making an impact on consumption of electricity[19].

A. Heat Pump Data Analysis

The Department of Business, Energy and Industrial Strategy (BEIS) previously known as Department of Energy and Climate Change (DECC), funded Buildings Research Establishment (BRE) to monitor the performance of HPs at the residential sector. This detailed monitoring campaign monitored 700 HPs at different residential locations, with 2 minute electricity data collected from 31st October 2013 to 31st March 2015 [21]. The data was large and rich with other information including, the type of HP, type of property, age of property, number of bedrooms, type of heat emitters and the power rating of the HP.

B. Data Processing

From this highly informative dataset, the first stage was to see what type of HP should be considered for analysis, the dataset showed that out of total 700 HP only 171 were GSHP and 525 were ASHP and 6 were missing. Hence for analysis ASHP were only considered for forecasting. The type of property selected was a detached house, with more than 4+ rooms and made after 2000 to see its maximum impact on the electrical grid.

The selected HP dataset also contained electricity consumption from not only the HP unit alone but also from domestic hot water, circulation pump and the whole system boost hence all of the consumptions were sum and converted in kW from kWh. Second stage was to up sample the data from 2 minute frequency to 1 minute frequency with the repetition of the same value, and change the date-year from 2012 to 2008 and 2013 to 2009, when changing the year, the starting year was 2008 from (21/08/2012 13:00:00 to 31/05/2013 23:58) where 2013 is 2009 now. Anything above 21/08/2008 13:00:00 was changed to be 2009 NOT 2008 and was eventually came under 31/05/2013 23:58 as 31/05/2009 23:58. There was also no missing values in the dataset however the total period was slightly lower than the other dataset hence a repetition of the trend technique was used to make it ready for aggregation can be illustrated in the Fig 5.

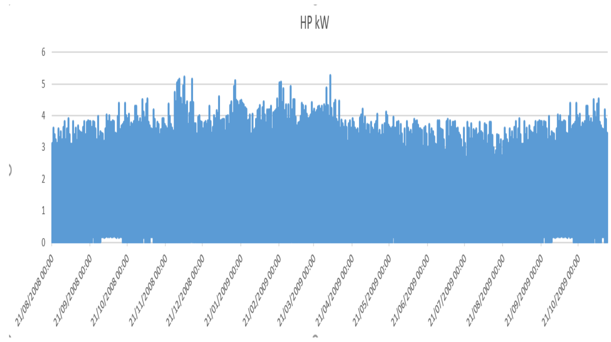

Figure 5. Processed dataset of HP system at two minutes

V. Aggregated Energy

TABLE I. Conversion and Aggregation of Energy

	Residential Energy	**PV Energy**	**HP Energy**
Frequency	One minutes	Hourly to a minute	2 minutes to a minute

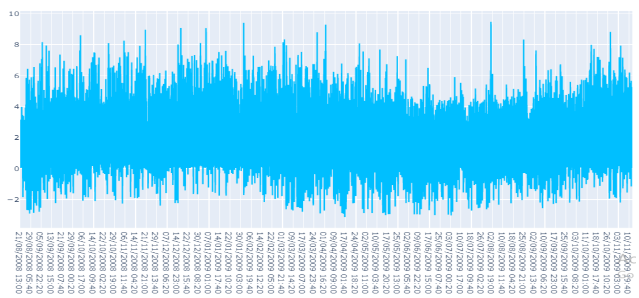

Figure 6. Aggregated data set of all the energy

The datasets from three different sources were prepared in the last step and now ready to be merged. The data is for the same period in all dataset from "21-August-2008" and ends at "19-November-2009", the data frequency is in minutes, and there is no missing values in the dataset. This may be seen in the table 1.

Since PV only generates, its value should be subtracted from the power consumed by the consumer, whereas the HP only consumes energy hence this will be added into the main as illustrated in the Fig 6.

VI. Forecasting models comparison

There are various models used to make forecast starting from the most common using statistical models[9] that typically use basic tools and skills. A far greater advancement been made in machine learning and deep learning allowed forecaster to use artificial intelligence-based models such as the one shown below, this permits greater accuracy. However, the forecasting technique do not rely on the type of the models used rather than it is more dependent on needs to be achieve. It is easier to accurately predict short term values with far more confidence but to make a long term prediction, the same model might fail hence a very different technique needed to be used. Everything matters starting from how much data is available, the size of the data with respect to its frequency, how the resampling of the data is made, and all the previous requirements from missing values to errors in the dataset.
After analysis of the type of models available which would perform better on a time series, a selection was made to consider some basic statistical models and some machine learning to deep learning models. This includes, moving average, Autoregressive Integration Moving Average (Arima), Seasonal Autoregressive Integration Moving Average (Sarima), Facebook prophets and LSTM. Following are the detailed forecasted results on the dataset.

A. Moving Average

Generally speaking, moving average is said to be a technique that gives idea of overall trends through given data sets[9]. It takes average of the data sets. Moving average is related to time series as it considers time to make predictions. To calculate moving average a new time series is constructed with values which are average of raw observations in the original series. Predictions are made by not looking at average of time rather interval of time "move" where the change occurred during considered time period is considered. It is advantageous in energy forecasting because it is reliable method for commodities which have constant demand with slight seasonality. It is simple method and opts out random variants[22]. Also, it gives constant forecast where plotting more than one trend is also possible. Disadvantages of methods are that it overlooks seasonal variations and upward demand trends. It also sometimes ignores complex relationships mentioned and requires maintaining complex time period history. For stock predictions and forecasting moving average is preferred to be used. The graphically forecasted values illustrated in Fig 7 while training 70% of the dataset and 30% for testing.

B. Autoregressive Integration Moving Average (ARIMA)

As the name indicates ARIMA is the combination of auto regression and moving average, where auto regression covers the varying variables that regress on their own prior values, while integration part focuses on differencing of raw observations to make time series stationary, similarly moving average here mentions the dependency of observation and error from average calculated to lagged observations. ARIMA makes the time series stationary, by differencing the data collected this makes predictions to be deduced easily[23]. ARIMA after making time series stationary evaluates the strength of dependent variable to the changing or varying variables. This makes the seasonality to be removed and predictions can be made. ARIMA covers the seasonal fluctuations not covered by MA alone by making time series stationary[24]. However, it can only found good fit model if time series are small, uncorrelated and randomly distributed otherwise seasonality may appear. For repeating cycles, it doesn't cover seasonality in data. This model is preferred by the users in sales forecasting, stock predictions and similar predictions. Due to the size of the data, the data was resampled to hourly, 70% was used to train the data, result illustrated in Fig 8.

C. Seasonal Autoregressive Integration Moving Average (SARIMA)

It is an extension of ARIMA model, where it was designed cover the drawbacks of the ARIMA model. So it is also known as seasonal ARIMA. As the name indicated it deals with making time series stationary then studying it for a single changing variable with a seasonal component. Thus SARIMA demands studying the hyper parameters for both trend elements which are same for ARIMA model including trend auto regression, integration and moving average and seasonal elements which do not exist for ARIMA model including seasonal auto regression, seasonal integration, seasonal moving average and number of time phases for single seasonal period[11]. Thus advantage of using SARIMA model is that it not only covers difference, auto regression, moving average as ARIMA but also covers the seasonal components of time series[23]. SARIMA may not be applicable for long term predictions as seasonality may reappear so to avoid that usually single variant and relatively shorter time duration is preferred. It is usually preferred for electricity load forecasting as it involves seasonal fluctuations[23]. Fig 9 presents the forecasting graph, the dataset is down sampled due to its size and same 30% is used to test the data.

D. Facebook Prophets

Prophet is forecasting method which is mainly restricted to python and R. It is termed so because it is released by Facebook data science team. Prophet is fully automated methods of forecasting; however it can be turned manual if needed by analysts. The relation of prophet with the time series is that it works well when strong seasonal effects are involved that too with involvement of several seasonal historical data. It gives an additional model with non-linear trends involving every kind of seasonality from daily, weekly to yearly basis even holiday effect is also considered[12]. The advantages of using prophet is that it is automated and fast, it's reliable and accurate[25]. Also, it can be implemented for forecast in whatsoever language that suits a person as in R and python. Forecasting can be adjustable as one can choose parameters as needed. The accuracy however still is a problem in this case as no holt-winters and box-jerkins models available for prophet[25]. As it is quite new more research needs to be carried out carried out. It has been used to forecast sales and purchase rates[26]. Similar approach used forecast the time series using Facebook prophets, despite is unsuitability for such dataset it still predicted linear result displayed in Fig 10.

E. Long Short-Term Memory (LSTM)

LSTM or Long Short-Term Memory network is a type of recurrent neural network used in sequence prediction problems for learning order dependence. Complex problems such as speech recognition and machine translation make use of this behavior since it lies in a complex region of deep learning. The primary advantage that LSTM provide over other methods are the exploding gradients and the vanishing gradients (both of which are associated with the way the network is trained)[27]. During the training, the aim is to minimize loss which is seen in the outputs when data for training is sent to it. The gradient is calculated, the weights are adjusted accordingly, and the process is iterated until an optimal set of weights is observed for which the loss is minimum. Since the network is suited for learning long sequences of observations, this network type is a perfect candidate for time series forecasting[25]. Although LSTM can solve the problem of vanishing gradients, it has been observed that they do not succeed at removing it completely. In addition to this, they are inefficient in terms of resources required since they need a lot of power to get trained before they can predict real-world problems. The graph shown in Fig 11, illustrates how close LSTM predicts compares to other models used in this paper.

VII. RESULTS AND DISCUSSION

The results shows LSTM predicts energy consumption, while sharply following the seasonal trend line without missing many peaks and valleys. Whereas, Facebook prophet results shows it did follow the seasonality but completely ignore any peaks and valleys which are very important in energy forecasting. ARIMA and SARIMA results take a good care of trends, cycles, errors and non-stationary aspects of a data set when making forecasts, especially the seasonality is very much improved in SARIMA as it use a linear combination of seasonal past values. Moving average forecasted graphs shows how well it predicted values under zero using basic averaging approach.

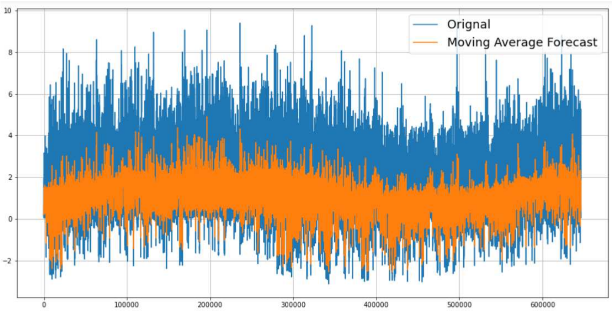

Figure 7. Moving Average Forecast

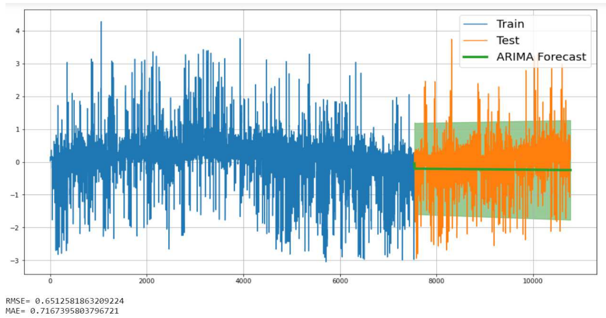

Figure 8. Forecasting using ARIMA

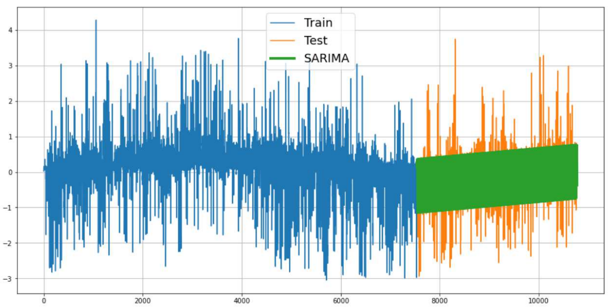

Figure 9. Forecasting using SARIMA

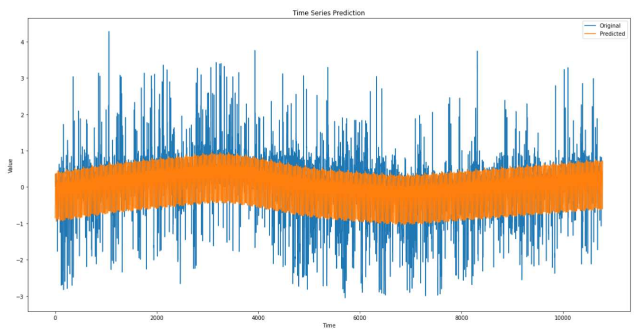

Figure 10. Forecasting using Prophets

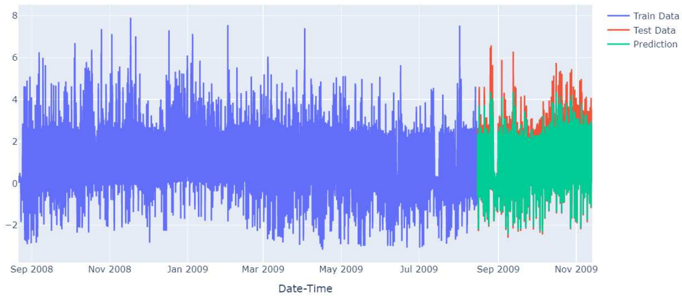

Figure 11. Forecasting using LSTM

VIII. CONCLUSION

Forecasting of energy consumption is an important element for DNOs, the accuracy varies from model to model but depends more on how the model is being used for a particular dataset. Most importantly how rich the dataset is and how well it is processed in order to extract the desired information and used that information in forecasting is the ultimate key. The other factor which influences forecasting models would be the parameters used, the training and testing set for validation and lastly the period covered.

From the result taken it maybe concluded that LSTM method maybe used to forecast the electrical energy consumption despite aggregating the residential energy with either the PV and/or HP. In future, the forecasted values will be used on a simulated electrical distribution network to model the impact of PV and HP.

REFERENCES

[1] A. Rafi, T. Lee, and W. Wu, "Impact of low-carbon technologies on the power distribution network," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 329, no. 1, 2019.

[2] M. Akmal, B. Fox, D. J. Morrow, and T. Littler, "Impact of high penetration of heat pumps on low voltage distribution networks," *PowerTech, 2011 IEEE Trondheim*, pp. 1–7, 2011.

[3] L. Hutton, "Decision on the 2020 Gas and Electricity Network Innovation Competition," 2020.

[4] BEIS, "Digest of UK Energy Statistics (DUKES) 2018 Chapter 5: Electricity," 2018.

[5] E. W. Paper, S. Küfeoğlu, M. Pollitt, and M. Pollitt, "The impact of PVs and EVs on Domestic Electricity Network Charges : a case study from Great Britain," no. May 2018, 1830.

[6] "the Renewable Heat Incentive : a Reformed Scheme the Renewable Heat Incentive : a Reformed," no. December, 2016.

[7] B. Fox, J. D. Morrow, M. Akmal, and T. Littler, "Impact of heat pump load on distribution networks," *IET Gener. Transm. Distrib.*, vol. 8, no. 12, pp. 2065–2073, 2014.

[8] J. Gershuny and O. Sullivan, "United Kingdom Time Use Survey, 2014-2015," *UK Data Serv.*, no. 8128, p. SN: 8128, 2017.

[9] F. Zheng and S. Zhong, "Time series forecasting using an ensemble model incorporating ARIMA and ANN based on combined objectives," *2011 2nd Int. Conf. Artif. Intell. Manag. Sci. Electron. Commer*, pp. 2671–2674, 2011.

[10] T. Panapongpakorn and D. Banjerdpongchai, "Short-Term Load Forecast for Energy Management Systems Using Time Series Analysis and Neural Network Method with Average True Range," *2019,* pp. 86–89, 2019.

[11] S. I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou, and A. G. Bakirtzis, "Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting," *2016 IEEE Int. Energy Conf. ENERGYCON 2016*, 2016.

[12] L. G. Rocha, E. D. Goias, S. G. S. Alcala, and L. P. Garces Negrete, "Short-term electric load forecasting using neural networks: A comparative study," *2020 IEEE PES Transm. Distrib. Conf. Exhib. - Lat. Am. T D LA 2020*, 2020.

[13] M. Y. Zhai, "A new method for short-term load forecasting based on fractal interpretation and wavelet analysis," *Int. J. Electr. Power Energy Syst.*, vol. 69, pp. 241–245, 2015.

[14] I. Richardson and M. Thomson, "Intagrated simulation of Photovoltaic Micro-Generation and Domestic Electricity Demand: A one minute resolution open source model," *CREST - Dep. Electron. Electr. Eng. Loughbrgh. Univ. UK.*, 2010.

[15] L. F. Ochoa and P. Mancarella, "Low-carbon LV networks: Challenges for planning and operation," *IEEE Power Energy Soc. Gen. Meet.*, pp. 1–2, 2012.

[16] Y. H. Hsiao, "Household Electricity Demand Forecast Based on Context Information and User Daily Schedule Analysis From Meter Data," *IEEE Trans. Ind. Informatics*, vol. 11, no. 1, pp. 33–43, 2015.

[17] J. W. Taylor, L. M. de Menezes, and P. E. McSharry, "A comparison of univariate methods for forecasting electricity demand up to a day ahead," *Int. J. Forecast.*, vol. 22, no. 1, pp. 1–16, 2006, doi: 10.1016/j.ijforecast.2005.06.006.

[18] F. Succetti, A. Rosato, R. Araneo, and M. Panella, "Multidimensional Feeding of LSTM Networks for Multivariate Prediction of Energy Time Series," *Proc. - 2020 IEEE Int. Conf. Environ. Electr. Eng. 2020 IEEE Ind. Commer. Power Syst. Eur. EEEIC / I CPS Eur. 2020*, 2020.

[19] DECC, "UK Renewable Energy Roadmap," *Carbon N. Y.*, vol. 5, no. July, pp. 293–298, 2011, doi: 10.1021/es00108a605.

[20] UKPN, "Validation of photovoltaic connection assessment tool - Closedown report," no. March, 2015.

[21] R. Lowe *et al.*, "Analysis of Data From Heat Pumps Installed Via the Renewable Heat Premium Payment (Rhpp) Scheme," no. 8151, pp. 2013–2015, 2017.

[22] S. Hansun, "A new approach of moving average method in time series analysis," *2013 Int. Conf. New Media Stud. CoNMedia 2013*, pp. 0–3, 2013, doi: 10.1109/conmedia.2013.6708545.

[23] K. Goswami and A. B. Kandali, "Electricity Demand Prediction using Data Driven Forecasting Scheme: ARIMA and SARIMA for Real-Time Load Data of Assam," *2020*, pp. 570–574, 2020.

[24] S. Mehrmolaei and M. R. Keyvanpour, "Time series forecasting using improved ARIMA," *2016 Artif. Intell. Robot. IRANOPEN 2016*, pp. 92–97, 2016.

[25] W. X. Fang, P. C. Lan, W. R. Lin, H. C. Chang, H. Y. Chang, and Y. H. Wang, "Combine Facebook Prophet and LSTM with BPNN Forecasting financial markets: The Morgan Taiwan Index," *Proc. - 2019 Int. Symp. Intell. Signal Process. Commun. Syst. ISPACS 2019*, pp. 0–1, 2019.

[26] V. D. M. Z. İ. Yelmen, "Drug Sales Prediction with ACF and PACF Supported ARIMA Method," *IEEE*, pp. 6–10, 2020.

[27] M. Imani, "Long short-term memory network and support vector regression for electrical load forecasting," *5th Int. Conf. Power Gener. Syst. Renew. Energy Technol. PGSRET 2019*, pp. 26–27, 2019.