Short-Term Forecasting of Uncertain Parameters for Virtual Power Plants

Mahdi Abbasi
Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
mahdi.abbasi@lis-lab.fr

Amirabbas Asadi
Bu-Ali Sina University, Computer Engineering Dept., Hamedan, Iran.
a.asadikorom@eng.basu.ac.ir

Seifeddine BenElghali
Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France,
seifeddine.benelghali@lis-lab.fr

Mohamed Zerrougui
Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France,
mohamed.zerrougui@lis-lab.fr

2021 IEEE 26th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD) | 978-1-6654-1779-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/CAMAD52502.2021.9617792

***Abstract*— Forecasting uncertain parameters are of great importance in the control scenarios of virtual power plants (VPPs). Short-term forecasting of energy consumption, air temperature, and market price is of great importance in the optimization engines in VPPs. Though with the advent of telemetering devices and climate forecasting systems, the energy consumption profiles of consumers and the climate profiles have been collected easily, and consequently the required input data for forecasting these parameters have been available easily, the accuracy of forecasting algorithms has remained an issue. Also, due to the impact of several actors on the market value, no fixed model can be used for all types of markets. This paper focuses on the effect of the time-series analysis methods on developing effective short-term forecasting models for these uncertain variables in a VPP. For this purpose, several models of energy forecasting were built and tested on the Formentera dataset. Our results show that by building the models based on the analysis results, the simple forecaster's models with lower complexity can accurately forecast the uncertain variables of VPPs.**

Keywords—Virtual Power Plant, Forecasting, Timeseries analysis, Short-term.

I. INTRODUCTION

The current decade has perceived significant evolutions in the electric power area. One of the prominent evolutions is the replacement of conventional fossil fuel power plants with renewable and clean energy resources. Significantly, with the accelerated growth in using renewable energy resources (RESs) such as wind farms and photovoltaic farms, the power systems increasingly have shifted from a centralized structure to a distributed manner [1]. In addition, the recent developments in knowledge and the initiation of smart meters have created the substrate for the concept of virtual power plants [1]. Virtual power plants are distributed generation devices such as microchips, wind power, small hydropower plants, renewable energy generators, and the like, all operated by a central or distributed control unit. The purpose of a virtual power plant is to relieve load in the network with the intelligent distribution of power generated by distributed energy resources (DERs) and RESs during peak load periods. In addition, the combined power generation and power consumption of grid units in the VPP are traded on the energy exchange according to the mechanisms provided by the VPP controller [2, 3].

Each VPP provides several services to different operators, companies, utilities, electricity supplier companies, power traders and network operators. VPP business activities include day-ahead and intraday market transactions. The VPP aggregator can contract with several small-scale DERs and gather their generation and participate in the electricity market on behalf of those. This mechanism may bring additional benefits for various stockholders. The VPP aggregator can coordinate the DERs scheduling and participate to different markets in order to obtain more profit. From DER's point of view, they can play a more active role in the power grid under the umbrella of VPP. Furthermore, the VPP is more efficient and flexible from the system operator perspective. All transaction is scheduled according to the predictions made for the amount energy production by each DER or RES, energy consumption and the market prices [4].

Short-term forecasting of uncertain variables is a vital process in the VPPs. Real-time management of VPPs is a complex problem due to diverse energy sources and the uncertainties associated with them. To do this, the VPP controller needs the value of some uncertain parameters in predictive intervals [5]. Especially, producing an accurate forecast of uncertain variables like demand, production, and market value in the next 24 hours is essential for VPPs, since it can have a direct impact on the optimal hourly planning of the generation units, as well as on purchase/sale in energy market [6].

This paper shows the importance of time series analysis selecting the most accurate model to forecast the uncertain parameters of VPPs. Also, we show that to forecast uncertain data with discoverable trends in their time-series, simple models are function more efficiently that the complex models. The paper is organized as follows section II explains the basic structure of a VPP, section III discusses the features that influence short term energy forecasting in a VPP. Section IV describes the time series analysis of the Formentera dataset. Section V reviews the forecasting models used in this study. The structure of formed times series for uncertain variables and the results of their forecast by different forecasting models are presented in section VI. Finally, section VII concludes the paper.

II. STRUCTURE OF A VIRTUAL POWER PLANT

Fig. 1 shows the fundamental structure of a VPP. In this structure, a forecasting engine is required. This module forecasts the weather, renewable generation, market price, and energy demand and delivers the forecasted series to an optimization engine. Based on these forecasts, the optimization engine carries out generation management, load

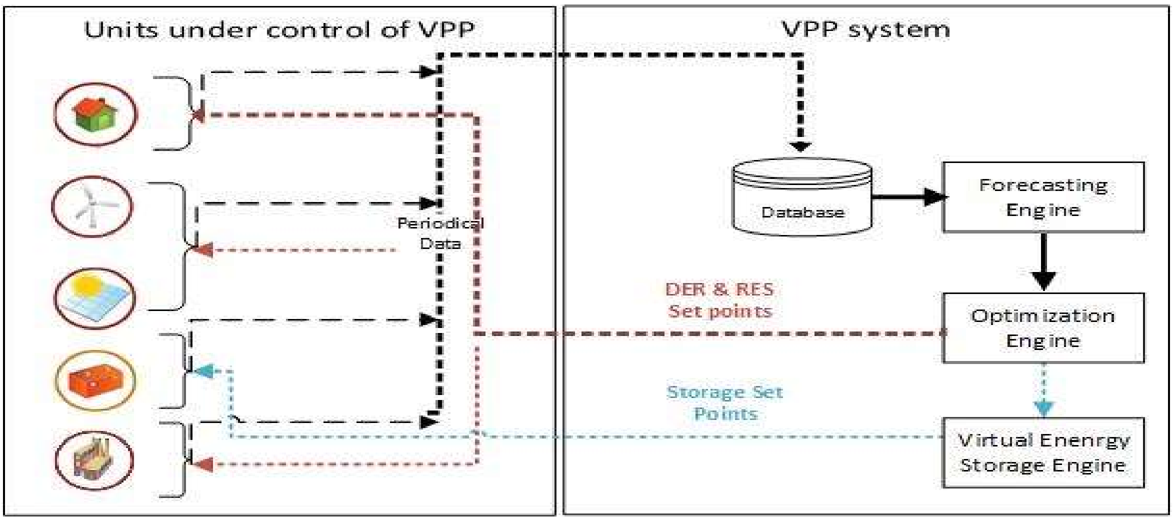

Fig. 1. Basic structure of a virtual power plant

management and scheduling according to different scenarios. The optimization engine while performing generation management has to consider the technical constraints of the DERs and RESs, the actual state of the DERs and its active and reactive power set points, and while performing load management, the number of controllable loads available with a given consumer, market prices and benefits of stakeholders, etc. The produced setpoints are dispatched to the corresponding units via communication mechanisms.

III. Parameters Impacting the Accuracy of Forecasting

Energy consumption is a dynamic process inclined by a wide range of features. For accurate prediction of energy demand, the forecasting model should take into account the variables that influence the consumption of energy. In addition to the traditional factors, energy demand in a VPP is influenced by several other factors. The following are the traditional factors that affect short-term energy forecasting [7, 8].

Weekly behaviors: The day of the week, whether it is a weekday or holiday, and hour of the day, influence energy forecasting significantly.

Seasonal Behaviors: In any season, the trend of energy consumption is likely same. For in each winter the consumption, and demand are the result of using heater instruments.

Climatic behaviors: conditions considered may include temperature, humidity, irradiation of sunlight, rainfall, cloud coverage and some special events like typhoons, etc.

Holiday behaviors: Local events like holidays may affect the energy demand. These events may lead to either an increase or decrease in demand. The influences of these events are usually local and restricted to specific geographical zone.

Also, there are some parameters that are inherently resulted by the services inside the VPP and hence must be considered in the model of energy forecasting in VPPs. These parameters include virtual energy storage restrictions and the amount of the flexible loads. In the following section, we provide the time series analysis of the Formentera dataset. Next, we show how the parameters mentioned in this section influence the trend and accuracy of the forecasting.

IV. Time Series Analysis of Formentera Dataset

The Formentera database contains 20 variables, the values of which have been measured based on the energy consumption and climate of Formentera Island for four years. For each of these variables, a value is recorded every ten minutes. The purpose of this study is to predict the short-term value of the variables Demand, Air Temp and Market Value. The variables in this dataset can be classified to the following groups:

1. Power production and consumption: The variables in this group are:
   a. Demand: shows the actual power demand in the specified period of time.
   b. Gasoline Turbine: shows the actual power produced by gasoline turbines.
   c. Diesel_generators: shows the actual power produced by diesel generators.
   d. Formentera_Ibiza_link: shows the actual power transmitted by the sea cable from Ibiza Island.
   e. PVPC: Shows the electrical energy produced by the photovoltaic farm.
2. Market:
   a. Market Value: this variable shows the market price in the periods of time.
3. Weather:
   a. Air_temp: shows the temperature in the periods of time.
   b. Azimuth: shows the compass direction from which the sunlight is coming.
   c. Dni: Shows the direct normal irradiance of the sun.
   d. Dhi: shows the diffuse horizontal irradiance o the sun.
   e. Gtifixedtilt: shows the total radiation received on a surface.

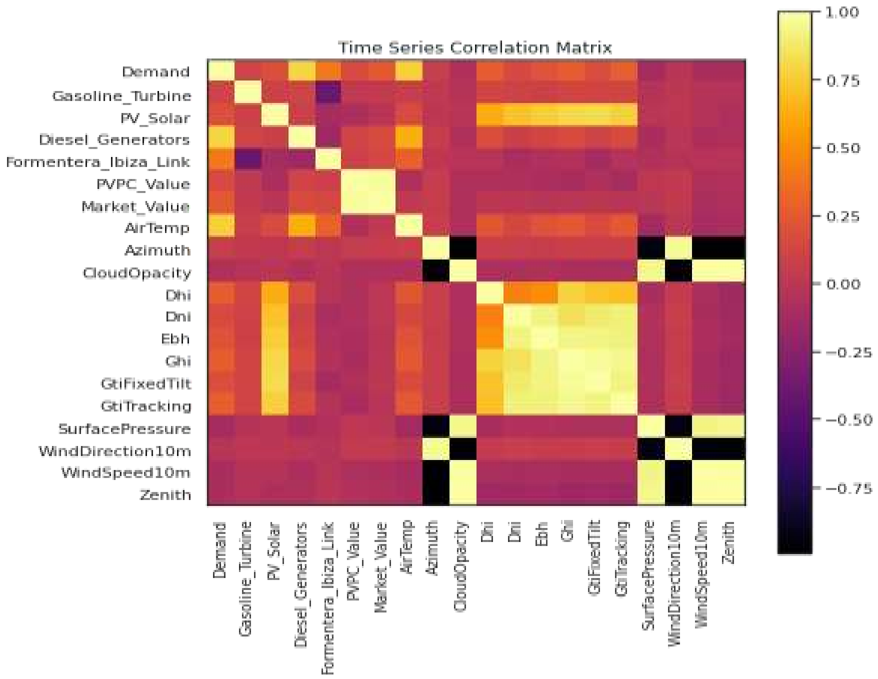

Fig. 2. Correlation of the time-series of uncertain variables in the Formentera dataset.

f. Gtitracking: shows the total tracking irradiance of a surface in front of the sunlight.

g. SurfacePressure: shows the atmospheric pressure at a location on the earth's surface.

h. WindDirection10m: shows the dominant direction of the wind in the period.

i. WindSpeed10m: Shows ten meters above the top of the tree canopy.

As explained in the introduction of the paper, many variables have a direct or indirect impact on the forecasting accuracy of the demand. Fig. 2 illustrates the value of the Pearson correlation factor for the variables of the Formentera dataset. As shown in Fig. 2, the correlation between demand and the variables of Diesel-generators and Airtemp is considerable. Also, the correlation between the demand and Ibiza-link shows that the Ibiza Sea cable has been used to compensate for the power requirements. Moreover, a correlation can be seen between the demand and market value. The plots of Figs .3 and 4 show the trends of correlation in the variables mentioned earlier in the period starting from Jan 2017 and ending in Jan 2018.

According to the observation of trends in the correlation of the uncertain parameters, we choose and adapt seven powerful forecasting models to predict the three target variables of Demand, AirTemp, and Diesel_generators.

V. Forecasting Models

In the following, we briefly introduce the models and their specifications. The reader can find more explanations on the referred papers.

A. Seasonal Naive

The Seasonal Naive model predicts the value of the variable $y(T)$ at time T as follows:

$$y(T) = y(T-h) \tag{1}$$

In this formulation, h denotes the length of the seasonal pattern. According to the analysis results, the value of h is selected equal to the last 24 hours of variable profile in the Formentera dataset. As a result, this model repeats the pattern of last night's prediction [1, 9].

B. Prophet

The Prophet model [10-12] is a prediction model used in this study for prediction. This model provides a method for predicting time series data based on the add-on model in which nonlinear trends are tailored to the annual, weekly and daily seasonality plus holiday effects. It performs best with time series facing strong seasonal influences and several historical data seasons. The Prophet is strong in losing data and changing trends, and usually handles away from the market well.

C. Linear Regression

In this study, a linear regression model is used to implement the rolling window method of forecasting [13]. The features used to predict this model are the values of the three target variables in the last six-time steps and a number of time-based features including month, day of the week and time of day. In the next two models, the same set of features is used for prediction.

D. Random Forest

We also use the Random Forest (RF) model to implement the rolling window method of forecasting [14-16]. This method is a tree-based ensemble method for supervised classification. In this study, rather than selecting a subset of features, which may cause information loss, all features are

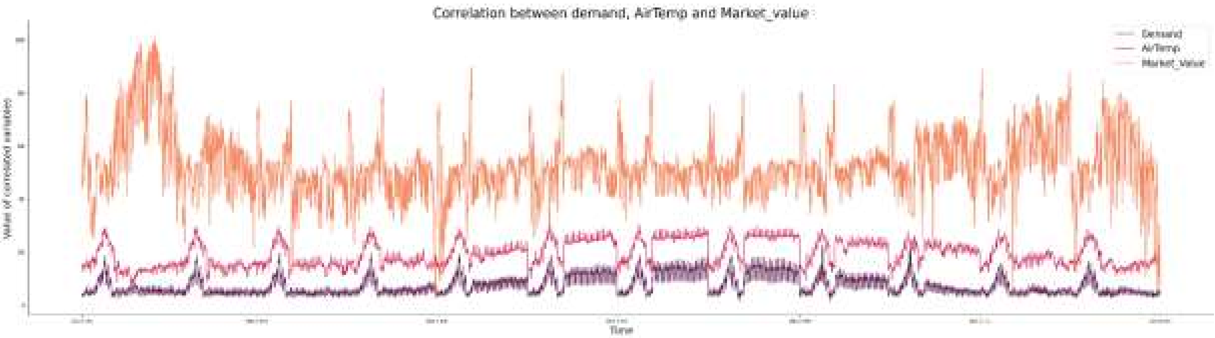

Fig. 3. Trends of Demand, Airtemp, and Market Value during one year

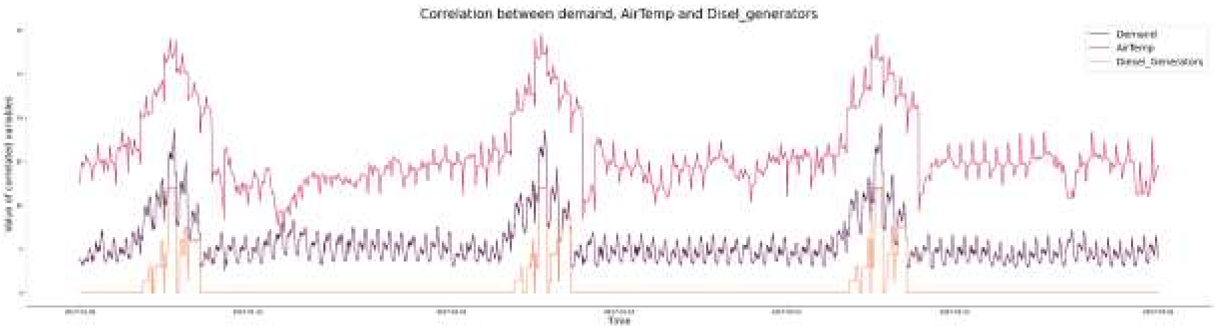

Fig. 4. Trends of Demand, Airtemp, and Diesel_generators during one year

considered in the proposed method. The RF method has better robustness to noise and missing data, and has a faster learning speed [17, 18].

E. Gradient Boosting

Gradient Boosting is an algorithm widely used for machine learning. In fact, this method is a machine learning technique for regression, classification and other tasks that produces a prediction model in the form of a set of poor prediction models, typically decision trees. This method has shown advanced results about tasks containing heterogeneous features, complex and noisy data such as air forecasting [19, 20].

F. Feedforward Neural Network

This model is an ANN, in which the connection between its constituent units does not form a cycle. In fact, this network is different from recursive neural networks. This ANN is the first and easiest type of artificial neural network. In this network, information only moves in a direction that is forward. In this study we use a Perceptron model of feedforward ANN with a hidden layer containing 10 neurons [21-22].

G. Gaussian Process Vector Autoregressive (GPVAR)

This model combines a Re-occurrent Neural Network (RNN)-based time series model with a low-rank covariance Coppola Gaussian model to reduce computational complexity and manage non-Gaussian marginal distributions. This allows the number of parameters to be drastically reduced, thus allowing the modeling of different time correlations of thousands of time series [23].

VI. PREPARATION OF DATASET AND EVALUATION OF FORECASTING MODELS

In this study, the time series related to the Formentera database is divided into 201 time series, each of them corresponding to one week and six hours. In the following evaluations, one-week data was used for training and the last six hours were used as the test data. The Python APIs of the models are used to provide the required codes.

TABLE 1. THE ERROR RATE OF DIFFERENT FORECASTING MODELS IN TERMS OF SMAPE

Model	AirTemp Error	Demand Error	Market_Value Error
Seasonal Naive	2.206	2.725	6.650
Prophet	1.673	2.654	6.170
Linear Regression	1.491	2.010	4.424
Random Forest	1.748	5.874	3.842
Gradient Boosting	1.608	3.614	3.766
Feed Forward Network	1.658	6.422	6.034
GPVAR	2.369	3.995	5.607

To compare the results, the Symmetric Mean Absolute Percentage Error (SMAPE) is used to evaluate the prediction of the models. This metric is defined as:

$$SMAPE = \frac{100}{n}\sum_{1}^{n}\frac{|F_t - A_t|}{|F_t| + |A_t|} \quad (2)$$

where A_t is the actual value, and F_t is the forecast value. Also, for probabilistic models, the median of prediction is used for evaluation. The results of evaluations are show in Table 1.

According to Table 1, the linear regression method is the most accurate method in forecasting the AirTemp and

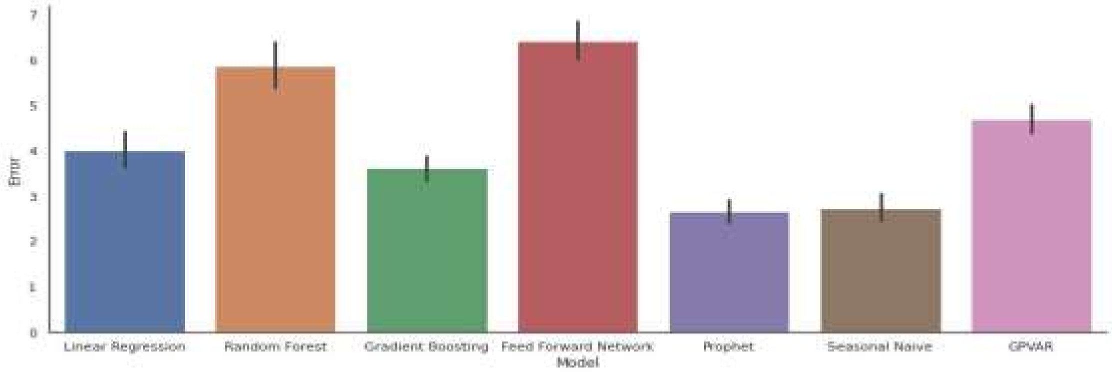

Fig. 5. Comparing the error of different forecasting models in forecasting the demand variable

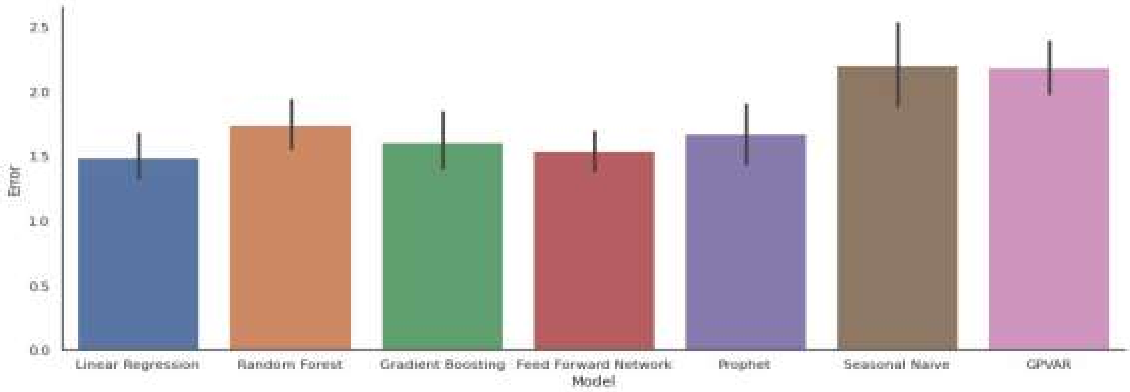

Fig. 6. Comparing the error of different forecasting models in forecasting the Airtemp variable.

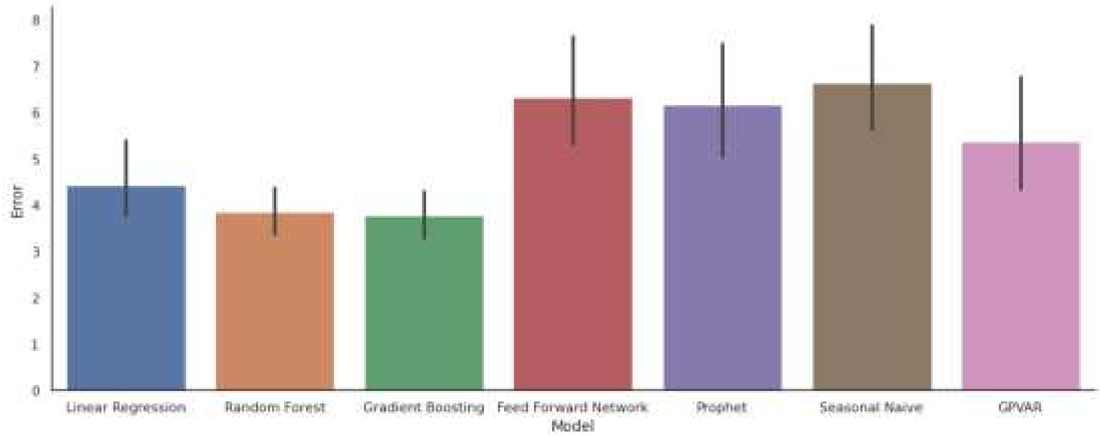

Fig. 7. Comparing the error of different forecasting models in forecasting the Market_value variable.

Demand variables with 1.491 and 2.010 percent of errors respectively. As shown in Figs. 3 and 4, these two variables have seasonal behaviors and are correlated with each other. For Market_value forecasting, the gradient boosting model is the best. As shown by Figs. 3, the Market_Value is not as correlated with the other uncertain variables.

Figs. 5 to 7 compare the accuracy of the competitor methods in terms of the error SMAPE in forecasting three uncertain variables correspondingly. As shown by Table 1 and confirmed by these figures, the Linear Regression, Random Forest, and Gradient Boosting are the best model for forecasting. Also, Fig. 8 compares the actual and frecasted values for the uncertain variables during six hours, every 10 minutes. For this short-term forecasting, the profile of the variables in the previous 24 hours, which includes 144 time steps of 10 minutes, was used to train the model. The plots of Fig. 8 confirm the superiority of simple models to the complex ANN-based models.

VII. Conclusion

Short-term forecasting of uncertain variables is a vital process in the VPPs. Especially, for real-time management of VPPs, a complex multi-objective optimization problem is formed that requires the accurately forecasted values of uncertain parameters. In this paper showed that precisely observing the attributes of the time series of the uncertain parameters of demand, temperature, and market price can considerably help in selecting the most efficient forecaster

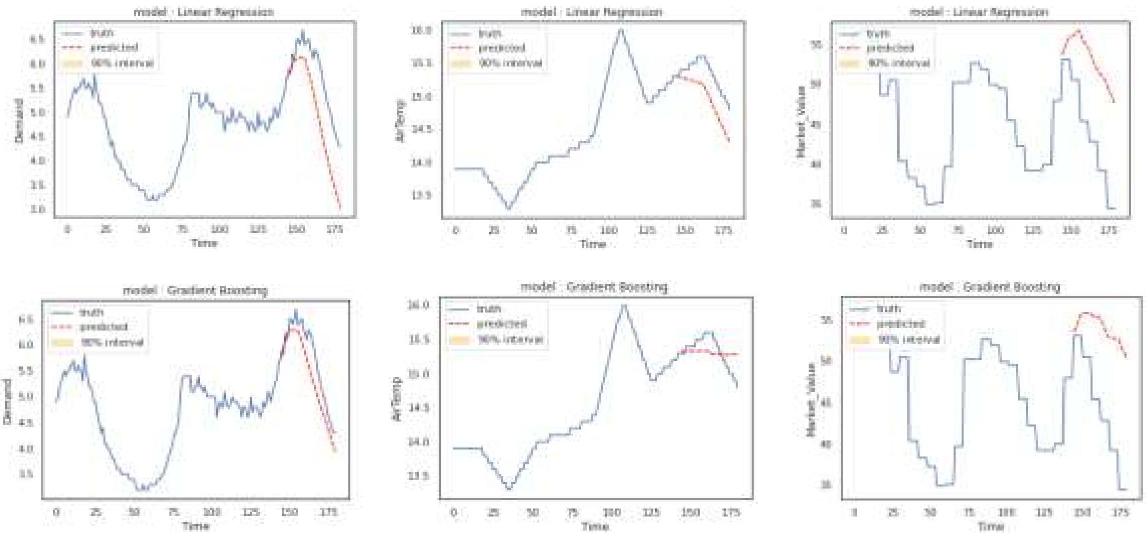

Fig. 8. Results of short-term forecasting.

models. We used seven powerful models of forecasting to predict the values of the short-term series of the demand, temperature, and market prices. Based on the results of the time-series analysis of the dataset, we divided the profiles to suitable train, and test sets. The results show that, enriched by analysis, the simple methods like regression and gradient boosting are more efficient as compared to complex ANN-based models.

ACKNOWLEDGMENT

This work has been carried out in the framework of the European Union's Horizon 2020 research and innovation program under grant agreement No 957852 (Virtual Power Plant for Interoperable and Smart isLANDS'-VPP4ISLANDS).

REFERENCES

[1] [1] A. Alexandrov, K. Benidis, M. Bohlke-Schneider, V. Flunkert, J. Gasthaus, T. Januschowski, et al., "GluonTS: Probabilistic and Neural Time Series Modeling in Python," J. Mach. Learn. Res., vol. 21, pp. 1-6, 2020.

[2] [2] K. Mahmud, B. Khan, J. Ravishankar, A. Ahmadi, and P. Siano, "An internet of energy framework with distributed energy resources, prosumers and small-scale virtual power plants: An overview," Renewable and Sustainable Energy Reviews, vol. 127, p. 109840, 2020.

[3] [3] M. M. Othman, Y. Hegazy, and A. Y. Abdelaziz, "A Review of virtual power plant definitions, components, framework and optimization," International Electrical Engineering Journal, vol. 6, pp. 2010-2024, 2015.

[4] [4] Z. Wei, S. Yu, G. Sun, Y. Sun, Y. Yuan, and D. Wang, "Concept and development of virtual power plant," Automation of Electric Power Systems, vol. 37, pp. 1-9, 2013.

[5] [5] A. Alahyari, M. Ehsan, D. Pozo, and M. Farrokhifar, "Hybrid uncertainty-based offering strategy for virtual power plants," IET Renewable Power Generation, vol. 14, pp. 2359-2366, 2020.

[6] [6] M. Rahimiyan and L. Baringo, "Real-time energy management of a smart virtual power plant," IET Generation, Transmission & Distribution, vol. 13, pp. 2015-2023, 2019.

[7] [7] R. Fildes and N. Kourentzes, "Validation and forecasting accuracy in models of climate change," International Journal of Forecasting, vol. 27, pp. 968-995, 2011.

[8] [8] S. Khatoon and A. K. Singh, "Effects of various factors on electric load forecasting: An overview," in 2014 6th IEEE Power India International Conference (PIICON), 2014, pp. 1-5.

[9] [9] Z. Pala and R. Atici, "Forecasting sunspot time series using deep learning methods," Solar Physics, vol. 294, pp. 1-14, 2019.

[10] [10] A. I. Almazrouee, A. M. Almeshal, A. S. Almutairi, M. R. Alenezi, and S. N. Alhajeri, "Long-term forecasting of electrical loads in kuwait using prophet and holt–winters models," Applied Sciences, vol. 10, p. 5627, 2020.

[11] [11] S. J. Taylor and B. Letham, "Forecasting at scale," The American Statistician, vol. 72, pp. 37-45, 2018.

[12] [12] P. E. Argo and I. J. Rothmuller, "PROPHET: An application of propagation forecasting principles," in Solar-Terrestrial Predictions Proceedings, 1979, pp. 312-321.

[13] [13] E. Zivot and J. Wang, "Rolling analysis of time series," in Modeling Financial Time Series with S-Plus®, ed: Springer, 2003, pp. 299-346.

[14] [14] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001/10/01 2001.

[15] [15] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine Learning, vol. 63, pp. 3-42, 2006/04/01 2006.

[16] [16] Y.-Y. Cheng, P. P. Chan, and Z.-W. Qiu, "Random forest based ensemble system for short term load forecasting," in 2012 international conference on machine learning and cybernetics, 2012, pp. 52-56.

[17] [17] D. Niu, K. Wang, L. Sun, J. Wu, and X. Xu, "Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study," Applied Soft Computing, vol. 93, p. 106389, 2020/08/01/ 2020.

[18] [18] S. R. Joelsson, J. A. Benediktsson, and J. R. Sveinsson, "Feature Selection for Morphological Feature Extraction using Randomforests," in Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG 2006, 2006, pp. 138-141.

[19] [19] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Annals of statistics, pp. 1189-1232, 2001.

[20] [20] A. Malinin, L. Prokhorenkova, and A. Ustimenko, "Uncertainty in gradient boosting via ensembles," arXiv preprint arXiv:2006.10562, 2020.

[21] [21] T. L. Fine, Feedforward neural network methodology: Springer Science & Business Media, 2006.

[22] [22] F. Saâdaoui, "A seasonal feedforward neural network to forecast electricity prices," Neural Computing and Applications, vol. 28, pp. 835-847, 2017.

[23] K. Kolokythas and A. Argiriou, "An application of a feed-forward neural network model for wind speed predictions," International Journal of Sustainable Energy, pp. 1-18, 2021.