International Conference on Applied Energy 2021
Nov. 29 - Dec. 2, 2021, in Bangkok, Thailand
Paper ID: 844

A Unique Three-Step Weather Data Approach in Solar Energy Prediction Using Machine Learning

Tolulope Olumuyiwa Falope[1], Liyun Lao[1*], Dawid Hanak[1*]

1 Energy and Power, School of Water, Energy & Environment, Cranfield University, MK41 0AL, UK

[*] Corresponding authors. Email: l.lao@cranfield.ac.uk, d.p.hanak@cranfield.ac.uk

ABSTRACT

The importance of renewable energy sources like solar energy in reducing carbon emissions and other greenhouse gases has contributed to an increase in grid integration. However, the intermittent nature of solar power causes reliability issues and a loss of energy balance in the system, which are barriers to solar energy penetration. This study proposes a unique three-step approach that identifies weather parameters with moderate to strong correlation to solar radiation and uses them to predict solar energy generation. The combination of an on-site weather station and a reliable local weather station produces relevant data that increases the accuracy of the forecasting model irrespective of the machine learning algorithm used. This data source combination is tested, along with two other scenarios, using the exponential Gaussian Process Regression machine learning algorithm in MATLAB. It was found to be the most effective algorithm with a Normalized Root Mean Square Error of 1.1922, and an R^2 value of 0.66.

Keywords: renewable energy sources, variable renewable energy, machine learning, solar energy forecasting, Gaussian Process Regression

NONMENCLATURE

Abbreviations	
GPR	Gaussian Process Regression
NWP	Numerical Weather Predictions
PV	Photovoltaic
RE	Renewable Energy
RES	Renewable Energy Sources
RMSE	Root Mean Square Error
NRMSE	Normalized Root Mean Square Error
VRE	Variable Renewable Energy

1. INTRODUCTION

While the influx of renewable energy (RE) technologies has provided both off-grid and on-grid capabilities for power generation, the intermittent and variable nature of renewable resources poses a challenge for grid operators in terms of forecasting and meeting load demands as energy generated does not always coincide with consumption. As penetration levels increase, utility operators and prosumers must balance supply and energy demand to maintain a reliable system. Grid integration is the practice of developing efficient and cost-effective ways of incorporating variable renewable energy (VRE) into the power system while maintaining or increasing system stability and reliability [1].

Renewable energy sources (RES) like solar and wind are relatively infinite in supply, have wide adaptations and are increasingly being deployed to satisfy energy demand. The ability to integrate solar into the grid will depend on the capacity to predict or forecast generation accurately [2]. Improved generation forecasting is critical to effective grid integration and planning [3].

Solar forecasting is the process of predicting future solar irradiance or solar power generated from historical and/or present meteorological observations [4]. These observations can be provided by on-site weather stations, local/regional weather stations, remote sensing (satellite imaging), to name a few.

There are largely four broad considerations to make in deciding a solar forecasting approach. These are the source and type of input data, mapping of the input data

Selection and peer-review under responsibility of the scientific committee of the 13th Int. Conf. on Applied Energy (ICAE2021).

or the forecasting architecture, the forecasting methodology, and the predicted outcome.

Weather data types are classified into forecast or historical and different weather stations provide one or the other and sometimes both. Apart from this classification, the source of the weather data is equally critical. Local, regional, and global weather data though interdependent, provide differing levels of accuracy. Majority of the studies reviewed get their weather data from either local or regional weather stations. While this data can be relatively accurate, they may not capture local conditions unique to that specific site. This weather data can be further refined by the presence of an on-site weather station [5]. This paper proposes using both a local weather station and an on-site weather station to increase the accuracy of the prediction model.

The forecasting architecture characterizes the mapping between input variables (meteorological observations) and output variables [4] which in our case refers to the mapping of the input weather data to the predicted solar power. Several forecasting architectures currently exist. This paper proposes a unique three-step approach that identifies weather parameters from both an on-site and a local weather station, with moderate to strong correlations to solar irradiation, and uses them along with the historical Photovoltaic (PV) output data as inputs into the machine learning algorithm to predict solar power generated. The logic is that the more relevant data included in the forecasting model, the better the accuracy. Figure 1 shows a graphical illustration of this unique three-step forecasting architecture.

Solar forecasting methods are divided into physical methods and statistical (or non-physical) methods. Physical methods use weather parameters like temperature, humidity, pressure, etc. from a specific region as inputs into numerical weather prediction (NWP) models to predict weather conditions like solar irradiation [6]. The Regression Learner App in the machine learning toolbox in MATLAB was used to train several models. The exponential Gaussian Process Regression (GPR) model produced the lowest Root Mean Square Error (RMSE) in all the scenarios tested.

This study contributes to research in the following ways:

- This study proposes a novel approach that combines relevant weather data from two dependable data sources as predictors to improve solar power prediction.
- The importance of relevant data to a forecasting problem.
- The proposed forecasting approach can be adapted to other RE sources that use weather for predictions for example wind and tidal.

2. MATERIAL AND METHODS

This section describes the data and method used to develop the forecasting model. Section 2.1 describes the data source and type. Section 2.2 briefly explains the correlation coefficient and matrix used to determine the most relevant weather parameters. Section 2.3 highlights NRMSE as a tool in selecting the most accurate model. Section 2.4 highlights the use of the Regression Learner App in MATLAB and the choice of the machine learning algorithm used to determine the predicted outcomes.

2.1 Data Source and Type

The source of the data used is critical to the performance of a model [7] and the more relevant data available, the better the prediction. With regards to data type, some studies pair the solar energy produced from solar panels with either exclusively historically observed data [8–18] or forecast data [19–22]. Others use only the historical PV output data [23,24]. Recall that the on-site weather station can capture local conditions that affect the weather that are not reflected in local or regional weather data. It is important to note that for the studies that used an on-site weather station [25–30], none of them combined the on-site data with data from a dependable local weather station; an approach this paper proposes.

In this study, nine weather parameters, including solar irradiation, were recorded every fifteen minutes by the Kisanhub weather station (KH075) (www.livinglab.ac.uk/data.html) located on site in Cranfield University. A power meter records half-hourly energy supply data (https://cranfield.energymanagerlive.com/) from the 1 MW solar farm located in the airfield of the Cranfield University airport. Data for twelve weather parameters was also provided on request, by The Metrological Office for Bedford, U.K which is the closest station to Cranfield. Hourly data points were collated for January to December 2019 across the two weather stations and the PV meter data. Days with non-available data were deleted from the data set. 70% of the data points were used to train the model while 30% were used for testing.

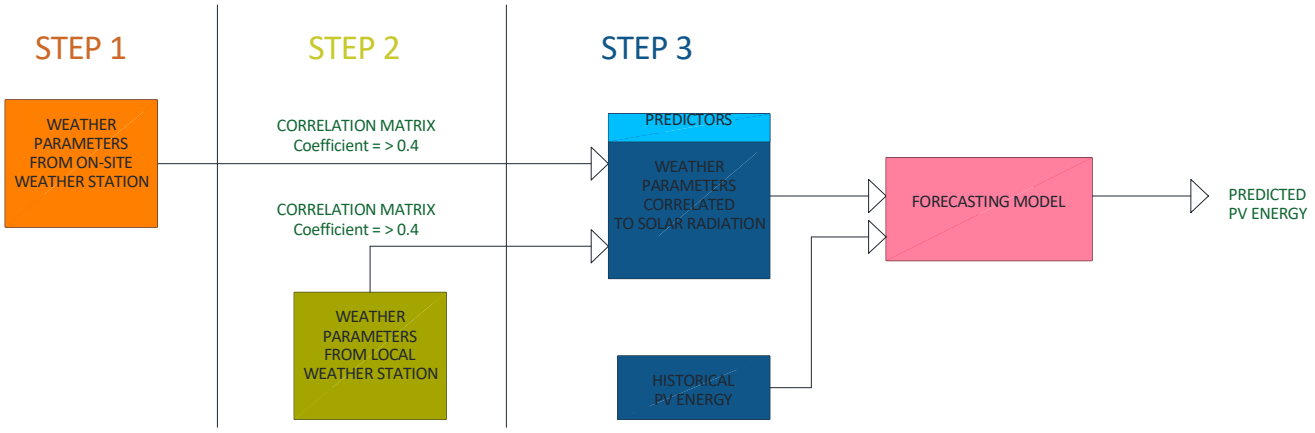

Fig. 1. Three-step forecasting architecture

2.2 *Data Comparison*

The correlation function in the data analysis tool pack of excel calculates the Pearson Coefficient between more than two measurement variables with N number of subjects and displays the output in a correlation matrix [31]. The Pearson Correlation Coefficient formula [32] is given by:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \quad (1)$$

2.3 *Model Comparison*

Though the Regression Learner App selects the best model based on RMSE, the NRMSE is a better measure of the performance of the model irrespective of the model type or data set used. The NRMSE formula [33] is given by:

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2}}{\frac{1}{N}\sum_{i=1}^{N} x_i} \quad (2)$$

2.4 *MATLAB: Regression Learner App*

After the data has been cleaned, the data is trained using nineteen different regression models using the PV energy supplied by the 1 MW solar farm as the "response" and the correlated weather data as the "predictors". The app selects the best prediction model based on the lowest RMSE value which in this case was the exponential GPR model.

3. THEORY/CALCULATIONS

Local weather station may capture relevant weather parameters to solar radiation that an on-site weather station may not capture and vice versa hence the reasoning in using both. This section puts forth a premise for the above statement. Section 3.1 describes the unique three-step forecasting architecture used to map input data to energy supply. Section 3.2 highlights a case study of three scenarios to test the effectiveness of the forecasting architecture.

3.1 *Forecasting Architecture*

[7] uses a three-step approach using an auxiliary model for the first step and a main model for the second. The auxiliary model is used to determine correlations between parameters in the forecast data and identified auxiliary variables in the weather data and outputs the result into the main model. The main model then uses the output result to predict the energy generated. The result is an overall increase in the accuracy of the forecasting model. Another approach is proposed by [34] where the observed data are used to adjust the forecast data rather than identifying auxiliary variables.

In the first step of this paper's unique approach, the correlation matrix of the nine weather parameters (inclusive of solar radiation) recorded by the on-site weather station in Cranfield is determined with respect to the solar radiation because solar radiation remains the most important predictor of energy supply generated. In the second step, the correlation matrix is determined with respect to the same solar radiation from the previous step, for twelve weather parameters recorded by the Bedford station. In the third and final step, the weather parameters from the two weather stations with

moderate to strong correlations (> 0.4) are filtered out as predictors for the forecasting model. These along with the response, which is the solar energy generated as recorded by the PV meter, are inputs into the machine learning program.

3.2 Case Study

To test the performance of the forecasting architecture, three different scenarios are studied and compared.

a. Scenario 1: The correlated data of the on-site Cranfield weather station alone.
b. Scenario 2: The correlated data of the Bedford weather station alone.
c. Scenario 3: A combination of the correlated data of both the on-site Cranfield weather station and the Bedford weather station are used.

4. RESULTS AND DISCUSSION

This section presents the results of the methods and case study highlighted in the previous sections. Section 4.1 presents the result of the correlation matrix of the weather parameters recorded on-site. Section 4.2 presents the correlation matrix of the weather parameters recorded by the Bedford weather station. Section 4.3 highlights the most accurate prediction model based on NRMSE and the R square values across the three case scenarios highlighted above. It also shows an example of the performance of the model in a next day prediction.

4.1 Correlation Matrix of the Cranfield On-Site Station

	Solar Radiation
Solar Radiation	1
Wind Speed	0.176545495
Relative Humidity	-0.60135733
Pressure	0.144238691
Precipitation	-0.071483573
Dew Point	0.265970205
Heat Index	**0.505402924**
Temperature	**0.518615159**
Wind Chill	**0.499023585**

Table 1. Correlation matrix of weather parameters in Cranfield

4.2 Correlation Matrix of Bedford Weather Station

	Solar Radiation
Solar Radiation	1
Hourly Pressure at Mean Sea Level (hPa)	0.124996778
Hourly Dewpoint Temperature (°C)	0.248986632
Hourly Rainfall Total (mm)	-0.084914974
Hourly Relative Humidity (%)	-0.624109028
Hourly Mean Wind Direction (o)	-0.032743693
Hourly Mean Windspeed (knots)	0.140534017
Hourly Maximum Gust (knots)	0.186917781
Hourly Temperature (°C)	**0.54943909**
Hourly Wet Bulb Temperature (°C)	**0.441551628**
Hourly Global Radiation (kJ/m^2)	**0.932498311**
Hourly Total Cloud Cover (oktas)	-0.144885133
Hourly Visibility (metre)	0.183733904

Table 2. Correlation matrix of weather parameters in Bedford

4.3 Performance of Forecasting Architecture and Model

Exponential GPR was selected as the best performing of all the regression models based on having the lowest NRMSE in each of the three scenarios. Table 3 shows the comparison of the 3 scenarios, with the proposed forecasting architecture highlighted in Scenario 3.

	Train		**Test**	
Scenario	**NRMSE**	**R^2**	**NRMSE**	**R^2**
Scenario 1	0.0057	1.00	1.4842	0.47
Scenario 2	0.0113	1.00	1.2308	0.64
Scenario 3	0.0006	1.00	1.1922	0.66

Table 3. Performance comparison of three-step forecasting architecture

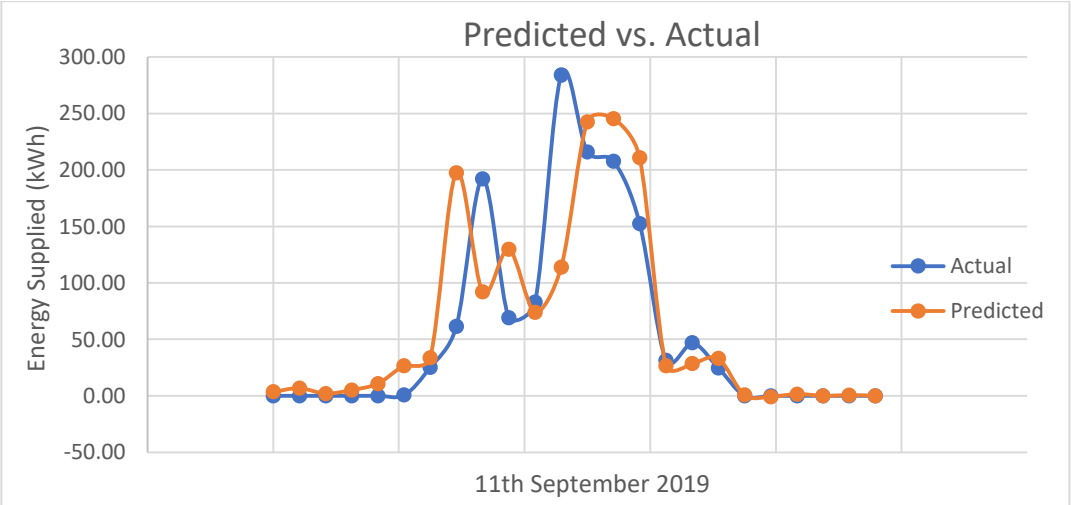

Fig. 2. Next day solar prediction

5. CONCLUSION

It is clear to see from the results that the three-step forecasting architecture used in mapping correlated weather data from both an on-site weather station and a nearby local weather station, referred to as scenario 3 in this study, outperforms the other scenarios. In both the training and testing phases, scenario 3 has the lowest NRMSE and highest R^2 values of all the scenarios. Combining useful data from an on-site weather station and a dependable local weather station increased the number of relevant weather parameters as predictors in the forecasting model.

ACKNOWLEDGEMENT

I wish to acknowledge the help and support of Mr. Gareth Ellis and Mr. Angus Murchie, both from the Energy & Environment Team, Cranfield University and Mark Beswick, Archive Information Officer from the Met Office National Meteorological Archive, without whom this research would not have been possible.

REFERENCE

[1] Katz J, Cochran J. INTEGRATING VARIABLE RENEWABLE ENERGY INTO THE GRID: KEY ISSUES GREENING THE GRID GRID INTEGRATION TERMINOLOGY. National Renewable Energy Laboratory 2015;NREL/FS-6A20-63033.

[2] Basmadjian R, Niedermeier F, de Meer H. Demand-Side Flexibility and Supply-Side Management: The Use Case of Data Centers and Energy Utilities 2017:187–204. https://doi.org/10.1007/978-3-319-65082-1_9.

[3] Cox S, Xu K. Integration of Large-Scale Renewable Energy in the Bulk Power System: Good Practices from International Experiences. National Renewable Energy Laboratory 2020.

[4] Li B, Zhang J. A review on the integration of probabilistic solar forecasting in power systems. Solar Energy 2020;210:68–86. https://doi.org/10.1016/J.SOLENER.2020.07.066.

[5] Agüera-Pérez A, Palomares-Salas JC, González de la Rosa JJ, Sierra-Fernández JM, Jiménez-Montero A. Nowcasting and Short-Term Wind Forecasting for Wind Energy Management. In: Moreno-Munoz A, editor. Large Scale Grid Integration of Renewable Energy Sources, Institution of Engineering and Technology; 2017, p. 59–85.

[6] Tian T, Chernyakhovskiy I. Forecasting Wind and Solar Generation: Improving System Operations, Greening the Grid. National Renewable Energy Laboratory 2016.

[7] Kim S-G, Jung J-Y, Sim MK. A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning. Sustainability 2019. https://doi.org/10.3390/su11051501.

[8] Ngoc-Lan Huynh A, Deo RC, Ali M, Abdulla S, Raj N. Novel short-term solar radiation hybrid model: Long short-term memory network integrated with robust local mean decomposition. Applied Energy 2021;298. https://doi.org/10.1016/j.apenergy.2021.117193.

[9] Pedregal DJ, Trapero JR. Adjusted combination of moving averages: A forecasting system for medium-term solar irradiance. Applied Energy 2021;298. https://doi.org/10.1016/j.apenergy.2021.117155.

[10] Ahmad T, Zhang D. Renewable energy integration/techno-economic feasibility analysis, cost/benefit impact on islanded and grid-connected operations: A case study. Renewable Energy 2021;180:83–108. https://doi.org/10.1016/J.RENENE.2021.08.041.

[11] Korkmaz D. SolarNet: A hybrid reliable model based on convolutional neural network and variational mode

decomposition for hourly photovoltaic power forecasting. Applied Energy 2021;300. https://doi.org/10.1016/j.apenergy.2021.117410.

[12] Najibi F, Apostolopoulou D, Alonso E. Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast. International Journal of Electrical Power and Energy Systems 2021;130. https://doi.org/10.1016/j.ijepes.2021.106916.

[13] Rodríguez F, Fleetwood A, Galarza A, Fontán L. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. Renewable Energy 2018;126:855–64. https://doi.org/10.1016/J.RENENE.2018.03.070.

[14] Wang J, Li P, Ran R, Che Y, Zhou Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree n.d. https://doi.org/10.3390/app8050689.

[15] Thukral MK. Solar Power Output Prediction Using Multilayered Feedforward Neural Network: A Case Study of Jaipur. 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), 2020, p. 1–6. https://doi.org/10.1109/iSSSC50941.2020.9358821.

[16] Yakoubi H, el Mghouchi Y, Abdou N, Hajou A, Khellouki A. Correlating clearness index with cloud cover and other meteorological parameters for forecasting the global solar radiation over Morocco. Optik 2021;242. https://doi.org/10.1016/j.ijleo.2021.167145.

[17] Zang H, Cheng L, Ding T, Cheung KW, Liang Z, Wei Z, et al. Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. IET Gener Transm Distrib 2018;12:4557–67. https://doi.org/10.1049/iet-gtd.2018.5847.

[18] Chen B, Lin P, Lai Y, Cheng S, Chen Z, Wu L. Very-Short-Term Power Prediction for PV Power Plants Using a Simple and Effective RCC-LSTM Model Based on Short Term Multivariate Historical Datasets. Electronics 2020;9. https://doi.org/10.3390/electronics9020289.

[19] Iyengar S, Sharma N, Irwin D, Shenoy P, Ramamritham K. SolarCast-A Cloud-based Black Box Solar Predictor for Smart Homes. BuildSys '14: 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, 2014, p. 40–9. https://doi.org/10.1145/2674061.2674071.

[20] Andrade JR, Bessa RJ. Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions. IEEE Transactions on Sustainable Energy 2017;8th:1571–80. https://doi.org/10.1109/TSTE.2017.2694340.

[21] Leva S, Dolara A, Grimaccia F, Mussetta M, Ogliari E. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. Mathematics and Computers in Simulation 2017;131:88–100. https://doi.org/10.1016/J.MATCOM.2015.05.010.

[22] Persson C, Bacher P, Shiga T, Madsen H. Multi-site solar power forecasting using gradient boosted regression trees. Solar Energy 2017;150:423–36. https://doi.org/10.1016/J.SOLENER.2017.04.066.

[23] Anaadumba R, Liu Q, Marah BD, Nakoty FM, Liu X, Zhang Y. A renewable energy forecasting and control approach to secured edge-level efficiency in a distributed micro-grid. Cybersecurity 2021;4. https://doi.org/10.1186/s42400-020-00065-3.

[24] Kushwaha V, Pindoriya NM. A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast | Elsevier Enhanced Reader. Elsevier Ltd 2019;140:124–39. https://doi.org/10.1016/j.renene.2019.03.020.

[25] Sarp S, Kuzlu M, Cali U, Guler O. An Interpretable Solar Photovoltaic Power Generation Forecasting Approach Using An Explainable Artificial Intelligence Tool. IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2021, p. 1–5. https://doi.org/10.1109/ISGT49243.2021.9372263.

[26] Aprillia H, Yang H-T, Huang C-M. Short-Term Photovoltaic Power Forecasting Using a Convolutional Neural Network-Salp Swarm Algorithm n.d. https://doi.org/10.3390/en13081879.

[27] Cannizzaro D, Aliberti A, Bottaccioli L, Macii E, Acquaviva A, Patti E. Solar radiation forecasting based on convolutional neural network and ensemble learning. Expert Systems with Applications 2021;181. https://doi.org/10.1016/j.eswa.2021.115167.

[28] Gu B, Shen H, Lei X, Hu H, Liu X. Forecasting and uncertainty analysis of day-ahead photovoltaic power using a novel forecasting method. Applied Energy 2021;299. https://doi.org/10.1016/j.apenergy.2021.117291.

[29] Nguyen NQ, Bui LD, Doan BV, Sanseverino ER, Cara DD, Nguyen QD. A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks a case study in Vietnam. Electric Power Systems Research 2021;199. https://doi.org/10.1016/j.epsr.2021.107427.

[30] Pan C, Tan J, Feng D. Prediction intervals estimation of solar generation based on gated recurrent unit and kernel density estimation. Neurocomputing 2021;453:552–62. https://doi.org/10.1016/j.neucom.2020.10.027.

[31] Microsoft. Use the Analysis ToolPak to perform complex data analysis. Microsoft Support 2021. https://support.microsoft.com/en-us/office/use-the-analysis-toolpak-to-perform-complex-data-analysis-6c67ccf0-f4a9-487c-8dec-bdb5a2cefab6 (accessed July 15, 2021).

[32] Microsoft. PEARSON function. Microsoft Support 2021. https://support.microsoft.com/en-us/office/pearson-function-0c3e30fc-e5af-49c4-808a-3ef66e034c18 (accessed October 15, 2021).

[33] Ghofrani M, Alolayan M. Time Series and Renewable Energy Forecasting. Time Series Analysis and Applications 2017. https://doi.org/10.5772/INTECHOPEN.70845.

[34] Kyliashkina IA, Eroshenko SA, Shelyug S. Intelligent Systems as a Tool for Predicting Electrical Energy and Power Generation. 2019 60th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), 2019, p. 1–5. https://doi.org/10.1109/ITMS47855.2019.8940721.