JOURNAL OF
EMERGING INVESTIGATORS

Article

An analysis of the feasibility of SARIMAX-GARCH through load forecasting

Anirudh Satheesh[1], Sateesh Sundararagan[2]
[1] Thomas Jefferson High School for Science and Technology, Alexandria, Virginia
[2] Development Finance Industry, Alexandria, Virginia

SUMMARY

Load forecasting is critical for energy sector planners and generation companies to predict the level of energy that should be generated to maximize energy security. Accurate predictions of future energy consumption and supply capacity increase public confidence in energy sector planners, and help make informed decisions on power generation infrastructure to reduce capital costs. To forecast load, energy sector planners have used variations of the AutoRegressive Integrated Moving Average (ARIMA) model, such as Seasonal AutoRegressive Integrated Moving Average with eXogenous factors (SARIMAX). However, the accuracy of these models is limited due to heteroskedasticity, where the variance of data is not constant. Consequently, ARIMA can be combined with General AutoRegressive Conditional Heteroskedasticity (GARCH), a model that forecasts variance, in the SARIMAX-GARCH model. We hypothesized that SARIMAX-GARCH will be more accurate in predicting load than SARIMAX, a variant of ARIMA. We trained SARIMAX-GARCH and SARIMAX and selected the best models using Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC), resulting in a Mean Absolute Percentage Error (MAPE) of 13.2% for SARIMAX and 11.0% for SARIMAX-GARCH. This shows that SARIMAX-GARCH is more accurate than SARIMAX for load forecasting data. The results indicate that SARIMAX-GARCH could potentially be improved through ensemble techniques and other exogenous variables. The results of our study will help energy sector planners and generation companies in forecasting energy consumption with more accurate predictions.

INTRODUCTION

Energy sector planners across countries strive to maintain energy security for all people. One of the factors that contributes to energy security is load forecasting, which predicts energy consumption in advance so that electrical utilities and generation companies can plan and make decisions regarding the amount of power produced at specific periods (1). Energy security risk has been exacerbated by the ongoing Russia-Ukraine crisis, so it is paramount that load forecasting on energy consumption and delegation are accurate (2).

Additionally, there has been a global increase in renewable energy, and while this contributes to clean energy and climate change goals, it also contributes to variability in energy generation as renewable energy cannot be generated in certain regions at all times of the year (3). Governments can combat this by using load forecasting to understand how such variable energy production can be integrated with battery storage, and how that energy can be delegated to maximize energy security. One example of using load forecasting for renewable energy is in Kazakhstan, where the Power the Future program predicted energy production of 22 different plants. Through load forecasting, the program was able to predict renewable energy consumption accurately and aims to increase renewable energy generation to 50% by 2050 (4).

To predict power consumption, time series models such as AutoRegressive Integrated Moving Average (ARIMA) can be used. These models input past energy consumption data from utilities, such as power substations, and output a prediction of power consumption at a specific time interval. ARIMA is made of two different models, AutoRegressive and Moving Average, which are integrated using differencing. There are three parameters in the ARIMA model: The lag order p represents the number of previous data points that are used to predict the load at the next time, the moving average window q, which tells how the data is smoothed over time to increase accuracy, and differencing term d, which is used because it increases stationarity, which is necessary for ARIMA (5). Variations of ARIMA are also used, such as Seasonal AutoRegressive Integrated Moving Average (SARIMA) and Seasonal AutoRegressive Integrated Moving Average with eXogenous factors (SARIMAX). SARIMA incorporates seasonality trends in the model, which are denoted by the letter s in the model parameters. In addition to s, SARIMA contains three extra parameters for the seasonal data: P, Q, and D, which represent the seasonal counterparts of p, q, and d (6). SARIMAX also contains an extra parameter for an exogenous variable, such as temperature, which is used along with the p, q, and d parameters to create a more optimal model (7). Another model that can be used is GARCH (General AutoRegressive Conditional Heteroskedasticity), which predicts the conditional variance of the dataset. GARCH models assume that the conditional variance of the dataset is not constant, also known as heteroskedasticity, which is often the case with time series data but can only have an input of the residuals after another time series model, such as ARIMA, has forecasted the data (8). One problem with variations of ARIMA is that conditional variance can lead to inaccurate results, as the model is not equipped to handle heteroskedasticity well. To combat conditional variance, many models combine the results from ARIMA and GARCH to create a more reliable output (9).

In this study, we used both SARIMAX and SARIMAX-GARCH models to forecast electric load, with the seasonality trend as the exogenous variable. We hypothesized that the SARIMAX-GARCH model would have a lower MAPE (mean absolute percent error) than the SARIMAX model. In this research, our team pre-processed the training and testing datasets of load forecasting in Panama, in which the training dataset had 36,000 data points and the testing dataset had 2,300 data points. We used this dataset to create both the SARIMAX model and the SARIMAX-GARCH model, which we compared using MAPE. We found that the SARIMAX-GARCH model had an MAPE of 11.0%, whereas the SARIMAX model had an MAPE of 13.2%, which signifies that SARIMAX-GARCH is a more accurate model.

RESULTS

Before training the SARIMAX and SARIMAX-GARCH models, we checked for stationarity by using the Augmented Dickey-Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. Stationarity is a property that the mean, variance, and covariance do not change with time. The ADF test has a null hypothesis that the training data is non-stationary and an alternative hypothesis it is stationary (11). The KPSS has the opposite null hypothesis and alternative hypothesis and was run to further test for stationarity (12). The ADF test returned a p-value of 0.56 and the KPSS test returned a p-value of 0.0017, showing that the data is non-stationary. However, since the SARIMAX and SARIMAX-GARCH models are variations of ARIMA, they have differencing parameters to transform non-stationary data into stationary data. This means that non-stationary data will not significantly affect the output of either model.

We trained both the SARIMAX and SARIMAX-GARCH models on the Panama dataset, which contains load consumption data from 2015-2019. We used 70% of the preprocessed dataset for training and the remainder for testing. Next, we generated all permutations of model parameters with values of either 0 or 1 and selected the model with the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for training. Typically, model parameters are generated with a larger range, but this would significantly increase the amount of time required to select the best model based on AIC and BIC, so parameters had values of either 0 or 1.

The results from training each type of model revealed that SARIMAX(1, 1, 0, 1, 0, 1, 12) and SARIMAX(1, 0, 1, 1, 0, 0, 12)-GARCH(1, 1) had the lowest Information Criterion. SARIMAX (1, 1, 0, 1, 0, 1, 12) had an AIC of 525.634 and a BIC of 551.672 and SARIMAX(1, 0, 1, 1, 0, 0, 12)-GARCH(1, 1) had an AIC of 511.357 and a BIC of 528.299. We forecasted energy loads using these models on the testing dataset and compared their accuracy using Mean Absolute Percentage Error (MAPE), since it is a percentage of the total error between the forecasted values and the data points and allows us to compare the error between different models.

The results of our study indicated that SARIMAX-GARCH had a MAPE of 11.0%, while SARIMAX had a MAPE of 13.2%, showing that SARIMAX-GARCH is the more accurate model (**Figure 1**). Thus, energy sector planners and generation companies could enhance their accuracy in forecasting by using SARIMAX-GARCH in their electrical utilities to maximize energy security. However, even with the additional computational power and reduction of heteroskedasticity, there was not a significant decrease in the MAPE.

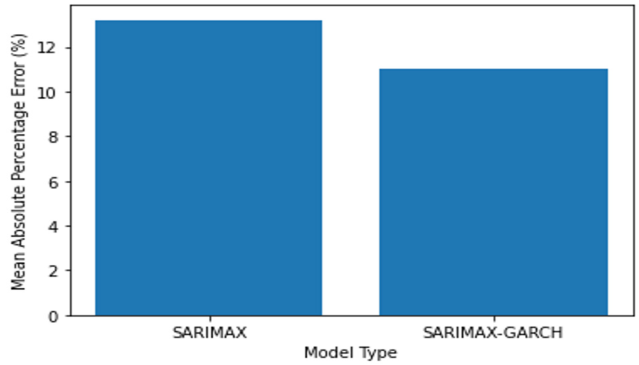

Figure 1: Mean Absolute Percentage Error of SARIMAX and SARIMAX-GARCH. The difference in Mean Absolute Percentage Error of SARIMAX and SARIMAX-GARCH, with SARIMAX-GARCH having a 2.2% decrease in MAPE as compared to SARIMAX.

DISCUSSION

After analyzing the models, the results showed that SARIMAX-GARCH was a better predictor than SARIMAX for energy load time series data, based on a lower MAPE. Since many energy sector planners use variants of ARIMA such as SARIMAX in load forecasting in addition to other models such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM), SARIMAX-GARCH may be a viable alternative (13). Our results agrees with our initial hypothesis that SARIMAX-GARCH would be more accurate, but the error for the SARIMAX-GARCH model could be reduced to offset the additional computational cost.

One possible way to decrease the error would be to transform the data to make it more stationary. Many time-series models, including variants of ARIMA, assume that each data point is independent of the previous lag values, which is not the case in non-stationary datasets. While stationarity was not needed to train both SARIMAX and SARIMAX-GARCH, as differencing would reduce non-stationarity to an extent, any increase in stationarity of the data would result in higher accuracy (14). To achieve this, we could take the natural logarithm of the data to remove any skews and exponential trends first, and then use the differencing operator (15). However, we also would have to run ADF tests and KPSS tests on the data to ensure stationarity.

Incorporating other exogenous variables could also improve the results of the load forecasting model. For example, factors such as weather patterns, temperature, and wind speeds affect the amount of load needed and may be impactful factors than solely the seasonal trend. The next step would be to include multiple exogenous factors, but this could have setbacks, as the model could prioritize the exogenous variables over the data points (16).

We could also apply ensemble techniques such as gradient boosting to find more optimal *p*, *q*, and *d* parameters. Gradient boosting minimizes the bias error in the training dataset, which would likely lead to a more accurate model. This is done by creating several weak learners and successively combining them to minimize the loss function . However, gradient boosting introduces complexities that could interfere with SARIMAX and SARIMAX-GARCH's predictability, while also being much more computationally expensive. In addition, gradient boosting may generalize poorly to the testing data due to overfitting, leading to an overemphasis on outliers (17). Finally, we could change the models entirely from statistical-based models to neural networks, which are typically superior to variations of ARIMA. For time series data such as load forecasting, a Long Short-Term Memory (LSTM) neural network, which is a type of Recurrent Neural Network (RNN), would be used. The advantage of using a LSTM model over ARIMA is that LSTM models have a significantly better ability to adapt to new conditions. As a result, LSTMs would be able to make more accurate predictions than ARIMA over longer forecasting periods because they are more robust to changes in the dataset. However, LSTMs require large amounts of data to train, so the feasibility of using this model would depend on the size of the forecasting dataset (18). Additionally, LSTMs require significantly more parameters in comparison to the *p, q,* and *d* parameters in ARIMA, so runtime would also be affected by tuning parameters, which is another disadvantage of converting to LSTMs. To maximize accuracy, variants of ARIMA should be used for short-term load forecasting and LSTMs for long-term load forecasting.

Overall, SARIMAX-GARCH can improve forecasting of energy supply and demand, help reduce redundancy in investments and improve energy security. With SARIMAX-GARCH, a power sector planner would be able to predict more accurately the consumer energy demand in the future. Without such accuracy, there could either be an overestimate or an underestimate of the actual demand in the future. If actual demand is more than the forecasted demand, then the planner must purchase the difference from the spot market at a higher energy cost and/or allow for shutdowns, either of which could have high-cost implications for the economy and potentially lead to energy security risk. On the other hand, if actual demand is less than the forecasted demand, any additional investments made in anticipation of higher actual demand in the future, such as building new capacities or locked-in long-term power purchase agreements, could incur redundant costs to the economy. For a 100 MW renewable energy plant, the estimated capital costs would be around $100 million. If the actual demand is lower, then it would render the newly built public investment redundant and become a sunk cost (19).

The results of our study show that the combination of SARIMAX and GARCH into SARIMAX-GARCH is a feasible model for load forecasting. However, its marginal increase in accuracy combined with the required resources to run several GARCH models and combining them with SARIMAX highlights that SARIMAX-GARCH could be further strengthened. By using other exogenous variables and ensemble techniques, we could reduce the error in the model. Additionally, other models such as LSTM can be considered for more accurate load forecasting. We hope that the results of this study will allow others to gain insight into combining time series models, specifically variations of ARIMA and variations of GARCH, and how energy generation companies and energy sector planners can understand the benefits and costs of using SARIMAX-GARCH.

MATERIALS AND METHODS

Dataset Processing

We first selected a load forecasting dataset to compare the effectiveness of SARIMAX-GARCH with SARIMAX. We used a Panama Kaggle dataset with load consumption data from 2015-2019 (10). In total, the Kaggle dataset contains 39072 rows, providing the models with sufficient data to analyze. In our research, all analysis on the Kaggle dataset was done in python.

Before analyzing the data with both models, we pre-processed the data to increase the efficiency and accuracy of the models. The Panama Kaggle dataset contains 12 columns, including the day of the week, holidays, and the temperature. First, we used the parser method in the dateutil package to set the index column as the date column in the pandas data frame. Next, we dropped all the columns in the pandas data frame except for the date index and the load measured at that date, as the remaining variables would be additional exogenous variables and SARIMAX only takes in one exogenous variable. We knew that 39,072 rows would take a significant amount of time to process, so we reduced the data to start from the year 2018. This decreased the number of rows to 11,135 and increased the relevancy of the output, as more recent data is passed through both models. By removing redundancies in the data, the shape of our data frame changed from 39702 rows and 12 columns to 11135 rows and 2 columns.

Analysis of Raw Data

After removing redundancies from the data, we plotted a graph of the dataset to identify trends and make initial assumptions (**Figure 2**). Our initial observations indicated that there was a slight overall trend upwards in energy load between 2018 and 2019, with a seasonal trend changing the load periodically. To confirm these trends, we performed an additive seasonal decomposition of the data to separate the actual trend, seasonal trend, and the residuals by using the seasonal_decomposition method in the statsmodels package (**Figures 3- 5**). We used the seasonal trend as the exogenous variable and the residuals were passed through the GARCH models.

Model Creation and Testing

Finally, we created both SARIMAX and SARIMAX-GARCH models and trained them on the load data. We first split the data into training data and testing data, in which the training data would be used to create and fit the model, and the testing data would be used to show the accuracy of each model. The training dataset consisted of 70% of the pre-processed Panama Kaggle data and the testing dataset consisted of the remaining 30%. To create both the SARIMAX and SARIMAX-GARCH models, we used the itertools package to create all possible combinations of *p*, *q*, and *d* parameters and seasonal *P*, Q, and *D* parameters, where the value for each parameter was either 0 or 1. Next, we created a SARIMAX model for every possible combination of these parameters, with the seasonal trend incorporated as an exogenous variable. To select the best model, we evaluated each model using AIC and BIC using the statsmodel library and chose the model with the lowest AIC and BIC. This meant that it had a high log likelihood, a measure of how well the model fits the data, to avoid underfitting and a small amount

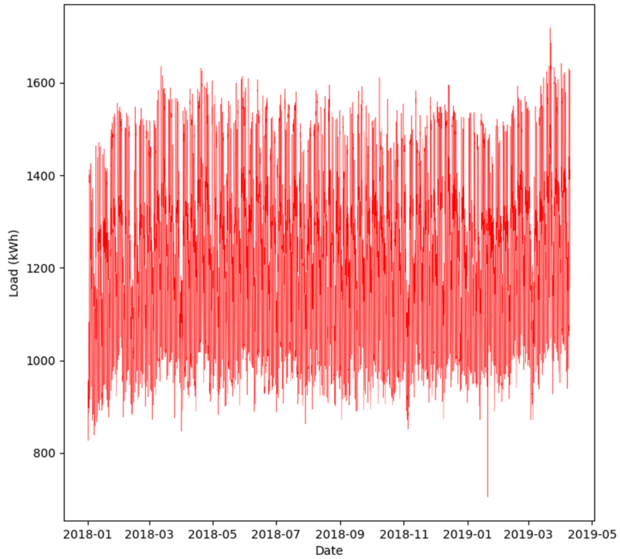

Figure 2: Raw load data of the Panama case study from 2018-2019. The load data from 2018 to 2019 from the Panama Kaggle dataset (10). The data was stored in a pandas dataframe after removing all features except time and load and plotted using

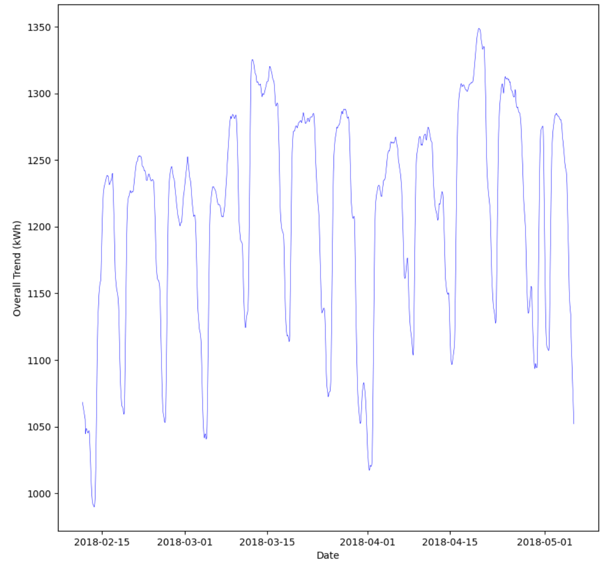

Figure 3: Overall base trend of the load dataset. The overall base trend of the load consumption data (kWh), excluding the seasonal pattern. There is a general sinusoidal trend to the data with the load increasing over time. The graph was generated with the python

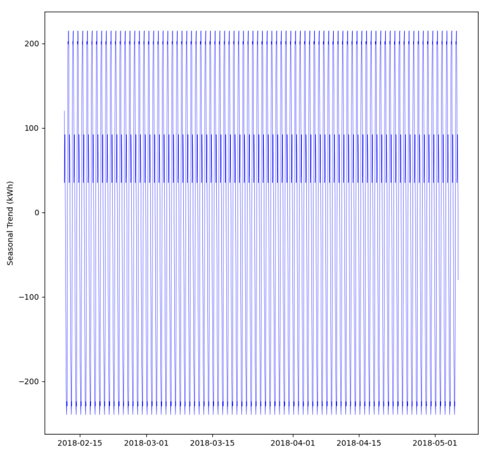

Figure 4: Seasonal trend of the dataset. The seasonal pattern of the data (kWh). This is used as the exogenous variable, as it allows us to see the impact of seasonality on the model.

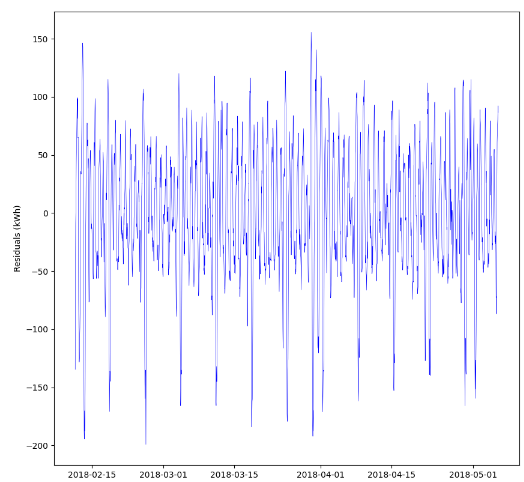

Figure 5: Residuals after performing a seasonal decomposition. The residuals from the seasonal decomposition. The residuals (kWh), which are the differences between the predicted value by the SARIMAX model and the observed value, are used in the GARCH models to combat conditional variance and improve the accuracy of the model.

of parameters to avoid overfitting. (20). Then, we fit the model on the training data and predicted the load using the model on the testing data. Our method to test the SARIMAX-GARCH model was similar to the method for the SARIMAX model, but we had to create each SARIMAX model and GARCH model separately and then combine them to test for AIC and BIC. The GARCH models had two parameters *p* and *q* generated by itertools in conjunction with SARIMAX and passed into each of the models. One key difference was that the GARCH model would pass in the residuals from the SARIMAX model as the data instead of values from the training dataset, as GARCH is used to measure heteroskedasticity, not to predict actual values. Each of the SARIMAX and SARIMAX-GARCH model's forecasted values were compared with the true values of the dataset by using MAPE.

Received: July 20, 2022
Accepted: September 11, 2022
Published: November 20, 2022

REFERENCES

1. Singh, Arunesh Kumar, et al. "Load Forecasting Techniques and Methodologies: A Review." 2012 2nd International Conference on Power, Control and Embedded Systems, Dec. 2012, doi:10.1109/ICPCES.2012.6508132.
2. Bond, Kingsmill, et al. "From Deep Crisis, Profound

Change." RMI, 7 Apr. 2022, rmi.org/insight/from-deep-crisis-profound-change/. Accessed 15 July 2022.
3. Castro, Rui, and João Crispim. "Variability and Correlation of Renewable Energy Sources in the Portuguese Electrical System." *Energy for Sustainable Development*, vol. 42, Feb. 2018, pp. 64-76, doi:10.1016/j.esd.2017.10.005.
4. USAID. "Variable Renewable Energy Generation Forecasting and Integration withDispatching." *USAID*, Feb. 2022, www.usaid.gov/energy/sure/vre-forecasting-webinar. Accessed 15 Apr. 2022.
5. Rotela Jr, Paulo, et al. "ARIMA: An Applied Time Series Forecasting Model for the Bovespa Stock Index." *Applied Mathematics,* vol. 05, no. 21, 2014, pp. 3383-91, doi:10.4236/am.2014.521315. Accessed 15 Apr. 2022.
6. Permanasari, Adhistya Erna, et al. "SARIMA (Seasonal ARIMA) Implementation on Time Series to Forecast the Number of Malaria Incidence." 2013 International Conference on Information Technology and Electrical Engineering (ICITEE), Oct. 2013, doi:10.1109/ICITEED.2013.6676239.
7. Perktold, Josef, et al. "Time Series Analysis by State Space Methods."Statsmodels, www.statsmodels.org/dev/statespace.html. Accessed 15 Apr. 2022.
8. Zhou, Bo, et al. "Traffic Modeling and Prediction Using ARIMA/GARCH Model."Modeling and Simulation Tools for Emerging Telecommunication Networks, pp. 101-21, doi:10.1007/0-387-34167-6_5.
9. Engle, Robert. "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics." *Journal of Economic Perspectives,* vol. 15, no. 4, 1 Nov. 2001, pp. 157-68, doi:10.1257/jep.15.4.157.
10. Madrid, Ernesto Anguilar. "Short-term electricity load forecasting (Panama case study)." Mendeley Data, doi.org/10.17632/byx7sztj59.1. Accessed 15 Apr. 2022.
11. Prabhakaran, Selva. "Augmented Dickey Fuller Test (ADF Test)." Machine Learning Plus, 2 Nov. 2019, www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/. Accessed 15 Apr. 2022.
12. Prabhakaran, Selva. "KPSS Test for Stationarity." Machine Learning Plus, 2 Nov. 2019, www.machinelearningplus.com/time-series/kpss-test-for-stationarity/. Accessed 15 Apr. 2022.
13. Kuster, Corentin, et al. "Electrical Load Forecasting Models: A Critical Systematic Review." *Sustainable Cities and Society*, vol. 35, Nov. 2017, pp. 257-70, doi:10.1016/j.scs.2017.08.009.
14. Schaffer, Andrea L., et al. "Interrupted Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models: A Guide for Evaluating Large-scale Health Interventions." *BMC Medical Research Methodology*, vol. 21, no. 1, 22 Mar. 2021, do:10.1186/s12874-021-01235-8.
15. Levendis, John D. "Non-stationarity and ARIMA(p,d,q) Processes." Springer Texts in Business and Economics, 2018, pp. 101-22, doi:10.1007/978-3-319-98282-3_5.
16. M. Xie, C. Sandels, K. Zhu and L. Nordström, "A seasonal ARIMA model with exogenous variables for elspot electricity prices in Sweden," 2013 10th International Conference on the European Energy Market (EEM), 2013, pp.1-4, doi: 10.1109/EEM.2013.6607293.
17. Agarwal, Rahul. "Using Gradient Boosting for Time Series Prediction Tasks."Towards Data Science, 17 Nov. 2019, towardsdatascience.com/using-gradient-boosting-for-time-series-prediction-tasks-600fac66a5fc. Accessed 15 Apr. 2022.
18. Siami-Namini, Sima, et al. "A Comparison of ARIMA and LSTM in Forecasting Time Series." 2018 17th IEEE International Conference on Machine Learning and Applications, doi:10.1109/ICMLA.2018.00227. Accessed 15 Apr. 2022.
19. IPIECA. "Combined Cycle Gas Turbines." IPIECA, 10 Apr. 2013, www.ipieca.org/resources/energy-efficiency-solutions/power-and-heat-generation/combined-cycle-gas-turbines/. Accessed 17 July 2022.
20. Kuha, Jouni. "AIC and BIC." Sociological Methods & Research 33 (2004): 188 - 229.