Forecasting Panama Electricity Load Using SARMAX

ECON 491 Term Project

University of Calgary

Saul Chirinos

2021-08-13

Abstract

Accurate forecasts of electricity demand are detrimental to grid operators. They are used for effective load management, scheduling, economic and business planning, and pricing, among others. The goal of this paper is to develop a statistical forecasting model to predict the last week of electricity load in Panama. The optimized model utilized follows a seasonal ARMAX(1, 2)x$(2, 2)_{24}$ order with temperature in degrees Celsius from David city as the exogenous regressor. To evaluate the fit and accuracy, model diagnostics were plotted, as well as the MAPE, MAE, MSE, and RMSE were calculated.

1 Introduction

Forecasting energy demand (load) is one of the main tools used by decision-makers in the energy sector. Studies have shown that demand forecasting plays an imperative role in power system management, economic and business planning, construction strategies, and pricing (Box et al., 2015; Elamin & Fukushige, 2018; Hu et al., 2013; Taylor & McSharry, 2007). Resources are wasted with costly expansion and planners are mislead if demand is overestimated (Son & Kim, 2017). However, failures and shortages are caused if demand is underestimated (Kavaklioglu et al., 2009). These miscalculations also has serious monetary costs towards businesses and can lead to a loss of millions of dollars (Bunn & Farmer, 1985;

Fan & Chen, 2006; Haida & Muto, 1994). In fact, Soares & Medeiros (2008) found that a one percent increase in forecast error increases operating costs by 10 million pounds per year for an electric utility in the United Kingdom. Therefore, during the last three decades there has been plenty of interest from leading statisticans around the world to increase accuracy (Caro et al., 2020). The time horizons of forecasting are usually short, medium, and long-term, however *very* short-term forecasting of minute-by-minute data is also studied (Taylor, 2008). The objective of short-term forecasting is to assist in the decision making of operations of a utility, economic dispatch, hydrothermal coordination, load management, and are typically in the ranges between hours to a few weeks (Al-Alawi & Islam, 1996). Medium-term forecasting is essential for planning fuel procurement, scheduling unit maintenance, energy trading and revenue assessment, and usually ranges between weeks to a few years (Al-Alawi & Islam, 1996). Finally, long-term forecasts are needed for decision making on the system generation and transmission expansion plans (Al-Alawi & Islam, 1996). As stated by Gonzalez-Romera et al. (2006), short-term load forecasting (STLF) has been the main focus point since it critical in the day-to-day operations of utility systems. However, due to the deregulation of energy markets, medium-term load forecasting (MTLF) has become essential as well (Hahn et al., 2009).

In forecasting, one must select relevant input variables that help explain the dependent variable in order to improve model accuracy. Preferably, feature selection should be based off of field expertise and variable correlation. Electricity demand is impacted by several weather, socio-economic, and demographic variables such as temperature, wind speed, humidity, gross domestic product (GDP), and population (Al-Alawi & Islam, 1996). Of all possible variables,

temperature and humidity are the most common in load forecasting (Arora & Taylor, 2013). Additionally, the forecasting time horizon is an indicator on deciding which input features to include and which to exclude from the model(s). In terms of STLF, weather factors highly impact demand, however social and economic factors have no impact since they have more medium to long-term effects (Al-Alawi & Islam, 1996). Upon completion of feature selection the forecaster must choose what error metrics to use in order to judge the accuracy of the model(s). For many forecasting studies, the mean absolute percentage error (MAPE) is the most used evaluation metric since the proportionality between the forecast error and the actual demand is captured, and because of its clear interpretability (Hippert et al., 2001). Other popular measurements include the mean absolute error (MAE), root mean squared error (RMSE), and mean squared error (MSE) (Camara et al., 2016; Caro et al., 2020; Dash & Dash, 2019; Taylor & McSharry, 2007). Equations for these metrics can be seen in Table 5.

This paper is divided into the following sections: Section 2 discusses methods and findings of other research papers along with their results; Section 3 explains the economic theory behind the ARIMA family; Section 4 describes the data used for forecasting; Section 5 explains the processes used to estimate the order of the model; Section 6 summarizes the results obtained from estimation and the metrics of accuracy; and Section 7 concludes the findings of the paper.

2 Literature Review

Ever since forecasting became known as a tool used for predicting the future, many techniques have been developed and are still being researched in both classical time series methods

(Bunn & Farmer, 1985; Dordonnat et al., 2008; Huang & Shih, 2003), and machine learning intelligence (Al-Musaylh et al., 2019; Cai et al., 2019; Hippert et al., 2001; Taylor et al., 2006). Examples of classical time series methods include seasonal autoregressive integrated moving average with exogenous regressors (SARIMAX) which can be broken down into simple AR, MA, or ARMA models, simple exponential smoothing (SES), and vector autoregressive moving average (VARMA). In fact, the research of George Box and Gwilym Jenkins originally popularized the ARIMA model (Box et al., 2015). In forecasting with machine learning models, there exists k-nearest neighbor (KNN) regression, support vector regression (SVR), and classification and regression tree (CART) to name a few (Ahmed et al., 2010).

Elamin & Fukushige (2018) estimated hourly load in a region of Japan through seasonal autoregressive integrated moving average with exogenous regressors (SARIMAX) containing only main effects and a SARIMAX with main and interaction effects. Both models are compared with an interaction based multiple linear regression (MLR) benchmark. The SARIMAX with main and interaction effects produces more accurate results than its competitors since the interaction effects in addition to the main effects aided in the detection of primary demand drivers. The model achieved a MAPE of 0.7, MAE of 21.7, and RMSE of 31.6. Another study by Kim (2013) estimated hourly demand in Korea with a double SARMA model with interactions in order to seize intraday and intraweek autocorrelations of consumer demand. The proposed model was benchmarked against several other models (SES model, SARMA model, 2 SARMAX models). The estimates demonstrated the proposed model outperforms all benchmark models in both the total forecasting and special-day forecasting period when evaluated on the RMSEs and MAPEs. Camara et al. (2016) modeled quarterly residential

consumption data from the U.S. using SARIMA with the Box-Jenkins methodology and an Artificial Neural Network (ANN) approach. Both models were evaluated on training and testing data with success metrics consisting of MAPE, MAE, and MSE. The results showed that both model performances were relatively similar, with the former having a slightly lower MAPE, MAE, and MSE for the training data and a lower MSE for the testing data. Sigauke & Chikobvu (2011) utilized purely classical time series methods. Short-term load forecasts were predicted for daily electricity demand in South Africa using SARIMA, SARIMA with generalized autoregressive conditional heteroskedastic errors (SARIMA-GARCH), and regression-SARIMA-GARCH models. The use of the GARCH is to account for possible heteroskedasticity since the data seemed to not possess a constant variance, thus violating the assumption of homoskedasticity (Sigauke & Chikobvu, 2011). Additionally, each model performance is evaluated on a benchmark piecewise linear regression model. Final model estimations conclude that the Reg-SARIMA-GARCH model outperformed its competitors with a MAPE of 1.42 and a RMSE of 554. The SARIMA and SARIMA-GARCH models resulted in MAPEs of 1.47 and 1.43, and RMSEs of 571 and 556, respectively. The piecewise linear regression performed the worst.

3 Economic Theory

The model of choice for this paper follows a $SARIMAX(p,d,q) \times (P,D,Q)_s$ order and can be found at the end of Section 5. This section delves deeper on SARIMAX models and their mathematical notations for better understanding. Beginning with an AR model of order p

with drift (intercept), a series y_t can be expressed as

$$y_t = c + a_1 y_{t-1} + ... + a_p y_{t-p} + \epsilon_t, \tag{1}$$

where $a_1, ..., a_p$ are the AR parameters, $y_{t-1}, ..., y_{t-p}$ is the series lagged by one up to p periods, and ϵ_t is the error at time t. By contrast, an MA model of order q with drift can be expressed as

$$y_t = c + \epsilon_t + m_1 \epsilon_{t-1} + ... + m_p \epsilon_{t-q}, \tag{2}$$

where $m_1, ..., m_p$ are the MA parameters, $\epsilon_{t-1}, ..., \epsilon_{t-q}$ are the error terms lagged by one up to q periods, and ϵ_t is the error at time t. Integrating both models into one creates an $ARMA(p, q)$ model as shown below.

$$y_t = c + a_1 y_{t-1} + ... + a_p y_{t-p} + \epsilon_t + m_1 \epsilon_{t-1} + ... + m_p \epsilon_{t-q} \tag{3}$$

The autocorrelation and partial autocorrelation functions are used in estimating the q and p order, respectively. The I in $ARIMA$ comes from the order of differencing required to make a series stationary. Extensive research has been applied in this area of time series analysis (Dickey & Fuller, 1979; Kwiatkowski et al., 1992; Perron, 1988). The most common method of detecting stationarity in a series is the augmented Dicky-Fuller (ADF) test which tests the null hypothesis that the series possesses a unit root (Dickey & Fuller, 1979).

$$y_t = \rho y_{t-1} + \epsilon_t, t = 1, 2, ..., T \tag{4}$$

If $|\rho| \geq 1$, then the series possesses a unit root, i.e. is not stationary and differencing is required. The second, less popular, stationarity detection method is the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test which tests the null hypothesis that the series *is* stationary (Kwiatkowski et al., 1992).

$$y_t = \xi t + r_t + \epsilon_t \tag{5}$$

Where r_t represents a random walk:

$$r_t = r_{t-1} + \epsilon_t, \tag{6}$$

and the ϵ_t are independent and indentically distributed (iid) with mean of zero and constant variance $(0, \sigma_\epsilon^2)$. The series will be stationary around a fixed level or have a fixed intercept if the data is stationary (Wang, 2006). Usually, most forecasting studies test for stationarity with the methodology of Dicky and Fuller. However, Kwiatkowski et al. (1992) argued that this is an issue. Originally, Perron (1988) tested the research of Dicky and Fuller by conducting the ADF test on 14 annual U.S. time series data. Every level series failed to reject the null hypothesis of a unit root except for one. This suggests that the normalized tests for a unit root are not strong enough against the alternatives and it is advised that both the unit root hypothesis and stationarity hypothesis is performed (Kwiatkowski et al., 1992). Having both results allows us to determine whether the level series is stationary more confidently. If both tests give contradicting results, then the data is not sufficiently informative to know whether it is integrated or not (Kwiatkowski et al., 1992).

Now with a clear understanding of $ARIMA(p, d, q)$ models, we can begin adding seasonal

affects for $SARIMA(p,d,q) \times (P,D,Q)_s$ models. Here, p is the AR order, d is the order of differencing, q is the MA order, P is the seasonal AR order, D is the seasonal differencing order, Q is the seasonal MA order, and s is the number of seasonal periods. An expression of this model is shown below:

$$y_t = c + \frac{(1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)(1 - \Theta_1 B^s - \Theta_2 B^{2s} - ... - \Theta_Q B^{Qs})}{(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)(1 - \Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_P B^{Ps})} \epsilon_t, \tag{7}$$

where $\phi_1, \phi_2, ...\phi_p$ are the AR weights, $\theta_1, \theta_2, ...\theta_q$ are the MA weights, $\Phi_1, \Phi_2, ...\Phi_P$ are the seasonal AR weights, $\Theta_1, \Theta_2, ...\Theta_Q$ are the seasonal MA weights, and B denotes the backshift operator. For example, $By_t = y_{t-1}$ or $B^s y_t = y_{t-s}$.

The inclusion of exogenous variables for a $SARIMAX$ model does not complicate notation:

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{i,t} + \frac{(1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)(1 - \Theta_1 B^s - \Theta_2 B^{2s} - ... - \Theta_Q B^{Qs})}{(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)(1 - \Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_P B^{Ps})} \epsilon_t, \tag{8}$$

where $\sum_{i=1}^{k} \beta_i X_{i,t}$ denotes the coefficient and exogenous variable from one up to k variables. Both SARIMA and SARIMAX model notations were referenced from Cools et al. (2009), with the former being the same as the latter, only without the exogenous variables included.

4 Data

The data demonstrated in this paper originates from Aguilar Madrid & Antonio (2021) who exercised a machine learning model in forecasting. Variables provided include historical hourly electricity load and weekly forecasts in Panama provided by the grid operator (CND),

calendar information relating to school periods provided by Panama's Ministery of Education, calendar information regarding holidays provided by "When on Earth?" website, and weather variables for three main cities provided by Earthdata. For a more detailed description, see Table 1. The historical hourly load begins on January 3, 2015 at 1:00 AM and ends on June 27, 2020 at 12:00 AM making a total of 48,048 observations recorded. Figure 1 depicts yearly, monthly in the year 2015, daily on the month of January, 2015, and hourly on January 3, 2015 demand plots. The choice of time frames were selected arbitrarily. Furthermore, the weekly forecasts provided by CND begin on January 2, 2016 at 12:00 AM and end on July 31, 2020 at 11:00 PM.

Observing the demand plots, they show a consistent seasonal fluctuation repeating every 24 hours and has a slight increasing long-run trend. Interestingly, Figure 2 presents two sharp drops in electricity load in the middle of 2017 and early 2019. Demand decreases to around 1,250 MWh in early 2020 and picks up again which is most likely due to the effect of COVID-19 and Panama's policy responses. As a result, the variance does not seem constant given the trend and shocks. Taking a closer look at the hourly graph, we notice demand reach a global minimum and maximum around 6:00 AM and 10:00 AM, respectively, after which electricity load begins to slowly decrease and then increase again to a local maximum around 8:00 PM, after which begins decreasing again.

In the next section, a discussion on model building and the tools utilized is spoken about. The data period used for the model building and modeling process begins on June 1, 2020 at 12:00 AM up to the final recorded observation on June 27, 2020 at 12:00 AM making 625 total observations. The data was filtered for a shorter time frame since many SARIMAX models

were too time consuming to fit and check diagnostics on the original 48,048 observations. The subset was then split into training and testing sets in which the latter begins on June 20, 2020 at 1:00 AM, making 457 hours of training samples and 168 hours of testing samples.

5 Model

In order to estimate the order of the model, the autocorrelation and partial autocorrelation functions in Figure 3 are computed on both the level (top) and differenced (bottom) series. Both functions are similar in that they measure the correlation between the period today and its previous periods. The latter however controls for the lagged values. Referring to Figure 3, in both the ACF and PACF it is noticed that there is clear evidence of daily seasonality, more so in the differenced series, as seen in lags 24 and 48. This suggests a seasonal autoregressive and moving average order. Additional inspection of the seasonal component for the level and differenced series is presented in Figure 4 and Figure 5, respectively. The time frame was arbitrarily selected. Further, The ACF in both series seems to be oscillating whereas the PACF cuts off after the first or second lag suggesting a non-seasonal autoregressive process with no moving average terms.

In terms of stationarity, visually it is inferred that demand is non-stationary as shown in Figure 1. However, the ADF and KPSS tests were conducted on the level series to determine its stationary characteristic. Unfortunately, both tests reject the null hypothesis and therefore give contradicting conclusions. The former resulted in a test statistic of -20.96 and a p-value significantly less than 0.05, thus rejecting the null hypothesis that the series possesses a unit root and is therefore trend-stationary. The latter returned a test statistic of 12.67 and a

p-value of 0.01 which is less than the 0.05 significance level. Thus, we reject the null hypothesis that the series is stationary. However, after taking first differences the KPSS fails to reject the null hypothesis. As a result, due to the difficulty in estimating the differencing order, the model is estimated through an iterative process using algorithmic automation of model selection and the Box-Jenkins methodology. The former operates on minimizing the Akaike Information Criterion (AIC) based on permutations and combinations of model orders. The latter uses an iterative three step process: model identification, estimation, and diagnostic checking (Caro et al., 2020). Furthermore, while univariate ARIMA models are sufficient for STLF, many studies suggest the inclusion of relevant input variables increases forecasting accuracy (Soares & Medeiros, 2008; Taylor et al., 2006). The exogenous variables $T2M\_dav$, $QV2M\_toc$, and $W2M\_dav$ were selected based off of support from the literature, the correlation matrix, and a Random Forest regression to obtain feature importance, presented in Figure 7 and Figure 8, respectively. The correlation matrix was calculated based off the Spearman method, rather than the more well known Pearson correlation method. The reason for this being that it has been found that the relationship between electricity load and weather variables is non-linear (Al-Alawi & Islam, 1996; Caro et al., 2020; Elamin & Fukushige, 2018; Hippert et al., 2001; Taylor, 2010), thus making Spearman a more appropriate choice. In addition, a natural log transformation was applied to the national load demand since the standard deviation, as seen in Table 3, is extremely high relative to other variables. This transformation resulted in a drastic increase in model performace.

After various model combinations and diagnostic checks, the optimized model to be estimated was a $SARIMAX(1,0,2) \times (2,0,2)_{24}$ with $T2M\_dav$ as the only exogenous regressor. The

mathematical expression is shown below:

$$lnat\_demand_t = \beta_0 + \beta_1 T2M\_dav_t + \frac{(1-\theta_1 B - \theta_2 B^2)(1-\Theta_1 B^{24} - \Theta_2 B^{48})}{(1-\phi_1 B)(1-\Phi_1 B^{24} - \Phi_2 B^{48})}\epsilon_t, \qquad (9)$$

where $lnat\_demand_t$ is the log transform of national electricity load (MWh) at time t, β_0 is the intercept, $T2M\_dav_t$ is the temperature in degrees Celsius at 2 meters in David city at time t, $(1-\phi_1 B)$ is the AR lag of order one, $(1-\theta_1 B - \theta_2 B^2)$ is the MA lag of order two, $(1-\Phi_1 B^{24} - \Phi_2 B^{48})$ is the seasonal AR lag of order two, $(1-\Theta_1 B^{24} - \Theta_2 B^{48})$ is the seasonal MA lag of order two, and ϵ_t is the error term at time t. Given that no differencing was pursued on both the non-seasonal and seasonal ends, the model can be simplified to a $SARMAX(1,2) \times (2,2)_{24}$.

6 Empirical Results

The estimated coefficients from the fitted model on the last month of data are shown below:

$$lnat\_demand_t = .0334 + .0169\, T2M\_dav_t + \frac{(1-.4141B - .1326B^2)(1+.2583B^{24} + .3348B^{48})}{(1-.8761B)(1-.5258B^{24} - .4331B^{48})}\epsilon_t \qquad (10)$$

A more detailed summary including standard deviations, z-statistics and p-values is presented in Table 3. The results show that the temperature in David city, a one hour AR lag, a one and two hour MA lag, and a 24 and 48 hour seasonal AR lag have a positive relationship with electricity load. In contrast, the 24 and 48 hour seasonal MA lags have a negative relationship with demand. All but the intercept, and two seasonal AR and MA lags are statistically significant since they have p-values less than 0.05. This is also evident from the

magnitude of the standard errors. Furthermore, the generated AIC and BIC were -2341.962 and -2300.716, respectively. As mentioned in Section 5, the exogenous variables $T2M\_dav$, $QV2M\_toc$, and $W2M\_dav$ were selected to begin model fitting and diagnostic checking. Throughout the process, it became evident that the inclusion of $QV2M\_toc$ and $W2M\_dav$ hindered model performance and thus were eliminated.

In addition, model diagnostics can be seen in Figure 9. Taking a look at the standardized residuals, the residuals seem to be white noise. That is, there is no observable pattern in the residuals. The histogram almost follows a normal distribution, as outlined by the kernal density estimation (KDE), however a statistical test of residual normality can be determined through the Jarque-Bera (JB) test. The p-value returned is 0.00, thus rejecting the null hypothesis that the residuals are normally distributed. The Q-Q plot however shows promising results since most of the residuals follow a straight line. This can also be statistically supported with the Ljung-Box test which determines whether there are no correlations in the residuals. The p-value generated is 0.90. Thus, we fail to reject the null hypothesis that the residuals are uncorrelated. Both the JB and Ljung-Box test statistics with their associated p-values can be seen in Table 4. Finally, the correlogram shows the lagged values in the residuals up to ten periods. Lags 3 and 9 seem to be statistically significant. This observation, along with the non-normally distributed residuals, suggests the model can still be improved further.

Forecasts are made in-sample from the beginning of June 20, 2020 at 1:00 AM to June 27 at 12:00 AM in order to visually observe model accuracy, see Figure 10. Recall that a log transform was applied to the electricity demand, however, an exponential transformation

was used to revert load and predictions back to its original scale for plotting. Refering to Figure 10, the model seems to capture the sesonal affects as well as accurately predict the drop in demand each day. However, the large peaks in demand were not predicted well. The metrics of error selected to evaluate model performance follow the literature, with a MAPE of 0.053, MAE of 59.75, MSE of 5942.69, and RMSE of 77.09. As mentioned in Section 4, weekly forecasts were produced by CND. A comparision between the forecasts produced in this paper and the forecasts made by CND is presented in Figure 11.

7 Conclusion

This paper estimated short-term electricity demand in Panama using data filtered to the last month. The reason load forecasting is critical in firm management is because millions of dollars are lost with even the slightest error. In other words, the firms monetary value can be retained if forecast error is reduced. Additionally, scheduling, load flow analysis, power system security, and load shifting from one substation to another is made more effective with accurate forecasting (Sigauke & Chikobvu, 2011). To forecast one week of in-sample load demand in Panama, a $SARMAX(1,2) \times (2,2)_{24}$ model was utilized. The model was selected based on an AIC minimization algorithm and the Box-Jenkins methodology. Evaluation metrics consisted of MAPE MAE, MSE, and RMSE. Of which resulted in 0.053, 59.75, 5942.69, and 77.09, respectively. Finally, the model diagnostics generated in Figure 9 suggest the model could be improved with further experimentation and research.

8 Tables

Table 1: Variable Descriptions

	Description
datetime	Date-time corresponding to Panama time-zone UTC-05:00
nat_demand	National electricity load (MWh)
T2M_toc	Temperature at 2 meters in Tocumen, Panama city (ºC)
QV2M_toc	Relative humidity at 2 meters in Tocumen, Panama city (%)
TQL_toc	Liquid precipitation in Tocumen, Panama city (l/m2)
W2M_toc	Wind Speed at 2 meters in Tocumen, Panama city (m/s)
T2M_san	Temperature at 2 meters in Santiago city (ºC)
QV2M_san	Relative humidity at 2 meters in Santiago city (%)
TQL_san	Liquid precipitation in Santiago city (l/m2)
W2M_san	Wind Speed at 2 meters in Santiago city (m/s)
T2M_dav	Temperature at 2 meters in David city (ºC)
QV2M_dav	Relative humidity at 2 meters in David city (%)
TQL_dav	Liquid precipitation in David city (l/m2)
W2M_dav	Wind Speed at 2 meters in David city (m/2)
Holiday_ID	Unique identification number
holiday	Holiday binary indicator
school	School period binary indicator

Table 2: Summary Statistics

	Mean	Standard Deviation	Minimum	25%	50%	75%	Maximum
nat_demand	1182.87	192.07	85.19	1020.06	1168.43	1327.56	1754.88
T2M_toc	27.40	1.68	22.95	26.16	27.12	28.56	35.04
QV2M_toc	0.02	0.00	0.01	0.02	0.02	0.02	0.02
TQL_toc	0.08	0.07	0.00	0.03	0.07	0.12	0.52
W2M_toc	13.39	7.30	0.01	7.54	12.18	18.66	39.23
T2M_san	26.92	3.02	19.77	24.77	26.17	28.71	39.06
QV2M_san	0.02	0.00	0.01	0.02	0.02	0.02	0.02
TQL_san	0.11	0.09	0.00	0.04	0.09	0.16	0.48
W2M_san	7.05	4.10	0.06	3.96	5.99	9.41	24.48
T2M_dav	24.72	2.41	19.93	22.95	24.00	26.24	34.22
QV2M_dav	0.02	0.00	0.01	0.02	0.02	0.02	0.02
TQL_dav	0.14	0.09	0.00	0.08	0.13	0.20	0.48
W2M_dav	3.57	1.71	0.02	2.30	3.41	4.67	10.29
holiday	0.06	NA	NA	NA	NA	NA	NA
school	0.73	NA	NA	NA	NA	NA	NA

Table 3: Summary Results from $SARMAX(1,2)\times(2,2)_{24}$

	Coefficient	Standard Error	z	P>|z|
intercept	0.0334	0.058	0.573	0.566
T2M_dav	0.0169	0.005	3.242	0.001
ar.L1	0.8761	0.025	34.468	0.000
ma.L1	0.4141	0.046	9.081	0.000
ma.L2	0.1326	0.046	2.905	0.004
ar.S.L24	0.5258	2.297	0.229	0.819
ar.S.L48	0.4331	2.226	0.195	0.846
ma.S.L24	-0.2583	2.298	-0.112	0.910
ma.S.L48	-0.3348	1.617	-0.207	0.836

Table 4: Ljung-Box and Jarque-Bera Test Statistics with P-values

	Test Statistic	P-value
Ljung-Box	0.02	0.90
Jarque-Bera	24.78	0.00

Table 5: Error Metrics

Error Type	Formula
Mean Absolute Percentage Error (MAPE)	$\frac{100}{n}\sum_{t=1}^{n}\lvert\frac{y_t-\hat{y}_t}{y_t}\rvert$
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{t=1}^{n}\lvert y_t-\hat{y}_t\rvert$
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}$
Mean Squared Error (MSE)	$\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2$

9 Figures

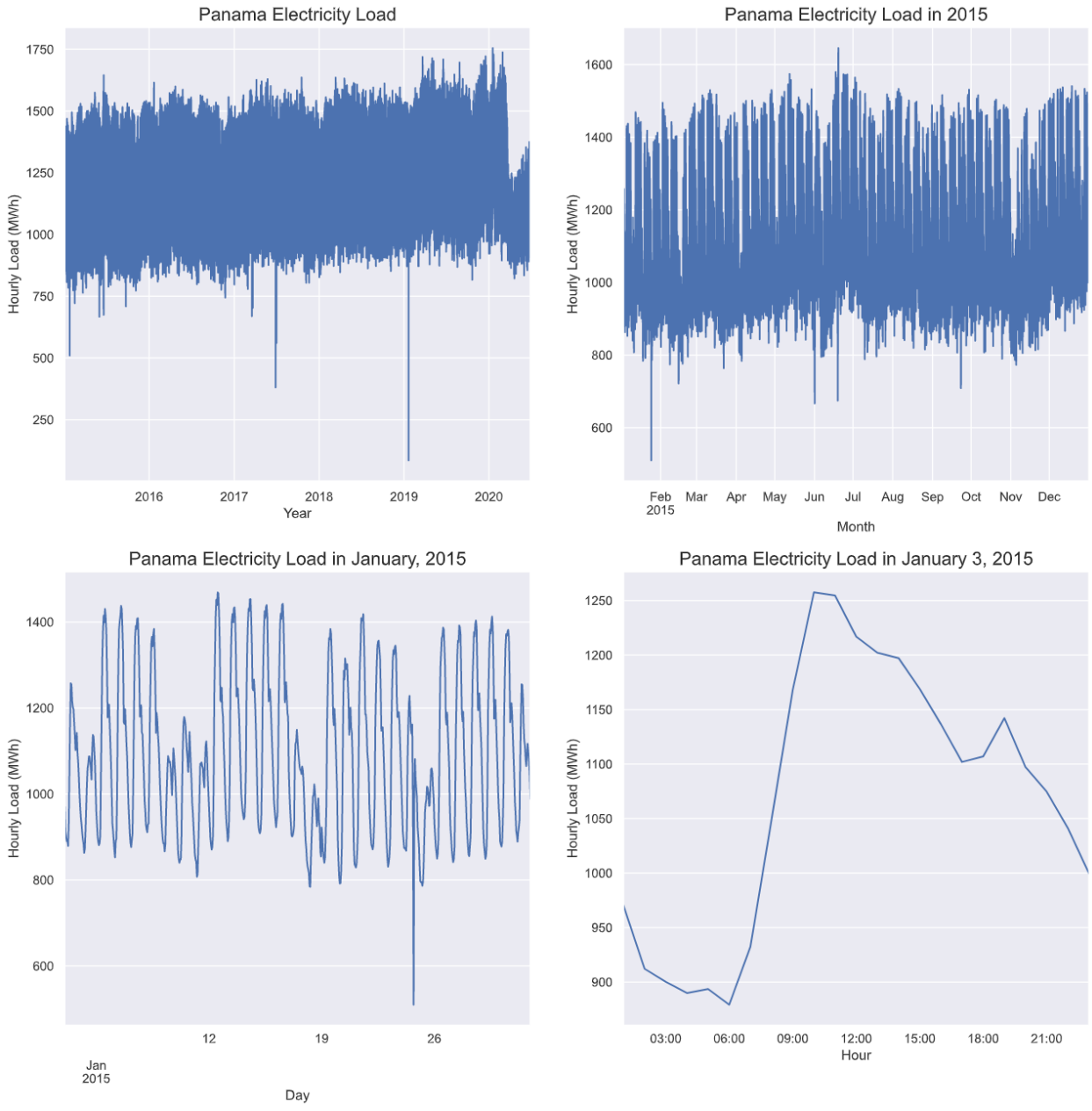

Figure 1: Electricity Demand in Different Time Horizons

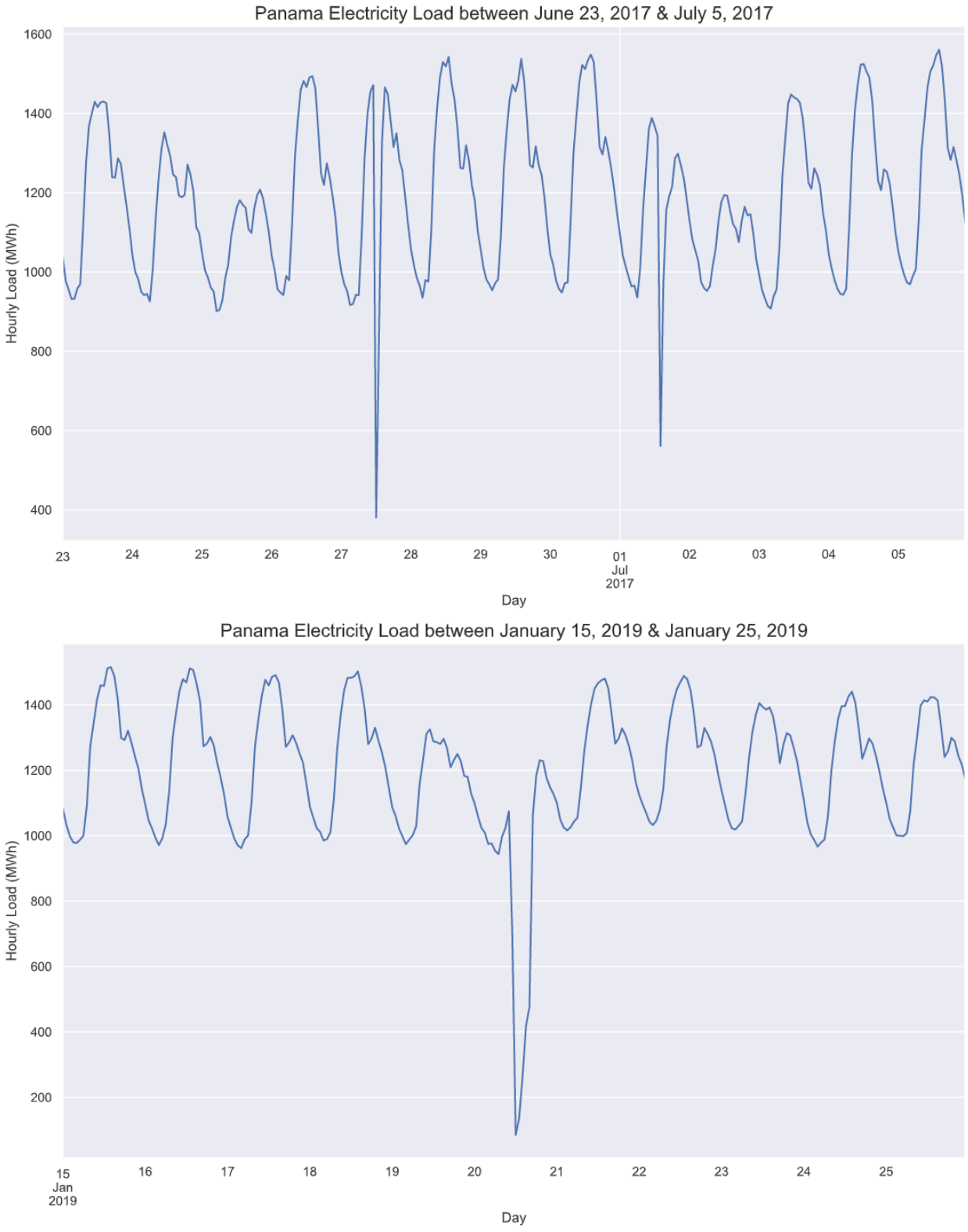

Figure 2: Magnification on Demand Drops

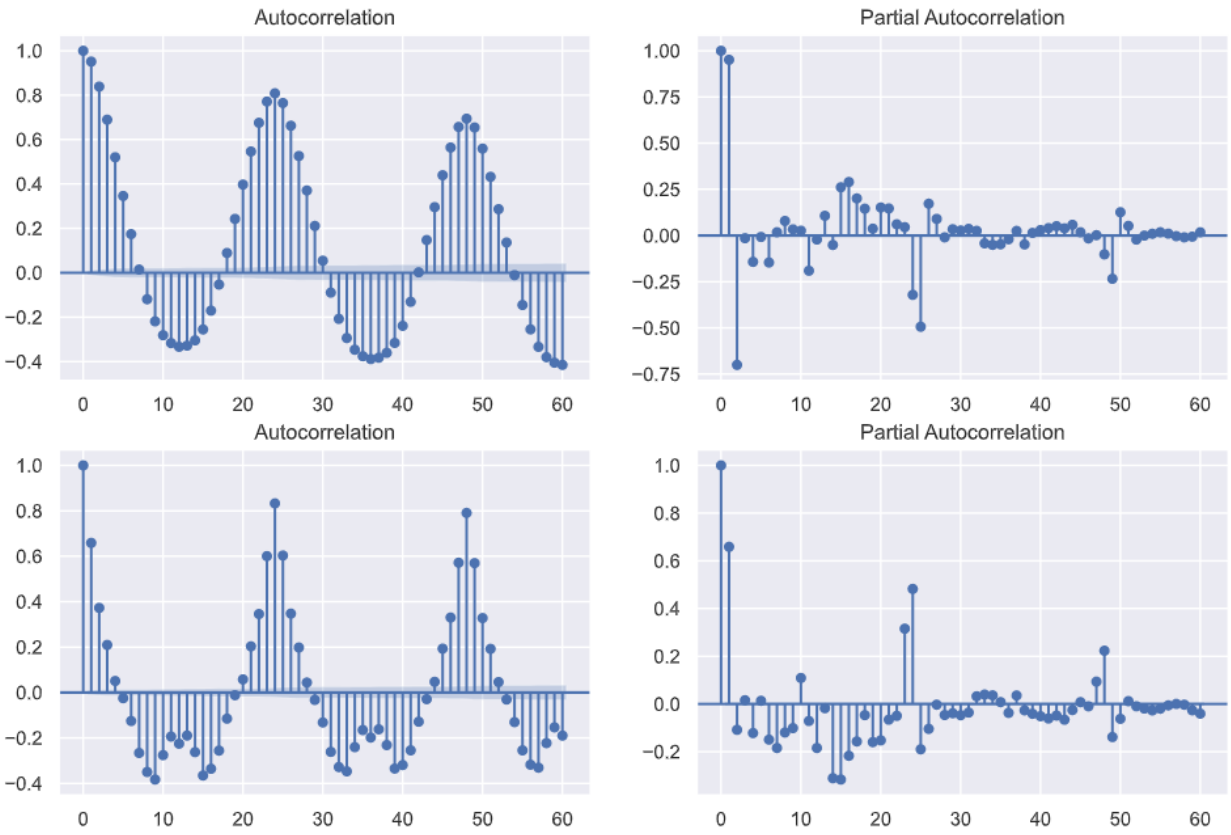

Figure 3: ACF and PACF for the Level (top) and Differenced (bottom) Series

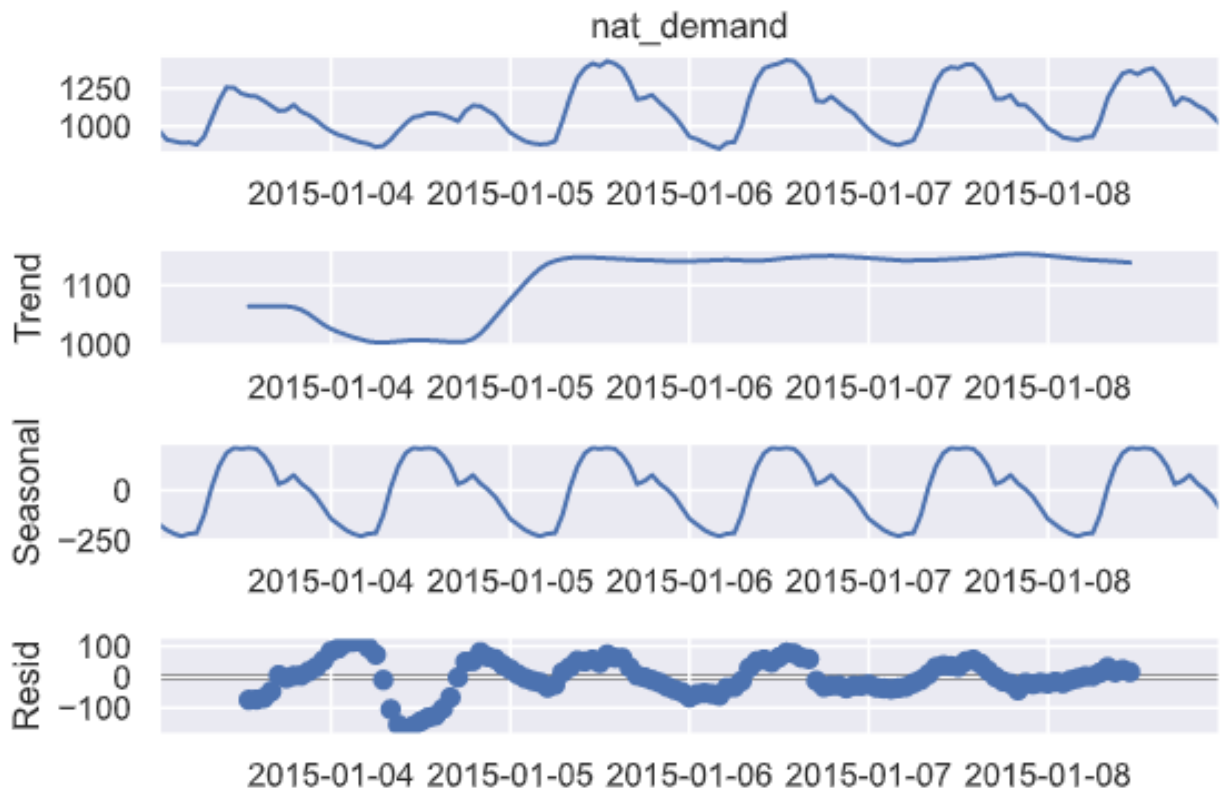

Figure 4: Seasonal Decomposition of Level Series between January 3, 2015 to January 8, 2015

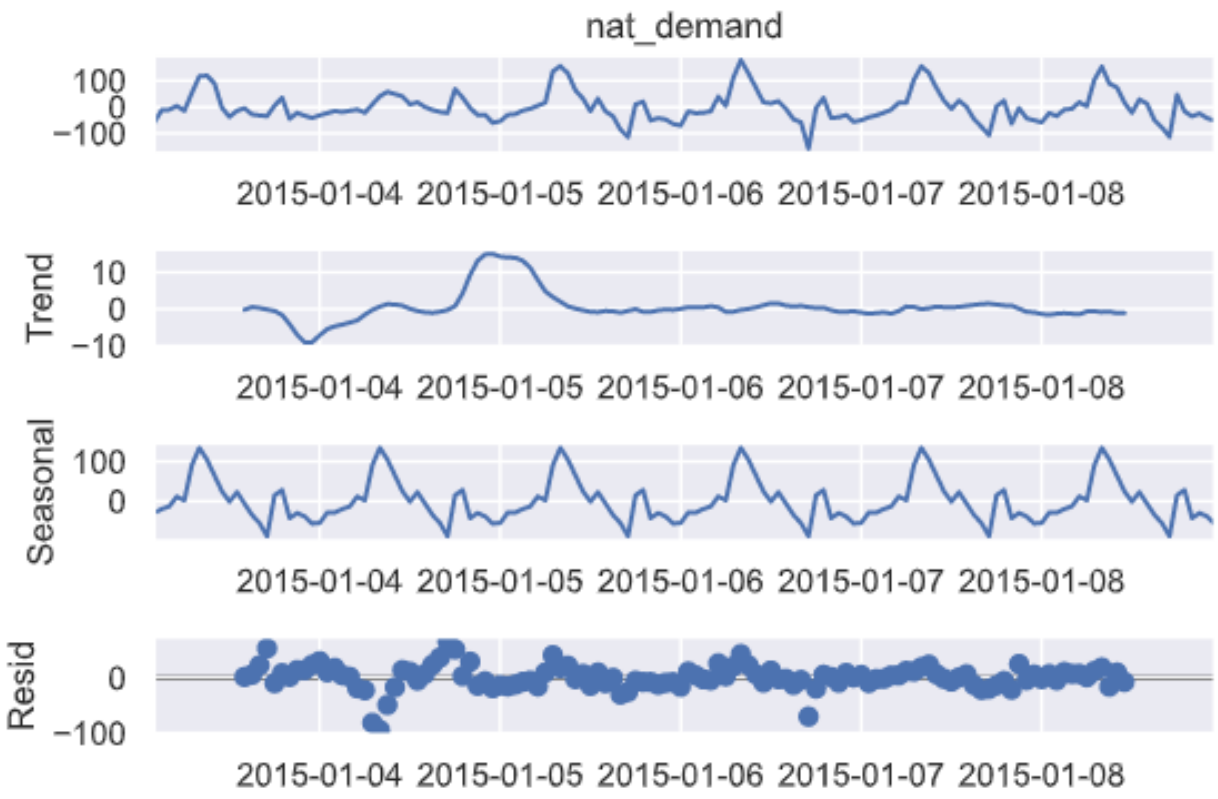

Figure 5: Seasonal Decomposition of Differenced Series between January 3, 2015 to January 8, 2015

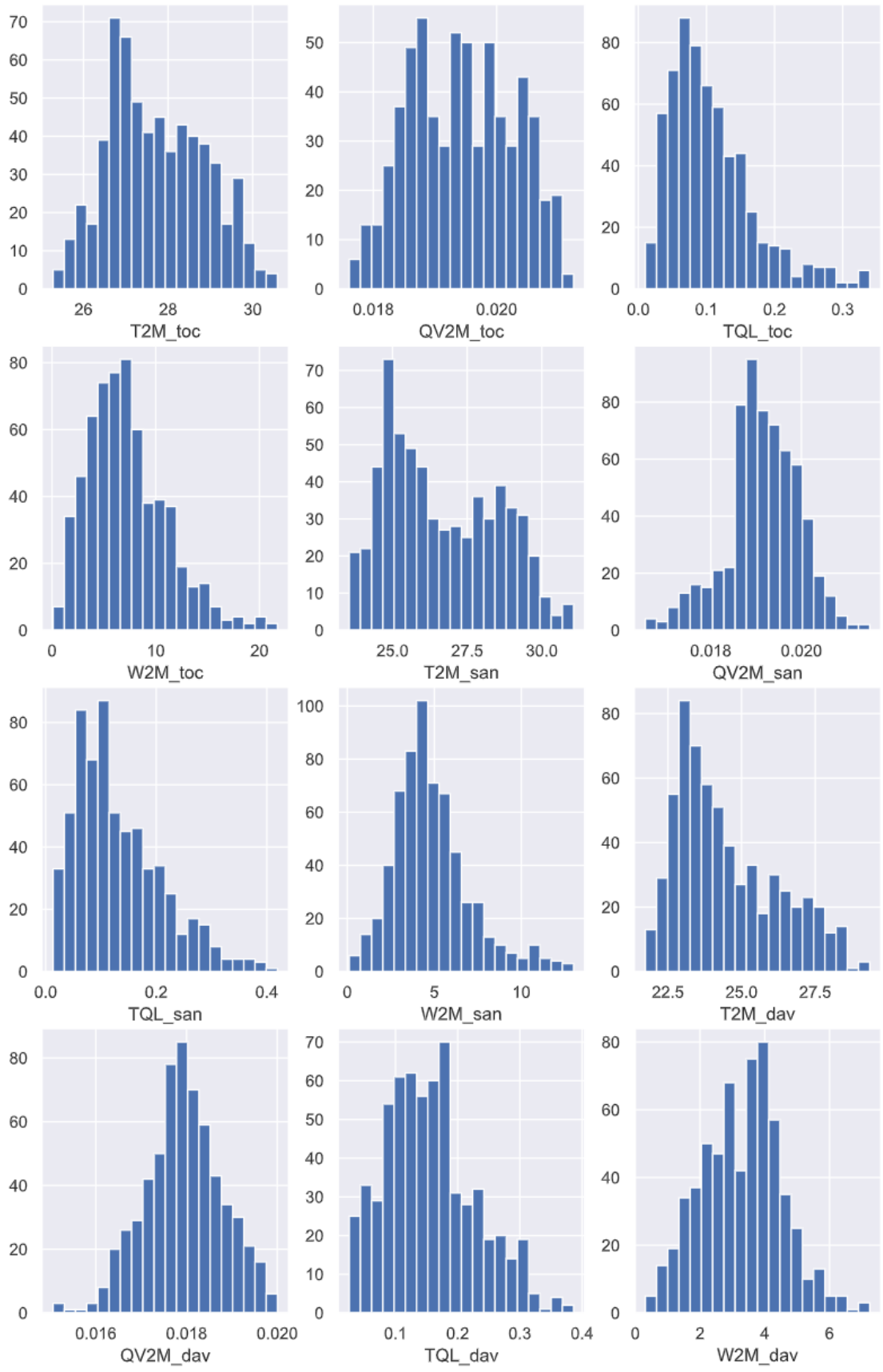

Figure 6: Distributions of Continuous Variables

Figure 7: Correlation Matrix

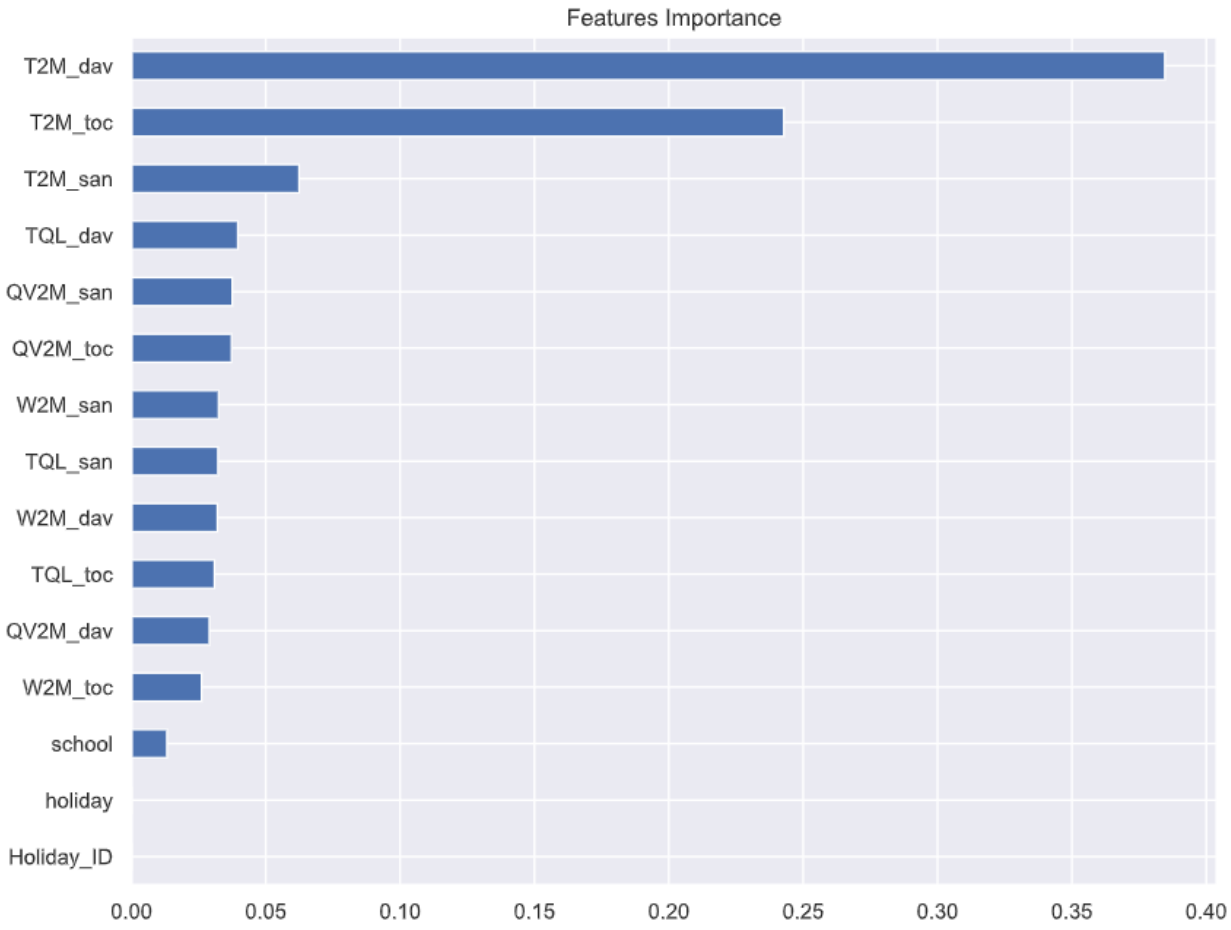

Figure 8: Feature Importance from Random Forest Regressor

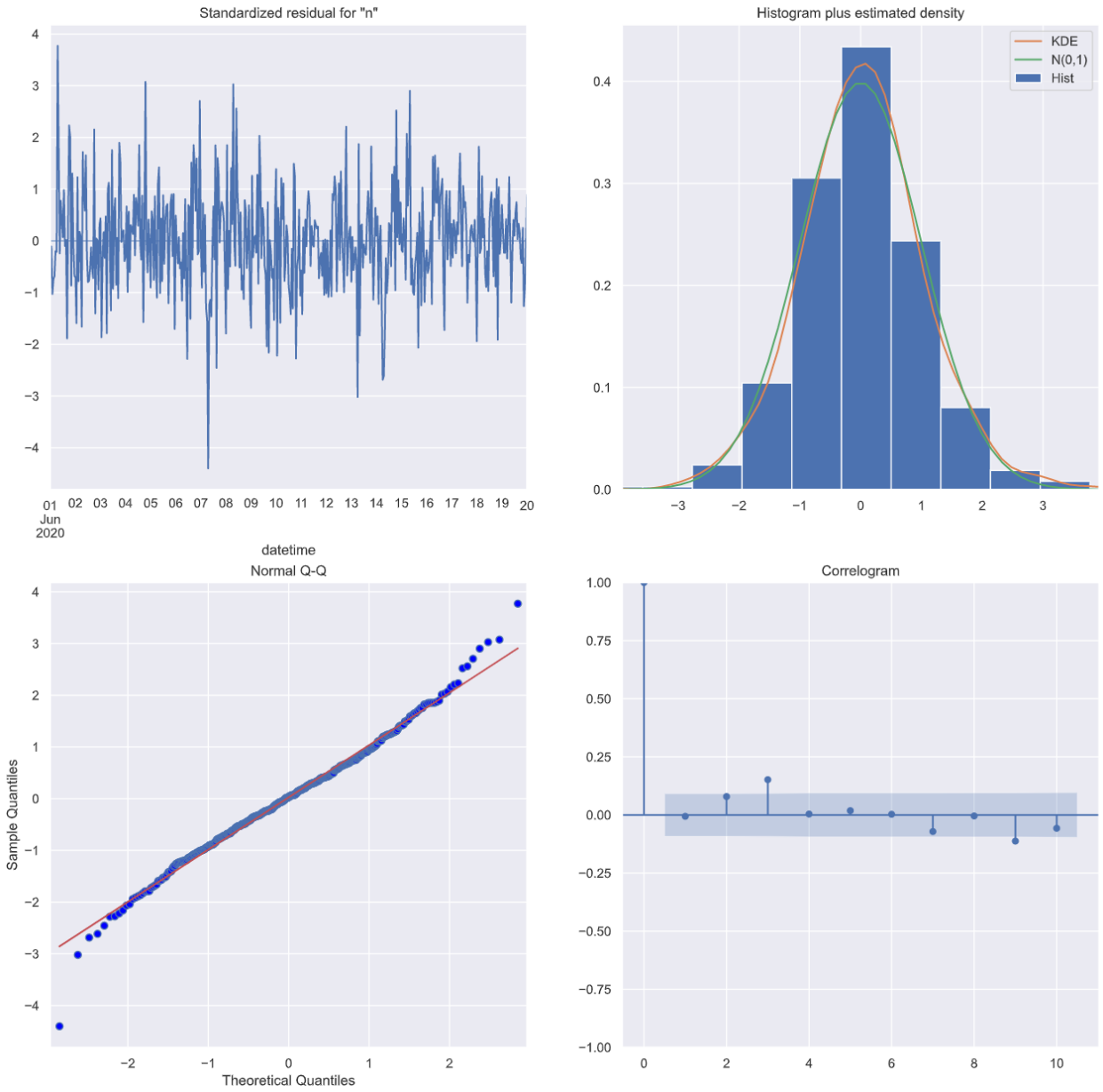

Figure 9: $SARMAX(1,2) \times (2,2)_{24}$ Diagnostics with Exogenous Variable $T2M\_dav$

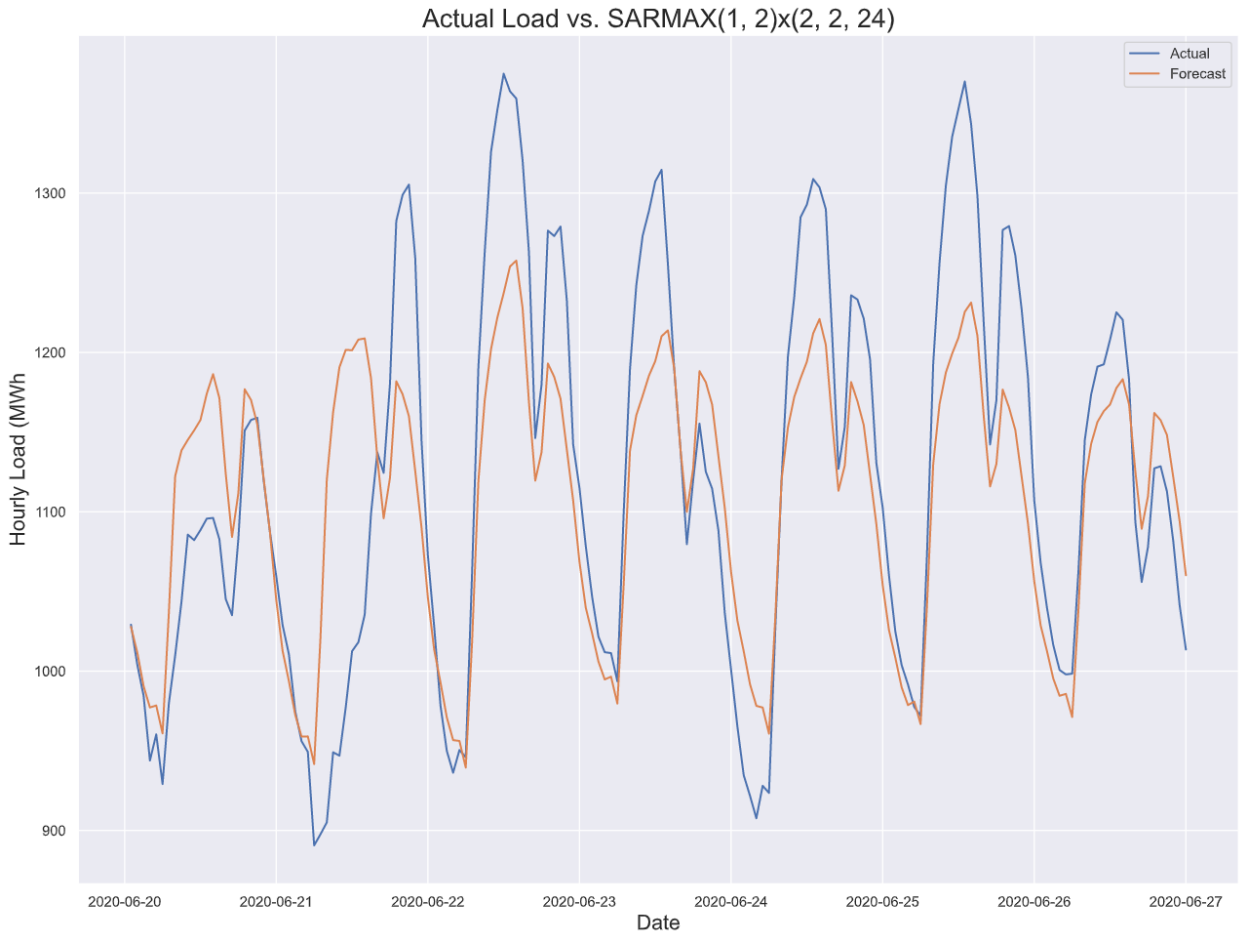

Figure 10: Electricity Load Forecasts from June 20, 2020 at 1:00 AM to June 27, 2020 at 12:00 AM

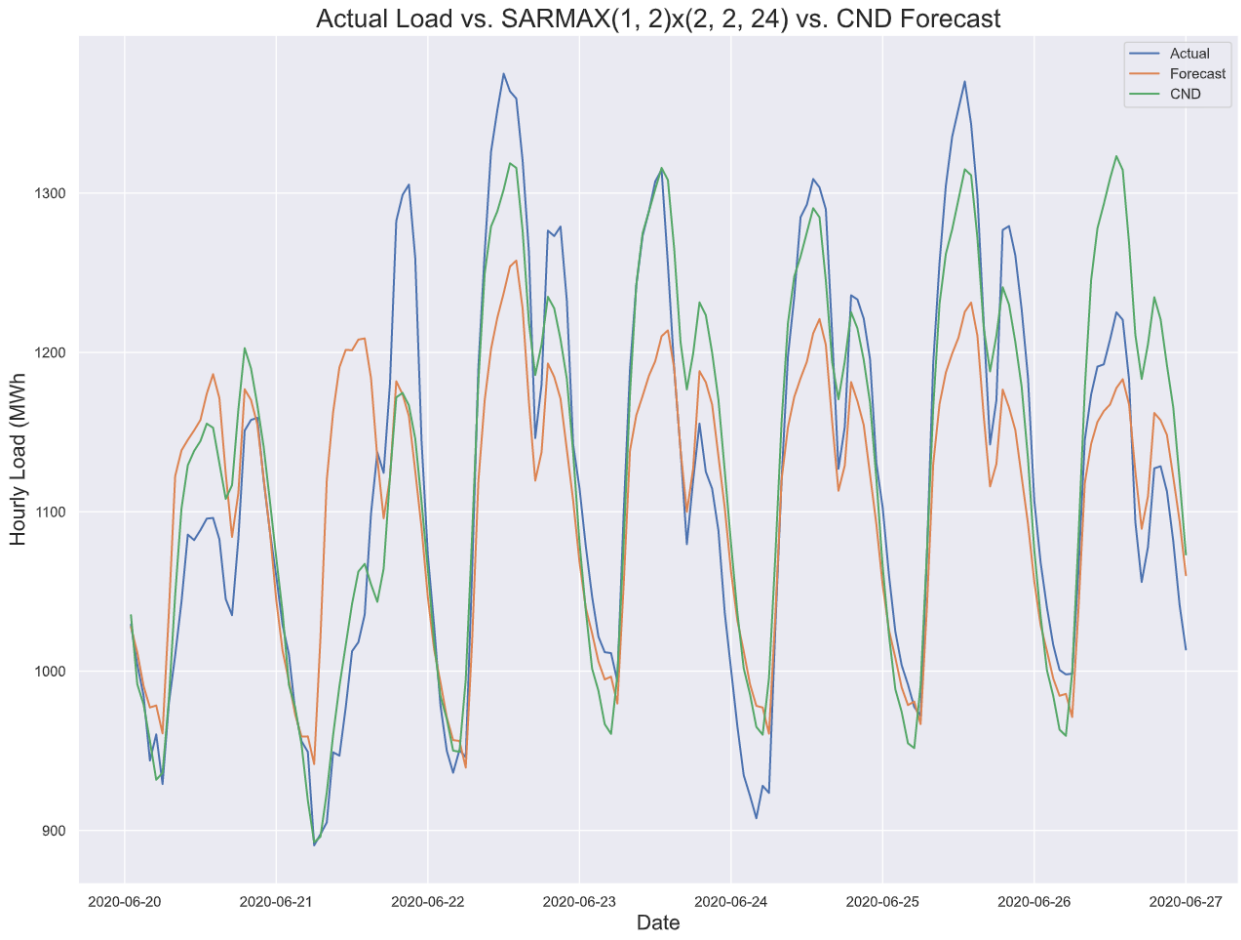

Figure 11: Comparison of Forecasts from June 20, 2020 at 1:00 AM to June 27, 2020 at 12:00 AM

References

Aguilar Madrid, E., & Antonio, N. (2021). Short-term electricity load forecasting with machine learning. *Information*, *12*(2), 50.

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*(5-6),

594–621.

Al-Alawi, S. M., & Islam, S. M. (1996). Principles of electricity demand forecasting. I. methodologies. *Power Engineering Journal*, *10*(3), 139–143.

Al-Musaylh, M. S., Deo, R. C., Adamowski, J. F., & Li, Y. (2019). Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF reanalysis atmospheric predictors in southeast queensland, australia. *Renewable and Sustainable Energy Reviews*, *113*, 109293.

Arora, S., & Taylor, J. W. (2013). Short-term forecasting of anomalous load using rule-based triple seasonal methods. *IEEE Transactions on Power Systems*, *28*(3), 3235–3242.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Bunn, D., & Farmer, E. D. (1985). *Comparative models for electrical load forecasting*.

Bunn, D., & Farmer, E. D. (1985). *Comparative models for electrical load forecasting*.

Cai, M., Pipattanasomporn, M., & Rahman, S. (2019). Day-ahead building-level load forecasts using deep learning vs. Traditional time-series techniques. *Applied Energy*, *236*, 1078–1088.

Camara, A., Feixing, W., & Xiuqin, L. (2016). Energy consumption forecasting using seasonal ARIMA with artificial neural networks models. *International Journal of Business and Management*, *11*(5), 231.

Caro, E., Juan, J., & Cara, J. (2020). Periodically correlated models for short-term electricity load forecasting. *Applied Mathematics and Computation*, *364*, 124642.

Cools, M., Moons, E., & Wets, G. (2009). Investigating the variability in daily traffic counts through use of ARIMAX and SARIMAX models: Assessing the effect of holidays on two site locations. *Transportation Research Record*, *2136*(1), 57–66.

Dash, S. K., & Dash, P. K. (2019). Short-term mixed electricity demand and price forecasting using adaptive autoregressive moving average and functional link neural network. *Journal of Modern Power Systems and Clean Energy*, *7*(5), 1241–1255.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427–431.

Dordonnat, V., Koopman, S. J., Ooms, M., Dessertaine, A., & Collet, J. (2008). An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, *24*(4), 566–587.

Elamin, N., & Fukushige, M. (2018). Modeling and forecasting hourly electricity demand by SARIMAX with interactions. *Energy*, *165*, 257–268.

Fan, S., & Chen, L. (2006). Short-term load forecasting based on an adaptive hybrid method. *IEEE Transactions on Power Systems*, *21*(1), 392–401.

Gonzalez-Romera, E., Jaramillo-Moran, M. A., & Carmona-Fernandez, D. (2006). Monthly electric energy demand forecasting based on trend extraction. *IEEE Transactions on Power Systems*, *21*(4), 1946–1953.

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, *199*(3), 902–907.

Haida, T., & Muto, S. (1994). Regression based peak load forecasting using a transformation technique. *IEEE Transactions on Power Systems*, *9*(4), 1788–1794.

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, *16*(1), 44–55.

Hu, Z., Bao, Y., & Xiong, T. (2013). Electricity load forecasting using support vector regression with memetic algorithms. *The Scientific World Journal*, *2013*.

Huang, S.-J., & Shih, K.-R. (2003). Short-term load forecasting via ARMA model identification including non-gaussian process considerations. *IEEE Transactions on Power Systems*, *18*(2), 673–679.

Kavaklioglu, K., Ceylan, H., Ozturk, H. K., & Canyurt, O. E. (2009). Modeling and prediction of turkey's electricity consumption using artificial neural networks. *Energy Conversion and Management*, *50*(11), 2719–2727.

Kim, M. S. (2013). Modeling special-day effects for forecasting intraday electricity demand. *European Journal of Operational Research*, *230*(1), 170–180.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1-3), 159–178.

Perron, P. (1988). Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of Economic Dynamics and Control*, *12*(2-3), 297–332.

Sigauke, C., & Chikobvu, D. (2011). Prediction of daily peak electricity demand in south africa using volatility forecasting models. *Energy Economics*, *33*(5), 882–888.

Soares, L. J., & Medeiros, M. C. (2008). Modeling and forecasting short-term electricity load: A comparison of methods with an application to brazilian data. *International Journal of Forecasting*, *24*(4), 630–644.

Son, H., & Kim, C. (2017). Short-term forecasting of electricity demand for the residential sector using weather and social variables. *Resources, Conservation and Recycling*, *123*, 200–207.

Taylor, J. W. (2008). An evaluation of methods for very short-term load forecasting using minute-by-minute british data. *International Journal of Forecasting*, *24*(4), 645–658.

Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, *204*(1), 139–152.

Taylor, J. W., De Menezes, L. M., & McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, *22*(1), 1–16.

Taylor, J. W., & McSharry, P. E. (2007). Short-term load forecasting methods: An evaluation based on european data. *IEEE Transactions on Power Systems*, *22*(4), 2213–2219.

Wang, W. (2006). *Stochasticity, nonlinearity and forecasting of streamflow processes*. Ios Press.