Online SARIMA applied for short-term electricity load forecasting

Quang Dat Nguyen
Hanoi University of Science and Technology

Nhat Anh Nguyen
Hanoi University of Science and Technology

Ngoc Thang Tran
Hanoi University of Science and Technology

Vijender Kumar Solanki
CMR Institute of Technology

Rubén González Crespo
Universidad Internacional de La Rioja

Thi Ngoc Anh Nguyen (✉ anh.nguyenthingoc@hust.edu.vn)
Hanoi University of Science and Technology   https://orcid.org/0000-0002-6555-9740

Research Article

Online SARIMA applied for short-term electricity load forecasting

Nguyen Quang Dat · Nguyen Nhat Anh · Tran Ngoc Thang · Vijender Kumar Solanki · Rubn Gonzlez Crespo · Nguyen Thi Ngoc Anh[*]

Abstract Short-term Load Forecasting (STLF) plays a crucial role in balancing supply and demand of load dispatching operation, ensures stability for the power system. With the advancement of real-time smart sensors in power systems, it is of great significance to develop techniques to handle data streams on-the-fly to improve operational efficiency. In this paper, we propose an online variant of Seasonal Autoregressive Integrated Moving Average (SARIMA) to forecast electricity load sequentially. The proposed model is utilized to forecast hourly electricity load of northern Vietnam and achieves a mean absolute percentage error (MAPE) of 4.57%.

1 Introduction

Electrical load prediction is an important task and fundamental to operate the power system of a nation [43]. Three common electrical load prediction categories are considered: short-term, medium term, and long-term [29,13]. However, short-term is the most important be-

Nguyen Quang Dat, Nguyen Nhat Anh, Tran Ngoc Thang, Nguyen Thi Ngoc Anh[*] (Corresponding author, Tel: +84989190192, E-mail: anh.nguyenthingoc@hust.edu.vn) Hanoi university of science and technology, Hanoi, Vietnam

Vijender Kumar Solanki
Department of Computer Science & Engineering, CMR Institute of Technology, Hyderabad, TS, India

Rubn Gonzlez Crespo
Department of Computer Science and Technology, Universidad Internacional de La Rioja, Logroo, Spain

cause of role in operation and planning of power systems in day-to-day scheduling [33,34].

Table 1 Review of several related works for ARIMA and variants. *Note: G.P. Zhang used scaled data.

Model	Metrics	Result	Ref.
SARIMA-Wavelets	MAPE	4.00%	[9]
SARIMA-SVM	MAPE	2.74%	[5]
SARIMA-ANN	MAPE	5.74%	[7]
SARIMA	Accuracy	95.0%	[28]
ARIMA	Accuracy	91.4%	[12]
ARIMA	Accuracy	95.0%	[21]
ARIMA-ANN	RMSE	$4.36 \cdot 10^{-5}$*	[46]
ARIMA	MAPE	6.38%	[10]
ARIMA, SARIMA and VARMA	RMSE	68.82	[16]

Thus many authors shown that SARIMA has its strength in real world applications for seasonal time series. Currently, we need to consider and process large data (big time series), spanning many years, with seasonal factors in the data. So SARIMA is a model that meets those requirements.

In the article [9], Choi et al. used several models to predict the sales in the past several decades. The authors compared the performance of single model SARIMA (pure model) with hybrid model Season ARIMA - Wavelets Transform and a prediction values on linear extrapolation with seasonal adjustment and Evolutionary Neural Networks. With the seasonal time series, the value of improvement on MAPE brought by SARIMA-Wavelets is 4% (which is equal to 9.1% of the percentage error reduction). In [5], Bouzerdoum et al. used the hybrid model SARIMA-SVM, and get result that MAPE is 2.7381%, better than SARIMA (5.1951%) and SVM (5.0706%). The authors shown that the SARIMA model is so good to analyze the linear components of the time

series and estimate them. In [7], Bozkurt O.O. et al. get the result MAPE 5.74% for the data of the Turkish Electricity Market (from Jan 01, 2013 to Dec 31, 2014) by using the SARIMA and artificial neural network model. This data is a very seasonal data. In [10], the authors applied the ARIMA model in expert system courses to predict the behavior of the servers' performance when the user traffic is high and achieved the mean errors around 5%.

Table 2 Review of related works in **electricity load**.

Model	MAPE(%)	RMSE	Ref.
DBN-ANN	2.3%		[11]
SWEMD-ENN	[2.14%; 3.95%]	[1.43; 3.11]	[25]

For the electricity load, in [25] (2017), Liu et al. proposed a hybrid model to forecast electricity load. They got MAPE in interval [2.14%;3.95%] and got RMSE in interval [1.43;3.11]. In other article, Dedinec et al. (2016) [11] shown a new hybrid method, combine from deep belief networks and neural network. Their MAPE is 8.00% for DBM model and is 2.3% for complex model DBM-ANN.

Currently, machine learning models need to have large time series data to be able to estimate and fix parameters of models. For some time series that are not large (or small data), it is difficult to good estimate parameters for the model. This reduces the accuracy of the proposed machine learning model. At that time, online machine learning model was used to overcome this weakness. Online models are learning models where we continuously update the training data points in sequential order. Online machine learning model is often used when calculations cannot be performed over entire data sets or practical applied models cannot wait until that has a large amount of data is collected. This machine learning method is in contrast to classical machine learning method, where all the training time series is available from the beginning for training.

Machine learning online with future data in time series in sequential order online machine learning is still performing the machine learning task with time series like traditional models. The goal of online machine learning is to update the data to the latest time, the latest data value, and to put that value into the training time series.

The online setting for time series prediction was first introduced by Anava et al. [3]. It is based on the online convex optimization framework where the goal is trying to minimize the regret over T iterations instead of directly minimizing the loss function. The authors applied the Online Gradient Descent (OGD) algorithm

[48] and the Online Newton Step (ONS) algorithm [19] to find the best autoregressive coefficients for the online ARMA model. In 2016, Liu et al. [24] extended this model to handle non-stationary time series, called online ARIMA. Based on the two previous works, the online VARMA model was introduced to deal with multivariate time series [45]. However, electricity load data often exhibit seasonal patterns such as daily pattern, weekly pattern, quarterly pattern, and annual pattern. The aforementioned researches didnt tackle the seasonality problem in time series. Therefore, it is important to develop an online technique to forecast time series with seasonality.

In this article, we use online data, constantly updated data to continuously conjecture the parameters of the model SARIMA. Since then, the proposed model is always consistent with the actual data, solving the problem and get the best results.

The article develops as follows. Section 2 presents the methodology of SARIMA model for time series and online machine learning methodology. From there we propose a new method to solve our problem: the Online SARIMA model. In section 3.1, we compare the results of the proposed model with the results of traditional models (we calculated on different models), thereby showing the advantages of the proposed model. export. In Section 3.2, we raise the problem we need to solve, the details of the forecasting method and present the forecast results based on the proposed model. In this section we discusses the comparison between methods. Comments and conclusions are contained in Section 4.

2 Methodology

2.1 ARIMA

Auto-Regressive Integrated Moving Average is a widely used model in information technology. An ARIMA model was represented in an important example from the research of Box and Jenkins [6] that was approached to the time series modeling. In 1991, Brockwell and Davis [31] have fully described this method. The ARIMA model was the most popular in the models building process and in forecasting in a time series.

The pure Auto-Regressive method and the pure Moving Average method can be used for prediction, but combining method (AR + MA) is better than pure method [46]. Salas et al. (in [36]) developed an estimation model with periodic coefficients of the ARIMA model.

A model ARIMA(p, d, q) for the time series y_t is expressed as:

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\ + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

where y_t are the actual values in data; ε_t is the random error at the moment t, $\{\phi_i\}$ are the AutoRegressive (AR) coefficients, $\{\theta_j\}$ are the Moving Average (MA) coefficients; the natural number p and q are orders that are often referred to as Auto-Regressive and moving average polynomials...

or follow function:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L^i)^d y_t = \theta_0 + \left(1 - \sum_{i=1}^{p} \phi_i L^i\right) a_t \quad (2)$$

where $\{\phi_i\}$ are the AutoRegressive coefficients; $\{\theta_i\}$ are the Moving Average coefficients; a_t is a white Gaussian process with zero mean and variance σ_a^2; the parameter θ_0 referres to as the deterministic trend term when d > 0.

ARIMA models can predict the future values in stationarity time series data with good accuracy. Montanari et al. in [28] predicted the flow of the Nile river with 95% confidence by the a seasonal ARIMA model. And in the other work [21], Najeeb Iqbal et al. shown the result, the ARIMA (1,1,1) model and the ARIMA (2,1,2) model can forecasts for wheat area and production with 95% confidence interval values, that used the data from Pakistan's government. In the article [12], the researchers predicted the gas and fuel in Turkey with accuracy 8.6%.

But, ARIMA models can only predict well for stationarity series [46,15]. With the non-stationarity, ARIMA procedures run, got the results that the not-high accuracy [47].

For the better accuracy, several methods was proposed. Some authors have improved the ARIMA method and have actually achieved better results.

In article [27], the author Helmut L. shown the model VAR and the method VARMA. By mathematical methods, the author has proved that these two methods in some cases give better results than model ARMA and model ARIMA. In the article [17], the author Hallin M. and Davy P. also demonstrated the same thing with the VARMA method.

Sheida et al. in [41] (2018) used the fuzzy ARIMA to forecast Irans steel consumption. The authors compared the predictive power of the fuzzy and non-fuzzy, the linear and non-linear forecasting models in cases estimate point and interval. In the case estimate interval forecasting, the Sheida's model is 59,77% better than ARIMA model. In the case estimate point, the authors'

model is 39,42% better than ARIMA (in the criteria RMSE) and 29,71% better in criteria MAE.

Other authors have come up with hybrid models, aimed at some models working well under one condition, others running well under other conditions. When using the hybrid model, the authors split the data of the problem into many small parts, each of which is solved by an appropriate method, thereby providing better combined results when using a single method.

First time, in 2003, Zhang P. [46] used the hybrid model, combine from ARIMA and ANN. He got very good result: the Zhang's model is 18.87% better (in MSE) or 7.97% (in MAD) than ARIMA only. In article [30], the authors combine two methods ARIMA and Support Vector Machine for stock price forecasting. The biggest MAPE value that they got, is 1.7988 for their hybrid model, but smaller than result of the ARIMA model. Aladag et al. in [2] got the result is 0.009 (in scaled data, MSE criteria). Khashei et al. in 2011 [22] used hybrid model for 3 public data series, and have received at lease 10.43% better (MAE) or 21.96% (MSE) than ARIMA. In 2014, Bubu et al. [4] used the hybrid ARIMA model and ANN model in moving-average filter to predict electricity price data in New South Wales from the Australian National Electricity Market and data of the prices of the Larsen and Turbo (L&T) company stock. They got the result that was 26% better than ARIMA in MAE criteria, was 58% better than ARIMA in MSE.

2.2 SARIMA

In particular, SARIMA is a linear models, is one of the most widely used for time series analysis and forecasting. A time series has N values that is denoted by $\{X_t | t = 1, 2, ..., N\}$ is generated by a SARIMA $(p, d, q) \times (P, D, Q)_s$ process by following equation:

$$\phi(L)(1 - L)^d \Phi(L^s)(1 - L^s)^D y_t = c + \theta(L)\Theta(L)\varepsilon_t \quad (3)$$

where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \\ \Phi(L) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \cdots - \Phi_P L^P \\ \theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q \\ \Theta(L) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \cdots - \Theta_Q L^Q \quad (4)$$

where p, d, q are hyper parameter of ARIMA; P, D Q are hyper parameters of seasonal and all hyper parameters are integers; the lag operator is denoted L; and s is the length of seasonal period.

$$SARIMA = ARIMA \underbrace{(p, d, q)}_{\text{Non-seasonal part}} \times \underbrace{(P, D, Q)_s}_{\text{Seasonal part}} \quad (5)$$

In this:

- s is the seasonal period.
- $\phi(L)$ is the Auto-Regressive operator (AR) of order p.
- $\Phi(L)$ is operator of Seasonal Auto-Regressive (S-AR) with order P.
- $\theta(L)$ is operator of the Moving Average (MA) with order q.
- $\Theta(L)$ is operator of the Seasonal Moving Average (S-MA) with order Q.
- d is regular differences number. In [39], d is shown that is less than 2.
- D is seasonal differences number. If $D = 0$ then the time series has no seasonality effect; if $D \geq 1$ then the time series has seasonality effect.
- ε_t is considered residual or white noise signal at the moment t. $\{\varepsilon_t\}_{t \geq 0}$ are identically independently Gaussian$(0, \sigma_\varepsilon^2)$ distributed.

Algorithm 1 Seasonal ARIMA algorithm

Input: Data $\{x_t\}$
1: Find s (the seasonal period of the data)
2: $aic = inf$
3: **for** (p, d, q, P, D, Q) **do**
4:      model $\leftarrow$ fit(arima(p,d,q)(P,D,Q)$_s$
5:      $aic_{current} \leftarrow$ calculate(AIC(model))
6:
7:      **if** $aic_{current} < aic$ **then**
8:          model$_{optimal} \leftarrow$ model
9:          $aic \leftarrow aic_{current}$
10:      **end if**
11: **end for**
12: Return model$_{optimal}$

In summary, we can build the SARIMA model through four steps:

- **1st step:** identify the SARIMA model. This step analyze the variables in order of the time series and verify its that is stationarity or not. If the time series is not stationarity, we need use the integral for them (fix the parameter d). In this step, a deserved combined model from Auto-Regression (AR) and Moving Average (MA) is determined;
- **2nd step:** estimate the unknown parameters of the SARIMA model. In this step we determine a SARIMA model with fixed parameters p and q. After this step, a model SARIMA has been created;
- **3rd step:** the validation phase of SARIMA model. This step tests the precision of the chosen model. Also in this step, the better enhancements are established;
- **4th step:** predict the time series by the fixed SARIMA model. This step, the future values of the time series will be forecast based on the known data with a confidence interval;

In the article [32], the researchers build and run SARIMA model on the air passenger data, and got the result MAPE 1.5142% and result RMSE 95.6635. In the article [40], the authors got result MAPE 9.05% for model SARIMA on data Beijing air passenger. In other article [23], the authors predict the solar PV power data of April 2017 to July 2017, and they got result RMSE 16.1320, MAPE 22.1810%. Luo et al. in 2013 [26] used SARIMA model for forecasting the vegetable price (Cucumber Price data). They have asserted that SARIMA is good at forecasting when their average fitting error is 17%, or the prediction data of 12 months (in the year 2011) is in line with the actual trend, and their average error reaches 25%.

In others articles, the researchers was combine the SARIMA with other model (or other models) to increase the accuracy of the result. In 2014, Ruiz et al. combine the SARIMA with ANN, and got result better than SARIMA only [35]. Other hybrid model, the SARIMA - linear regression, proposed by t. Fang et al. [14] got the result MAPE 9.85% and got result RMSE 22.85. In [42], Vagropoulos et al. predicted the data of hourly PV power generation in one PV plant at the Attica, outside Athens. In day-ahead forecasting, the result of the optimized combined model between SARIMA and SARIMAX was the best in all models in the case of average yearly (NRMSE = 10.25%, smallest in all results of all models).

In the article of Hazan, Agarwal, and Kale (2007) [19], the researchers build and run some online models (include ONS model), and proved that the time for running this model is O(n^2). Of caught, this model can be implemented in space O(n^2), too. In 2008, Agarwal et al. [1] used Kernel-based online machine learning. The proposed model was small for lower memory limit when compared to others models. In the article [8], the author Bubeck S. has achieved the same result with his ascertainment.

In the article [3] (2013), the authors confirm that can't predict the noise in the data. A good model for prediction will get an average error rate that is at least the variance of the noise (0.09 in article's setting). The ARMA-ONS model will better than the other online model. This model lower the regret in the authors' setting of the article, and quickly approaches the performance of the perfect forecasting values. One time again, ARMA-ONS is superior to the other models. The authors can clearly shown the robustness of online model(s) to correlated noise. They got a very good result on ARMA-ONS model: an average error rate, that

converges approximately to the variance of the noise, is 0.0833 for this online model.

In 2014, Hoi et al. [20] built a model and programmed a program that uses online machine learning to classify a data sets, with an accuracy no less than 0.332±0.012.

In the articles [24], the researchers was combine the Online model with other model (ARMA) to increase the accuracy of the result. The proposed models got the result better than single ARMA model only.

Schmidt et al. in 2018 [37] shown the model for detection anomaly events. The authors calculate the percentage of detected anomaly value for evaluation: how many false alarms in total per data point where given by the individual setups. With several time run model with difference parameters, in the worse case, accuracy is 60%, in the best case, accuracy is 90%.

In article [45], the authors compered three models: VARMA, VARMA-OGD and VRAMA-ONS. Criteria is running time and resource cost for running. For best case of VARMA, running time of VARMA model is 801.66% running time of VARMA-ONS, is 3206.66% running time of VARMA-ODG.

2.3 Online machine learning

Currently, in computer science, Online machine learning is a method solve a growing time series data [48, 18]. New values of the time series, in a sequential order, is used to update the model, thereby creating the best model for prediction for future data at each step. Online machine learning is opposed to offline learning techniques, which that predict by learning on the all of entire time series at once.

Online machine learning is used in the cases that is essential for the model to automatically adapt to new values in the time series, or when the data is upgraded.

Online machine learning can classify to some classes [19]:

- Online gradient descent: This algorithm is simple to implement. This method is the analogue of the Gradient Descent optimization method for the online setting [38]. In the online gradient descent model, the running time (per iteration given the gradient) is O(n).
- Online Newton step (ONS) is the online method from the NewtonRaphson method.

In this research, the Online Newton steps algorithm will be used for proposed model.

2.4 Online SARIMA

In this section, we assume the time series $\{X_t, t = 1, \ldots, T\}$ is a sequence of observations with seasonal period s. In online learning setting, this time series is not available all at once but comes one-by-one as a stream of data. Specifically, at each time t, only the history $X_1, X_2, \ldots, X_{t-1}$ is available to us. After we predict the time series at time t as $\tilde{X}_t$, the real value X_t is then observed and revealed to us.

The seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$ model assumes the time series is generated by:

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)\epsilon_t, \{\epsilon_t\} \sim WN(0, \sigma^2), \quad (6)$$

where the notation $\{\epsilon_t\} \sim WN(0, \sigma^2)$ indicates that $\{\epsilon_t\}$ is a white noise sequence with zero mean and variance σ^2,

$Y_t = (1 - B)^d(1 - B^s)^D X_t$ is the differenced time series,

$$\phi(B) = 1 - \overline{\phi}(B),$$
$$\Phi(B^s) = 1 - \overline{\Phi}(B^s),$$
$$\overline{\phi}(B) = \phi_1 B + \phi_2 B^2 + \ldots + \phi_p B^p,$$
$$\overline{\Phi}(B^s) = \Phi_1 B^s + \Phi_2 B^{2s} + \ldots + \Phi_P B^{Ps},$$
$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q,$$
$$\Theta(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs}. \quad (7)$$

In online learning context, at each iteration t, after we predict $\tilde{X}_t$ and X_t is observed, the loss function ℓ_t is revealed. For time series prediction problem, ℓ_t is usually the squared loss $\ell_t(X_t, \tilde{X}_t) = (X_t - \tilde{X}_t)^2$. Our objective is to try to keep the regret over T iterations, defined as:

$$R_T = \sum_{t=1}^{T} \ell_t(X_t, \tilde{X}_t) - \min_{\phi,\Phi,\theta,\Theta} \sum_{t=1}^{T} \ell_t(X_t, \tilde{X}_t(\phi, \Phi, \theta, \Theta)), \quad (8)$$

where $\min_{\phi,\Phi,\theta,\Theta} \sum_{t=1}^{T} \ell_t(X_t, \tilde{X}_t(\phi, \Phi, \theta, \Theta))$ is the total loss of the best fixed SARIMA model, as small as possible. To be specific, the growth of the difference between the cumulative loss of our predictions and that of the best fixed SARIMA model should be insignificant relative to T as T increases.

As mentioned in [3], predictions based on SARIMA model depend on the noise term $\{\epsilon_t\}$, which is unknown to us. Therefore, it is difficult to make predictions using SARIMA. Following [3,24], we modify the original model by omitting the moving average part θ, Θ. The

modified model $(p, d, 0) \times (P, D, 0)_s$ is given by the following formula

$$\phi(B)\Phi(B^s)Y_t = \epsilon_t, \tag{9}$$

where

$$Y_t = (1 - B)^d(1 - B^s)^D X_t, \text{ and } \{\epsilon_t\} \sim WN(0, \sigma^2). \tag{10}$$

From (9), the prediction $\tilde{Y}_t$ for the differenced time series at time t is generated via the formula

$$\tilde{Y}_t = \overline{\phi}(B)Y_t + \overline{\Phi}(B^s)Y_t - \overline{\phi}(B)\overline{\Phi}(B^s)Y_t. \tag{11}$$

The squared loss function is defined as follows

$$\begin{aligned}
f_t(\phi, \Phi) &= \ell_t(X_t, \tilde{X}_t(\phi, \Phi)) \\
&= (X_t - \tilde{X}_t(\phi, \Phi))^2 \\
&= (Y_t - \tilde{Y}_t(\phi, \Phi))^2 \\
&= [(1 - \overline{\phi}(B) - \overline{\Phi}(B^s) + \overline{\phi}(B)\overline{\Phi}(B^s))Y_t]^2 \\
&= [(1 - \overline{\phi}(B)(1 - \overline{\Phi}(B^s)Y_t]^2 \\
&= [\phi(B)\Phi(B^s)Y_t]^2.
\end{aligned} \tag{12}$$

For gradient-based methods, the gradient information is used to update the parameters at each iteration. The partial derivatives of the loss function with respect to ϕ_i and Φ_j are calculated as

$$\frac{\partial f_t}{\partial \phi_i} = -2(Y_t - \tilde{Y}_t)B^i\Phi(B^s)Y_t, \quad i = \overline{1, p}, \tag{13}$$

$$\frac{\partial f_t}{\partial \Phi_j} = -2(Y_t - \tilde{Y}_t)B^{js}\phi(B)Y_t, \quad j = \overline{1, P}, \tag{14}$$

$$\nabla_t = \nabla f_t(\phi, \Phi) = (\frac{\partial f_t}{\partial \phi_1}, \dots, \frac{\partial f_t}{\partial \phi_p}, \frac{\partial f_t}{\partial \Phi_1}, \dots, \frac{\partial f_t}{\partial \Phi_P})^T. \tag{15}$$

For the online gradient descent method, the parameters are updated by iteratively moving in the direction of the (negative) gradient $-\nabla_t$. It uses only first order derivative information. In contrast, the online Newton step method requires second order derivative information. It tries to approximate the Hessian. In particular, at each iteration, the algorithm chooses the direction of $-A_t^{-1}\nabla_t$ instead of $-\nabla_t$, where A_t is related to the Hessian [19]. This matrix is also iteratively updated by setting $A_t \leftarrow A_{t-1} + \nabla_t\nabla_t^T$.

Our proposed algorithm is as follows. The parameters $\phi = (\phi_1, \dots, \phi_p)^T$ and $\Phi = (\Phi_1, \dots, \Phi_P)^T$ of SARIMA model are updated using the online Newton step method.
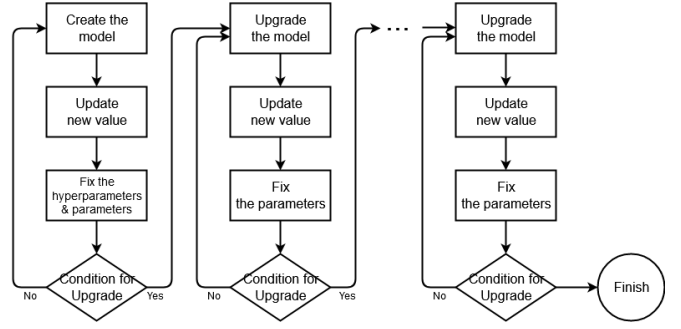

Fig. 1 The diagram of the online machine learning model.

Algorithm 2 Seasonal ARIMA ONS

Input: Regular order p; seasonal order P; seasonal period s; learning rate η; an initial $(p + P) \times (p + P)$ matrix A_0.
1: Choose $(\phi, \Phi)^1 \in K = \{\omega \in \mathbb{R}^{p+P}, |\omega_j| \leq 1, j = 1, \dots, p + P\}$ arbitrarily.
2: **for** $t = 1$ to $T - 1$ **do**
3:     Predict $\tilde{Y}_t = \overline{\phi}(B)Y_t + \overline{\Phi}(B^s)Y_t - \overline{\phi}(B)\overline{\Phi}(B^s)Y_t$
4:     Observe Y_t and compute loss $f_t((\phi, \Phi)^t)$
5:     Let $\nabla_t = \nabla f_t((\phi, \Phi)^t)$, update $A_t \leftarrow A_{t-1} + \nabla_t\nabla_t^T$
6:     Set $(\phi, \Phi)^{t+1} \leftarrow \Pi_K^{A_t}((\phi, \Phi)^t - \eta A_t^{-1}\nabla_t)$
7: **end for**

$\Pi_K^{A_t}$ is the projection onto K with the norm induced by A_t. Note that at each iteration, after we predict $\tilde{Y}_t$, the prediction $\tilde{X}_t$ of the original time series is obtained by expanding and rearranging equation $Y_t = (1 - B)^d(1 - B^s)^D X_t$.

2.5 Hyperparameters selection

We devise a procedure to select the best orders p and P for the model. Suppose we have historical data $X_1, \dots, X_{T_1}$. This time series is available all at once beforehand. The future time series $X_{T_1+1}, \dots, X_{T_2}$ is available one-by-one as in the online learning setting. We will choose the orders based on the historical data. First, the history is split into training and validation set in chronological order. Then we sweep the hyperparameters through a predefined range of values. Specifically, a grid search is performed over $\{0, 1, \dots, p_{max}\}$ for p and $\{0, 1, \dots, P_{max}\}$ for P, where p_{max} and P_{max} are predefined constant values. For each pair of orders (p, P), we train the model on the training set in online manner (observations come sequentially although the whole training set is available). The model is then evaluated on validation set and we choose the model with the best result. In this study, we use RMSE (Root Mean Square Error) as the evaluation metric and therefore the model with lowest MSE (Mean Squared Error) is chosen.
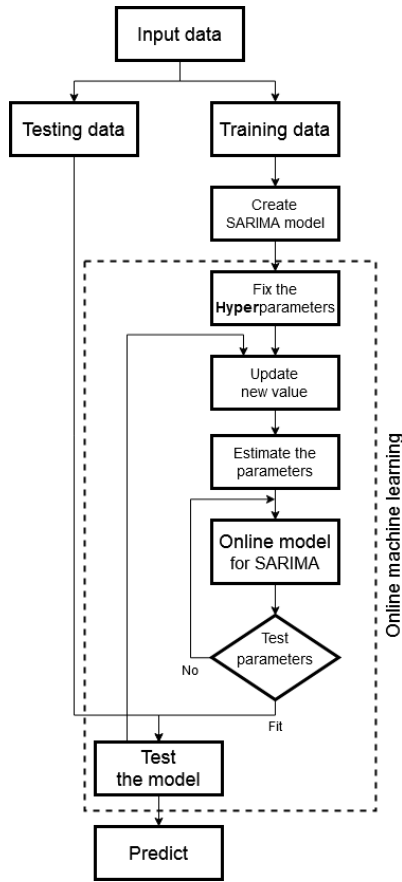
Fig. 2 The Online SARIMA model.

For differencing orders, we choose seasonal differencing order $D = 0$. Non-seasonal differencing order d is chosen based on augmented Dickey-Fuller test. If the p-value is greater than significance level $\alpha = 0.05$, we will choose $d = 1$. Otherwise, no differencing is required.

3 Experiment

In this section, we introduce the data set for experimenting.

3.1 Data set

Although we want to compare the forecast results for electricity load data with other models of the other authors. However, the electricity load data set with season characteristic is not available. In order to compare our proposed model Online-SARIMA with the other methods, the air passenger data set being used for forecasting seasonal models. Our experimental results show that our proposed model performance is better than the others with the same data set. The reason for better

results is online learning model updating the adaptive model as additional new data.

3.1.1 China Air Passenger

The data set **Monthly air transfer in China** (got public from website http://www.stats.gov.cn/) is tested our proposed model. This is the data that shows the air passenger numbers transferred from January 2005 to Aug 2018 in China (all airports in China) on a monthly basis. This data set have 164 time series records. Qin use 158 values in data for training set, and 6 value in data series for test set [32]. This data set is already of a time series class therefore no further class or date manipulation is required.

This data set, we used for compare the proposed model with model of Qin et al., the model was combining seasonal and trend component procedures based on loess with echo state network [32].

In the article [44], the authors used the same values of this data, but start at Feb 2005, finish at Feb 2018. We can consider that this is the same data set. The result of the Xu's model will be used for comparison.

3.1.2 Northern Vietnam Electricity Load

This data set contains electricity load measurements on working days (Monday to Friday) dated from 1 January 2015 to 30 August 2019. The data were collected at 1-hour intervals and consisted of 29208 instances. The data set is represented as a table. Each entry is corresponding to a date and the columns indicate the specific hour. We treat these 24 columns as 24 separate time series and the results are evaluated individually.

4 Results

In this section, we demonstrate the results of the proposed model.

4.1 China Air Passenger

This subsection will show the forecasting results in following tables. The following findings are shown on the table. Therefore, in this research, we will use the criteria MAPE, RMSE and some case on the criteria MAE.

In this comparison, we used data set in China and compare result of the proposed model with the models that were proposed by Qin et al. [32] and the model was proposed by Xu et al [44].

We run the model 10 times, and get the average value.

Table 3 Results of our proposed model and other methods on the **Monthly air transfer in China** data

	Model	MAPE(%)	RMSE	MAE	Public	
0	SARIMA	1.5142%	95.6635		(2019)	[32]
1	SARIMA-ANN		237.6600	207.1100	(2014)	[35]
1*	SARIMA-ANN (Qin et al. rebuild)	4.5500%			(2019)	[32]
2	SARIMA-SVR	2.3100%	131.1400	102.8000	(2019)	[44]
3	ESN	2.6360%	159.2997		(2019)	[32]
	PSO-ESN	2.3785%	135.3856			
	GWO-ESN	2.2646%	143.3343			
	WOA-ESN	1.7304%	108.8308			
	GOA-ESN	1.8496%	116.5840			
	EMD-ESN	2.2852%	148.3139			
	SEAM-ESN	4.7069%	66.8646			
	STL-ESN	1.4827%	83.1109			
4	Proposed model	**0.7780**%	**49.4118**	**35.2028**		
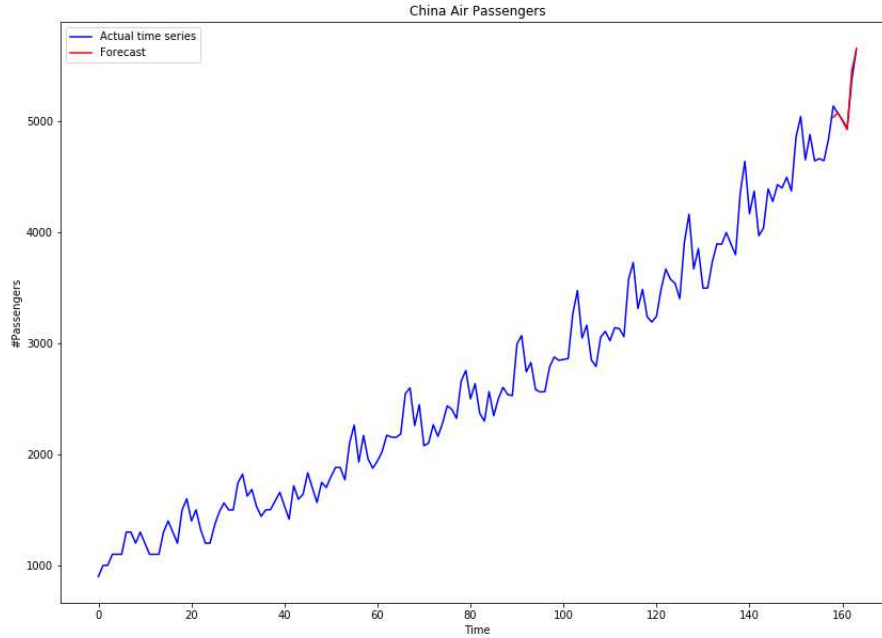
Fig. 3 The results of the **SARIMA-ONS** model.

We can see that the new method has yielded much better results than many methods that have been introduced in recent years. In the **table 4**, in the MAPE, the result of the our model is 0.7780%, smallest in all of other pure models, which shown the good result of the proposed model in forecasting for time series. This result of the proposed model is 83% better than Ruiz's model (4.5500%), is 66% better than Xu's model (2.3100%), is 48% better than best model of Qin (in the criteria MAPE, the best model in all of the Qin's model is the hybrid model STL-ESN with MAPE is 1.4827%).

By the criteria RMSE, the result of the proposed model is 49.4118, is 80% better than Ruiz's model (237.66), is 63% better than Xu's model (131.14), and is 27% better than best model of Qin (in the criteria RMSE, the best model in all of the Qin's model is the hybrid model SEAM-ESN with RMSE is 66.8646).

By the MAE criteria, the result of our proposed method is 35.2028, is 66% better than Xu's model (102.80), and 83% better Ruiz's model (207.11).
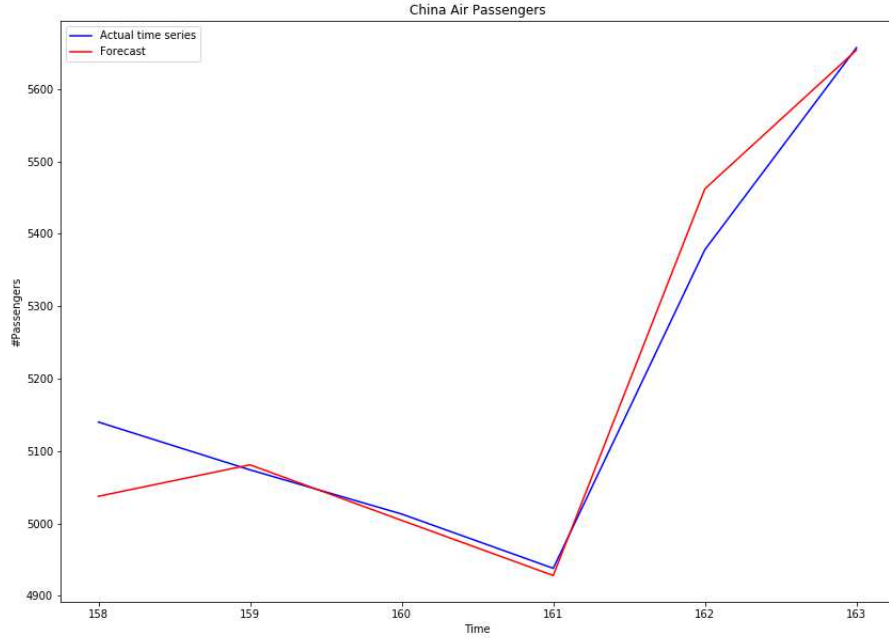
Fig. 4 The results of the **SARIMA-ONS** model.

4.2 Northern Vietnam Electricity Load

For the Northern Vietnam Electricity Load dataset, it is apparent that there is a seasonal pattern within a year. Since there are approximately 250 working days in a year, we chose $s = 250$. For each time series, 80% were used as training/historical data and 20% were used for testing. Each time series was split in chronological order. For hyper-parameters selection, we performed grid search with $p_{max} = 5$ and $P_{max} = 3$ and the selected model was compared with a $(1, 1, 0) \times (1, 0, 0)_{250}$ model. The models were trained on Google Colab and the average running time was 1.95 seconds per time series. We summarized the results in table 4.

The electricity load depend on the hour of a day. By experimental results, the MAPE using the historical data at the same hours of a day give the better performance, so each day, 24 models for each hour are build for prediction.24 models are divided by three categories by consuming electricity load that are low, medium and high. Three hours 5am, 13h, 21h are be-haft on low, high and medium respectively.

The figures 4.2 shows the optimization hyper parameters of SARIMA online for electricity load predictions at 05 o'clock AM being on behalf on low electricity load consumption. Concretely, the figures 4.2dimensions shows minimizing MAPE of SARIMA online

model base on historical data with two hyper parameter p for Auto-regression and hyper parameter and P seasonal Auto-regression.The best hyper-parameters $(p, P) = (1, 2)$. Similarly, The figures 4.2 presents the optimization hyper-parameters of SARIMA online for electricity load predictions at 13 o'clock PM being on behalf on high electricity load consumption. The best hyper-parameters $(p, P) = (0, 5)$. Furthermore, The figures 4.2 show the optimization hyper-parameters of Auto-regression and seasonal auto-regression for SARIMA online model at 21 o'clock PM being on behalf on medium electricity load consumption $(p, P) = (2, 0)$.

5 Conclusion

This paper proposes SARIMA-ONS model - an online version of SARIMA model to deal with streaming time-series data and apply to forecasting electricity load. Experiments show that SARIMA-ONS provides the competitive predictive performance when we compare it to the other methods with the same data sets.

Following the online learning approach, SARIMA-ONS model continuously updates the parameters to obtain the smallest forecast error. The second advantage is fast speed because there is no need to learn other offline models. However, the weakness of SARIMA-ONS
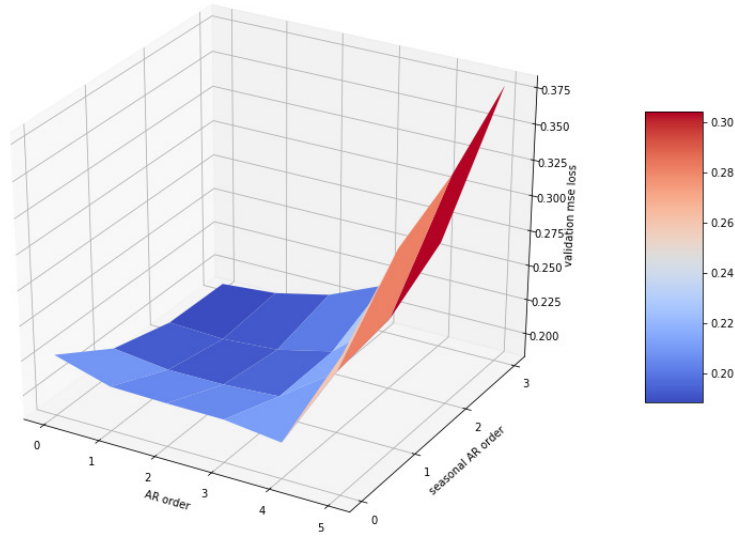
Fig. 5 Optimization hyper parameters Auto-regression and seasonal auto-regression for SARIMA online model at 05 o'clock AM being on behalf on low electricity load consumption. The best hyper-parameters $(p, P) = (1, 2)$.
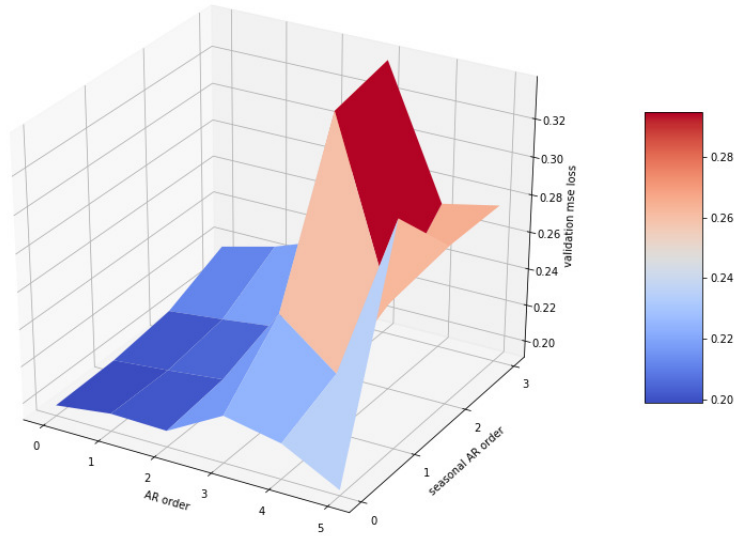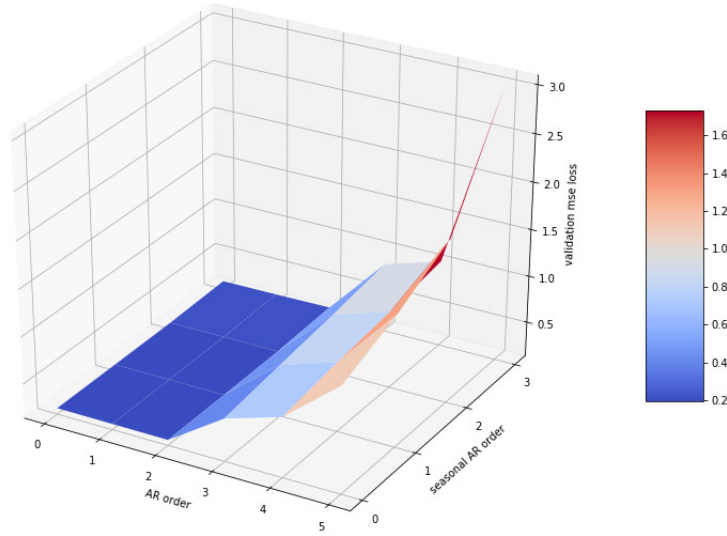

Fig. 6 Optimization hyper parameters Auto-regression and seasonal auto-regression for SARIMA online model at 13 o'clock PM being on behalf on high electricity load consumption. The best hyper-parameters $(p, P) = (0, 5)$.

algorithm is using ONS is more complexity and difficult to understand than other optimisation algorithm such as OGD.

The article has contributed and the main results as follows:

- Building successfully an online learning model for a seasonal auto-regssion moving average for time series.
- Succeeding in optimisation of hyper parameter based on performance evaluation of loss function.

- Forecasting successfully SARIMA-ONS for electricity load data that require online learning method.
- Obtaining the electricity load forecasting the MAPE is less than 5 percent.
- SARIMA- ONS making better than other methods by experimental results.
- SARIMA- ONS learns an average of 1.7 seconds, so the results are faster than SARIMA methods three times.

Fig. 7 Optimization hyper parameters Auto-regression and seasonal auto-regression for SARIMA online model at 21 o'clock PM being on behalf on medium electricity load consumption.The best hyper-parameters $(p, P) = (2, 0)$.

Hour	$(1, 1, 0) \times (1, 0, 0)_{250}$ model MAPE(%)	Proposed model	Proposed model MAPE(%)
1	6.01	$(4, 1, 0) \times (2, 0, 0)_{250}$	**5.71**
2	5.82	$(0, 1, 0) \times (2, 0, 0)_{250}$	**5.49**
3	5.60	$(0, 1, 0) \times (2, 0, 0)_{250}$	**5.27**
4	**5.21**	$(5, 1, 0) \times (0, 0, 0)_{250}$	5.44
5	4.92	$(1, 1, 0) \times (2, 0, 0)_{250}$	**4.68**
6	3.96	$(2, 1, 0) \times (2, 0, 0)_{250}$	**3.94**
7	**3.47**	$(2, 1, 0) \times (1, 0, 0)_{250}$	3.51
8	**3.62**	$(2, 1, 0) \times (1, 0, 0)_{250}$	3.64
9	**3.98**	$(2, 1, 0) \times (0, 0, 0)_{250}$	4.03
10	**4.26**	$(2, 1, 0) \times (0, 0, 0)_{250}$	4.35
11	**4.68**	$(2, 1, 0) \times (0, 0, 0)_{250}$	4.71
12	5.33	$(2, 1, 0) \times (0, 0, 0)_{250}$	**5.28**
13	5.65	$(5, 1, 0) \times (0, 0, 0)_{250}$	**5.62**
14	5.46	$(2, 1, 0) \times (0, 0, 0)_{250}$	**5.45**
15	**4.84**	$(2, 1, 0) \times (0, 0, 0)_{250}$	4.87
16	4.00	$(1, 1, 0) \times (0, 0, 0)_{250}$	**3.99**
17	**3.26**	$(2, 1, 0) \times (1, 0, 0)_{250}$	3.32
18	**2.83**	$(2, 1, 0) \times (0, 0, 0)_{250}$	2.84
19	**3.34**	$(3, 1, 0) \times (0, 0, 0)_{250}$	3.39
20	**3.62**	$(3, 1, 0) \times (0, 0, 0)_{250}$	3.73
21	**4.26**	$(2, 1, 0) \times (0, 0, 0)_{250}$	4.40
22	**5.07**	$(3, 1, 0) \times (1, 0, 0)_{250}$	5.14
23	5.42	$(2, 1, 0) \times (0, 0, 0)_{250}$	**5.40**
24	**5.30**	$(5, 1, 0) \times (0, 0, 0)_{250}$	5.41
Average	4.58	-	**4.57**

Table 4 Results of our proposed model on the Northern Vietnam Electricity Load dataset

For future work, SARIMA-ONS has many potential applications for forecasting problems in many different fields. In the near future, to increase the accuracy of the forecasting model, we will research some hybrid online models combining SARIMA-ONS with other online machine learning models such as ANN, RNN, LSTM, Spiking neural networks. The proposed hybrid model keep the advantage of single model and reduce the disadvantage of each to get better performance.
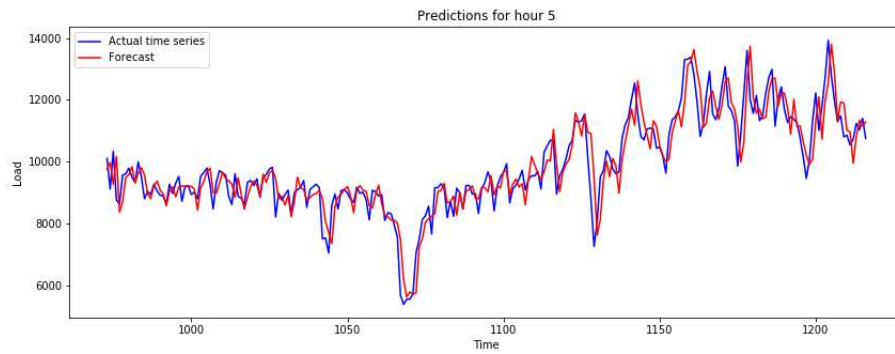
Fig. 8 The results forecast at 05 o'clock AM of the **SARIMA-ONS** model for Electricity load data.
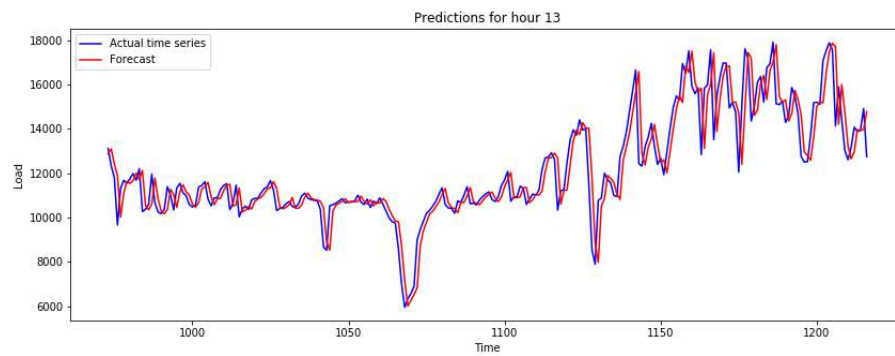

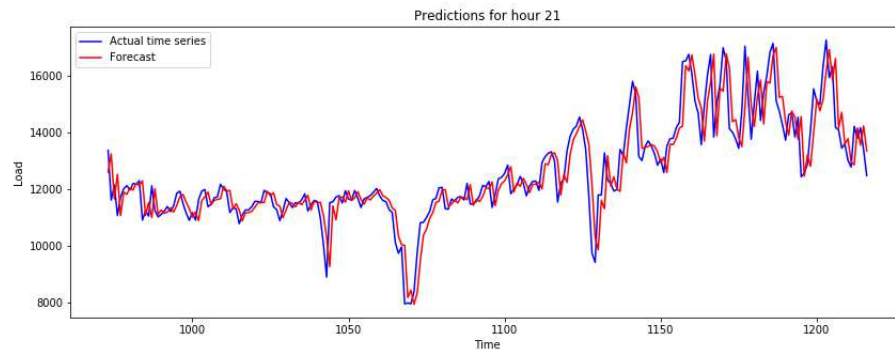**Fig. 9** The results forecast at 13 o'clock PM of the **SARIMA-ONS** model for Electricity load data.

Fig. 10 The results forecast at 21 o'clock PM of the **SARIMA-ONS** model for Electricity load data.

THE FUNDING SOURCE

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Agarwal, S., Saradhi, V.V., Karnick, H.: Kernel-based online machine learning and support vector reduction. Neurocomputing **71**(7), 1230 – 1237 (2008). DOI https://doi.org/10.1016/j.neucom.2007.11.023. URL http://www.sciencedirect.com/science/article/pii/S0925231208000581. Progress in Modeling, Theory, and Application of Computational Intelligenc

2. Aladag, C.H., Egrioglu, E., Kadilar, C.: Forecasting nonlinear time series with a hybrid methodology. Applied Mathematics Letters **22**(9), 1467 – 1470 (2009). DOI https://doi.org/10.1016/j.aml.2009.02.006. URL http://www.sciencedirect.com/science/article/pii/S0893965909001475

3. Anava, O., Hazan, E., Mannor, S., Shamir, O.: Online learning for time series prediction. JMLR: Workshop and Conference Proceedings p. 113 (2013). URL http://proceedings.mlr.press/v30/Anava13.pdf

4. Babu, C.N., Reddy, B.E.: A moving-average filter based hybrid arimaann model for forecasting time series data. Applied Soft Computing **23**, 27 – 38 (2014). DOI https://doi.org/10.1016/j.asoc.2014.05.028. URL http://www.sciencedirect.com/science/article/pii/S1568494614002555

5. Bouzerdoum, M., Mellit, A., Pavan, A.M.: A hybrid model (sarimasvm) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. Solar Energy **98**, 226 – 235 (2013). DOI https://doi.org/10.1016/j.solener.2013.10.002. URL http://www.sciencedirect.com/science/article/pii/S0038092X13004039

6. Box, G., Jenkins, G.: Time series analysis, forecasting and control. Holden-Day, San Francisco, CA. (1970)

7. Bozkurt, O.O., Biricik, G., Taysi, Z.C.: Artificial neural network and sarima based models for power load forecasting in turkish electricity market. PLOS ONE **12**(4), 1–24 (2017). DOI 10.1371/journal.pone.0175915

8. Bubeck, S.: "introduction to online optimization" (2011)

9. Choi, T.M., Yu, Y., Au, K.F.: A hybrid sarima wavelet transform method for sales forecasting. Decision Support Systems **51**(1), 130 – 140 (2011). DOI https://doi.org/10.1016/j.dss.2010.12.002. URL http://www.sciencedirect.com/science/article/pii/S0167923610002356

10. Crespo, R.G., Escobar, R.F., Aguilar, L.J., Velazco, S., Sanz, A.G.C.: Use of arima mathematical analysis to model the implementation of expert system courses by means of free software opensim and sloodle platforms in virtual university campuses. Expert Systems with Applications **40**(18), 7381 – 7390 (2013). DOI https://doi.org/10.1016/j.eswa.2013.06.054. URL http://www.sciencedirect.com/science/article/pii/S0957417413004491

11. Dedinec, A., Filiposka, S., Dedinec, A., Kocarev, L.: Deep belief network based electricity load forecasting: An analysis of macedonian case. Energy **115**, 1688 – 1700 (2016). DOI https://doi.org/10.1016/j.energy.2016.07.090. URL http://www.sciencedirect.com/science/article/pii/S0360544216310076. Sustainable Development of Energy, Water and Environment Systems

12. Ediger, V.S., Akar, S.: Arima forecasting of primary energy demand by fuel in turkey. Energy Policy **35**, 17011708 (2006). DOI 10.1016/j.enpol.2006.05.009. URL https://www.researchgate.net/publication/4947915

13. El-Baz, W., Tzscheutschler, P.: Short-term smart learning electrical load prediction algorithm for home energy management systems. Applied Energy **147**, 10 – 19 (2015). DOI https://doi.org/10.1016/j.apenergy.2015.01.122. URL http://www.sciencedirect.com/science/article/pii/S0306261915001592

14. Fang, T., Lahdelma, R.: Evaluation of a multiple linear regression model and sarima model in forecasting heat demand for district heating system. Applied Energy **179**, 544 – 552 (2016). DOI https://doi.org/10.1016/j.apenergy.2016.06.133. URL http://www.sciencedirect.com/science/article/pii/S0306261916309217

15. Franses, P., Franses, R., van Dijk, D., Dijk, D.: Non-Linear Time Series Models in Empirical Finance. Nonlinear Time Series Models in Empirical Finance. Cambridge University Press (2000). URL https://books.google.com.vn/books?id=5VQuzsGm9WAC

16. Gonzalez Crespo, R., Martnez, C., Escobar, R.: Modeling of network traffic in a digital video transmission via time series arima and sarima univariate and multivariate varma. In: International work-conference on Time Series (2016)

17. Hallin, M., Paindaveine, D.: Affine-invariant aligned rank tests for the multivariate general linear model with varma errors. Journal of Multivariate Analysis **93**(1), 122 – 163 (2005). DOI https://doi.org/10.1016/j.jmva.2004.01.005. URL http://www.sciencedirect.com/science/article/pii/S0047259X04000156

18. Hazan, E.: Introduction to Online Convex Optimization. Nonlinear Time Series Models in Empirical Finance. Foundations and Trends in Optimization (2015). DOI 10.1561/2400000013. URL http://mr.crossref.org/iPage?doi=10.1561%2F2400000013

19. Hazan, E., Kalai, A., Kale, S.: Logarithmic regret algorithms for online convex optimization. Math Learn **69**, 169 – 192 (2007). DOI https://doi.org/10.1007/s10994-007-5016-8. URL https://link.springer.com/article/10.1007/s10994-007-5016-8#citeas

20. Hoi, S.C., Wang, J., Zhao, P.: Libol: A library for online learning algorithms. Journal of Machine Learning Research **15**(1), 495–499 (2014). URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.675.147. Progress in Modeling, Theory, and Application of Computational Intelligenc

21. Iqbal, N., Bakhsh, K., Maqbool, A., Ahmad, A.S.: Use of the arima model for forecasting wheat area and production in pakistan. JOURNAL OF AGRICULTURE & SOCIAL SCIENCES **1 No.2** (2005)

22. Khashei, M., Bijari, M.: A novel hybridization of artificial neural networks and arima models for time series forecasting. Applied Soft Computing **11**(2), 2664 – 2675 (2011). DOI https://doi.org/10.1016/j.asoc.2010.10.015. URL http://www.sciencedirect.com/science/article/pii/S1568494610002759. The Impact of Soft Computing for the Progress of Artificial Intelligence

23. Kushwaha, V., Pindoriya, N.M.: Very short-term solar pv generation forecast using sarima model: A case study pp. 430–435 (2017). DOI 10.1109/ICPES.2017.8387332

24. Liu, C., Hoi, S.C., Zhao, P., Sun, J.: Online arima algorithms for time series prediction. In: Thirtieth AAAI conference on artificial intelligence (2016). URL https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/viewPaper/12135

25. Liu, Y., Wang, W., Ghadimi, N.: Electricity load forecasting by an improved forecast engine for building level consumers. Energy **139**, 18 – 30 (2017). DOI https://doi.org/10.1016/j.energy.2017.07.150. URL http://www.sciencedirect.com/science/article/pii/S0360544217313348

26. Luo, C.S., Zhou, L.Y., Wei, Q.F.: Application of sarima model in cucumber price forecast. In: Mechatronics, Robotics and Automation, *Applied Mechanics and Materials*, vol. 373, pp. 1686–1690. Trans Tech Publications Ltd (2013). DOI 10.4028/www.scientific.net/AMM. 373-375.1686

27. Ltkepohl, H.: Forecasting with varma models. Handbook of Economic Forecasting **1**, 287–325 (2006). DOI https://doi.org/10.1016/S1574-0706(05)01006-2. URL `https://www.sciencedirect.com/science/article/abs/pii/S1574070605010062`

28. Montanari, A., Rosso, R., Taqqu, M.S.: A seasonal fractional arima model applied to the nile river monthly flows at aswan. WATER RESOURCES RESEARCH **36 No.5**, 1249–1259 (2000)

29. Muzaffar, S., Afshari, A.: Short-term load forecasts using lstm networks. Energy Procedia **158**, 2922 – 2927 (2019). DOI https://doi.org/10.1016/j.egypro.2019.01. 952. URL `http://www.sciencedirect.com/science/article/pii/S1876610219310008`. Innovative Solutions for Energy Transitions

30. Pai, P.F., Lin, C.S.: A hybrid arima and support vector machines model in stock price forecasting. Omega **33**, 497 505 (2005). DOI 10.1016/j.omega.2004.07. 024. URL `https://www.sciencedirect.com/science/article/pii/S0305048304001082`

31. Peter, J.B., Richard, A.D.: "time series: Theory and methods" (1991)

32. Qin, L., Li, W., Li, S.: Effective passenger flow forecasting using stl and esn based on two improvement strategies. Neurocomputing **356**, 244 – 256 (2019). DOI https://doi.org/10.1016/j.neucom.2019.04. 061. URL `http://www.sciencedirect.com/science/article/pii/S0925231219306472`

33. Quan, H., Srinivasan, D., Khosravi, A.: Uncertainty handling using neural network-based prediction intervals for electrical load forecasting. Energy **73**, 916 – 925 (2014). DOI https://doi.org/10.1016/j.energy.2014. 06.104. URL `http://www.sciencedirect.com/science/article/pii/S0360544214008032`

34. Romero, Á., Dorronsoro, J.R., Díaz, J.: Day-ahead price forecasting for the spanish electricity market. International Journal of Interactive Multimedia and Artificial Intelligence **5**(4), 42–50 (2019). DOI 10.9781/ijimai.2018.04.008. URL `http://www.ijimai.org/journal/sites/default/files/files/2018/04/ijimai_5_4_5_pdf_14997.pdf`

35. Ruiz-Aguilar, J., Turias, I., Jimnez-Come, M.: Hybrid approaches based on sarima and artificial neural networks for inspection time series forecasting. Transportation Research Part E: Logistics and Transportation Review **67**, 1 – 13 (2014). DOI https://doi.org/10.1016/j.tre.2014. 03.009. URL `http://www.sciencedirect.com/science/article/pii/S1366554514000428`

36. Salas, J.D., Boes, D.C., Smith, R.A.: Estimation of arma models with seasonal parameters. Water Resources Research **18**(4), 1006–1010 (1982). DOI 10.1029/WR018i004p01006. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR018i004p01006`

37. Schmidt, F., Suri-Payer, F., Gulenko, A., Wallschlger, M., Acker, A., Kao, O.: Unsupervised anomaly event detection for vnf service monitoring using multivariate online arima. In: 2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), pp. 278–283 (2018). DOI 10.1109/CloudCom2018.2018. 00061. URL `https://ieeexplore.ieee.org/abstract/document/8591030`

38. Shalev-Shwartz, S.: Online learning and online convex optimization. Foundations and Trends in Machine Learning **4**(2), 107–194 (2012). DOI 10.1561/2200000018. URL `https://www.nowpublishers.com/article/Details/MAL-018`. Progress in Modeling, Theory, and Application of Computational Intelligenc

39. Shumway, R.H., Stoffer, D.S.: Very short-term solar pv generation forecast using sarima model: A case study pp. 47–82 (2006)

40. Sun, S., Lu, H., Tsui, K.L., Wang, S.: Nonlinear vector auto-regression neural network for forecasting air passenger flow. Journal of Air Transport Management **78**, 54 – 62 (2019). DOI https://doi.org/10.1016/j.jairtraman. 2019.04.005. URL `http://www.sciencedirect.com/science/article/pii/S0969699718303818`

41. Torbat, S., Khashei, M., Bijari, M.: A hybrid probabilistic fuzzy arima model for consumption forecasting in commodity markets. Economic Analysis and Policy **58**(1), 22 – 31 (2018). DOI https://doi.org/10.1016/j.eap.2017.12. 003. URL `https://www.sciencedirect.com/science/article/abs/pii/S031359261730067X`

42. Vagropoulos, S.I., Chouliaras, G.I., Kardakos, E.G., Simoglou, C.K., Bakirtzis, A.G.: Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting pp. 1–6 (2016). DOI 10.1109/ENERGYCON.2016.7514029

43. Wu, J., Cui, Z., Chen, Y., Kong, D., Wang, Y.G.: A new hybrid model to predict the electrical load in five states of australia. Energy **166**, 598 – 609 (2019). DOI https://doi.org/10.1016/j.energy.2018.10. 076. URL `http://www.sciencedirect.com/science/article/pii/S0360544218320668`

44. Xu, S., Chan, H.K., Zhang, T.: Forecasting the demand of the aviation industry using hybrid time series sarima-svr approach. Transportation Research Part E: Logistics and Transportation Review **122**, 169 – 180 (2019). DOI https://doi.org/10.1016/j.tre.2018.12. 005. URL `http://www.sciencedirect.com/science/article/pii/S1366554518308330`

45. Yang, H., Pan, Z., Tao, Q., Qiu, J.: Online learning for vector autoregressive moving-average time series prediction. Neurocomputing **315**, 9 – 17 (2018). DOI https://doi.org/10.1016/j.neucom.2018.04. 011. URL `http://www.sciencedirect.com/science/article/pii/S0925231218304259`

46. Zhang, G.: Time series forecasting using a hybrid arima and neural network model. Neurocomputing **50**, 159 – 175 (2003). DOI https://doi.org/10.1016/S0925-2312(01)00702-0. URL `http://www.sciencedirect.com/science/article/pii/S0925231201007020`

47. Zhang, G.P.: An investigation of neural networks for linear time-series forecasting. Computers & Operations Research **28**(12), 1183 – 1202 (2001). DOI https://doi.org/10.1016/S0305-0548(00)00033-2. URL `http://www.sciencedirect.com/science/article/pii/S0305054800000332`

48. Zinkevich, M.: Online convex programming and generalized infinitesimal gradient ascent. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003) pp. 928–936 (2003). URL `https://www.aaai.org/Papers/ICML/2003/ICML03-120.pdf`. Progress in Modeling, Theory, and Application of Computational Intelligenc
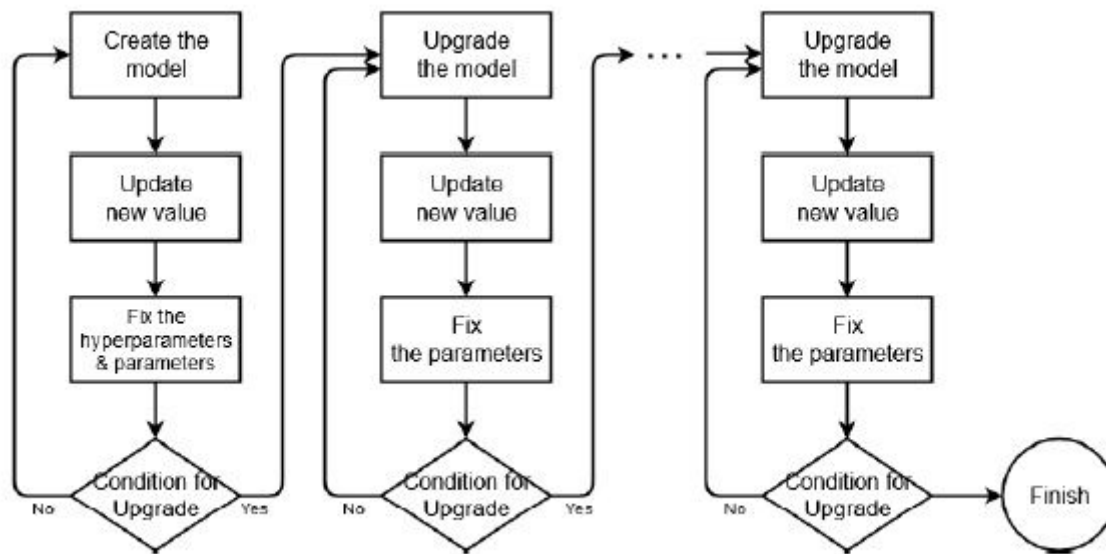
Figures

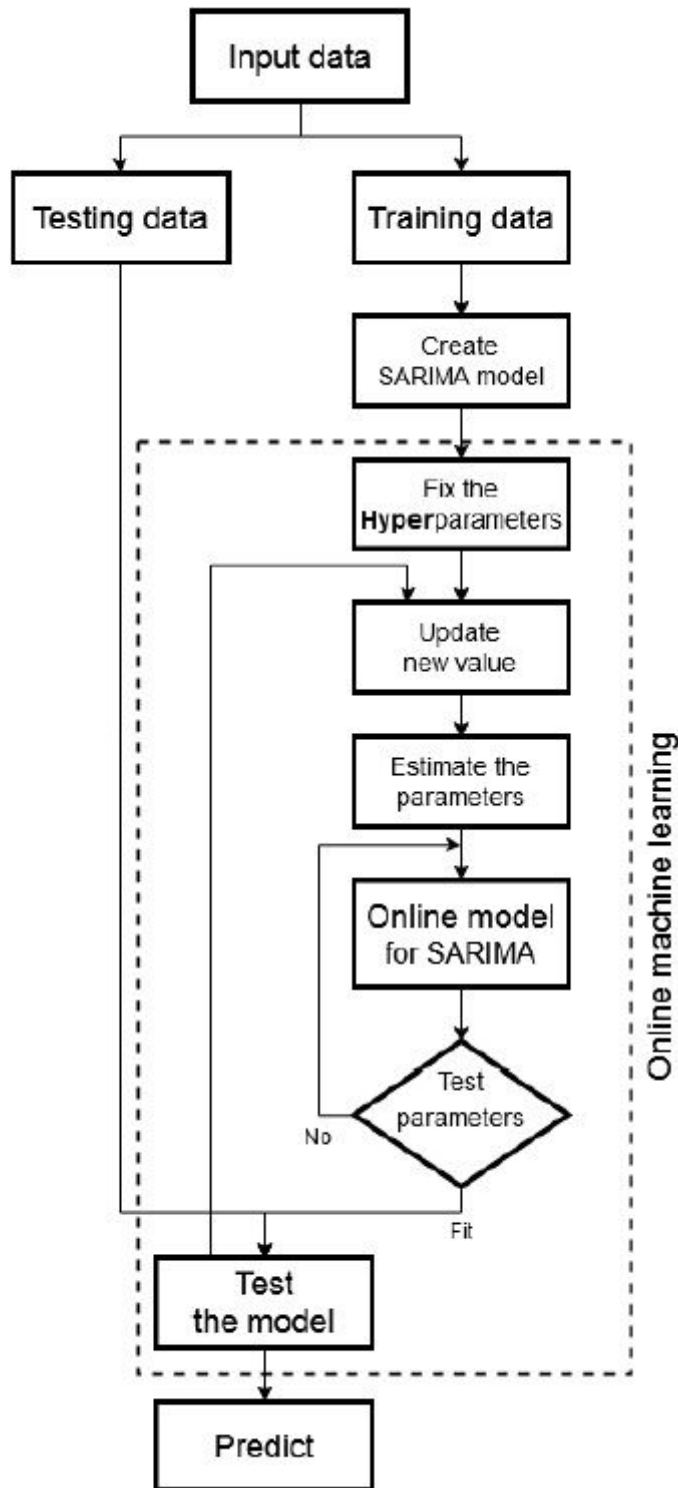**Figure 1**

The diagram of the online machine learning model.

Figure 2

The Online SARIMA model.
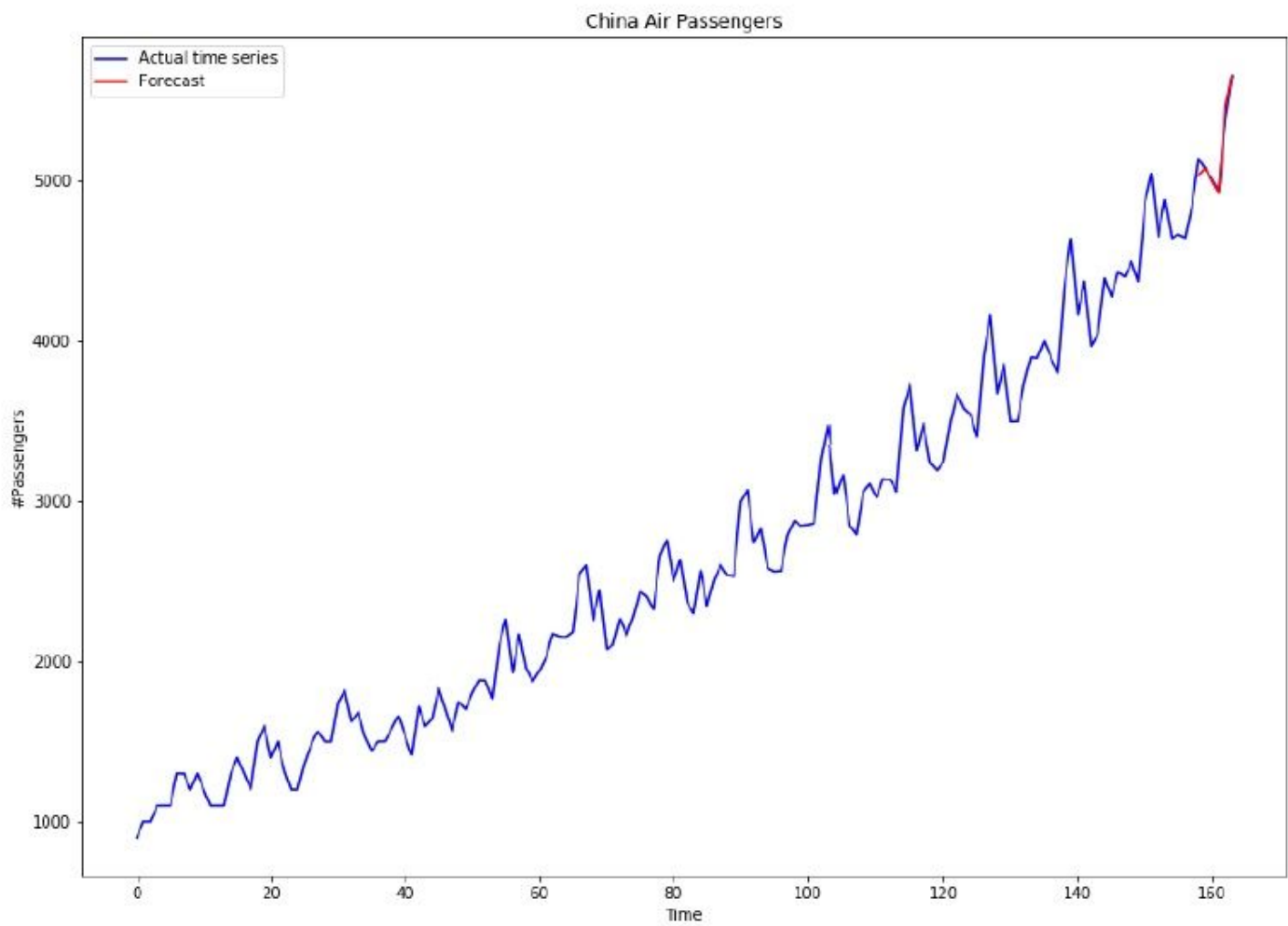
China Air Passengers

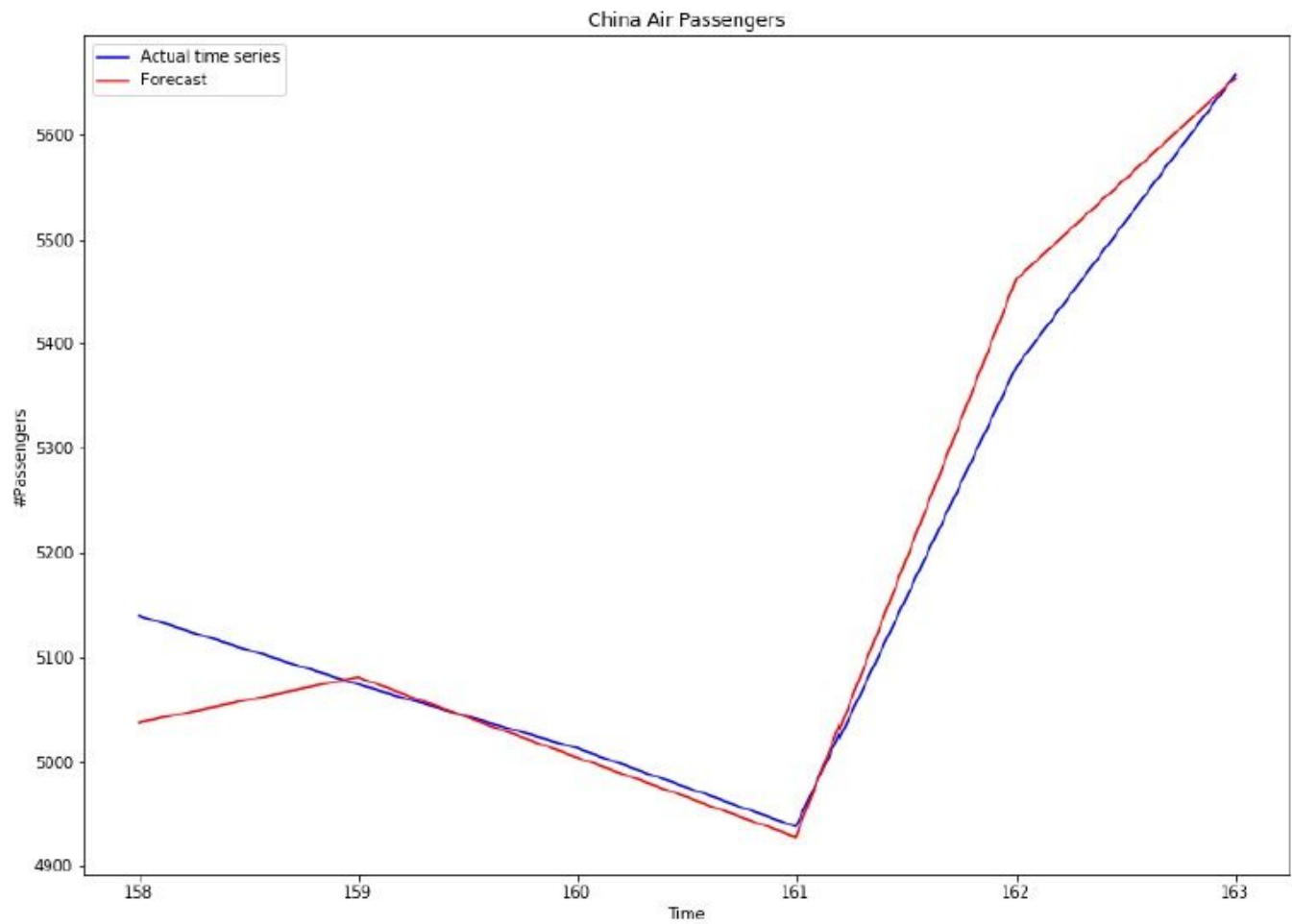**Figure 3**

The results of the SARIMA-ONS model.

Figure 4
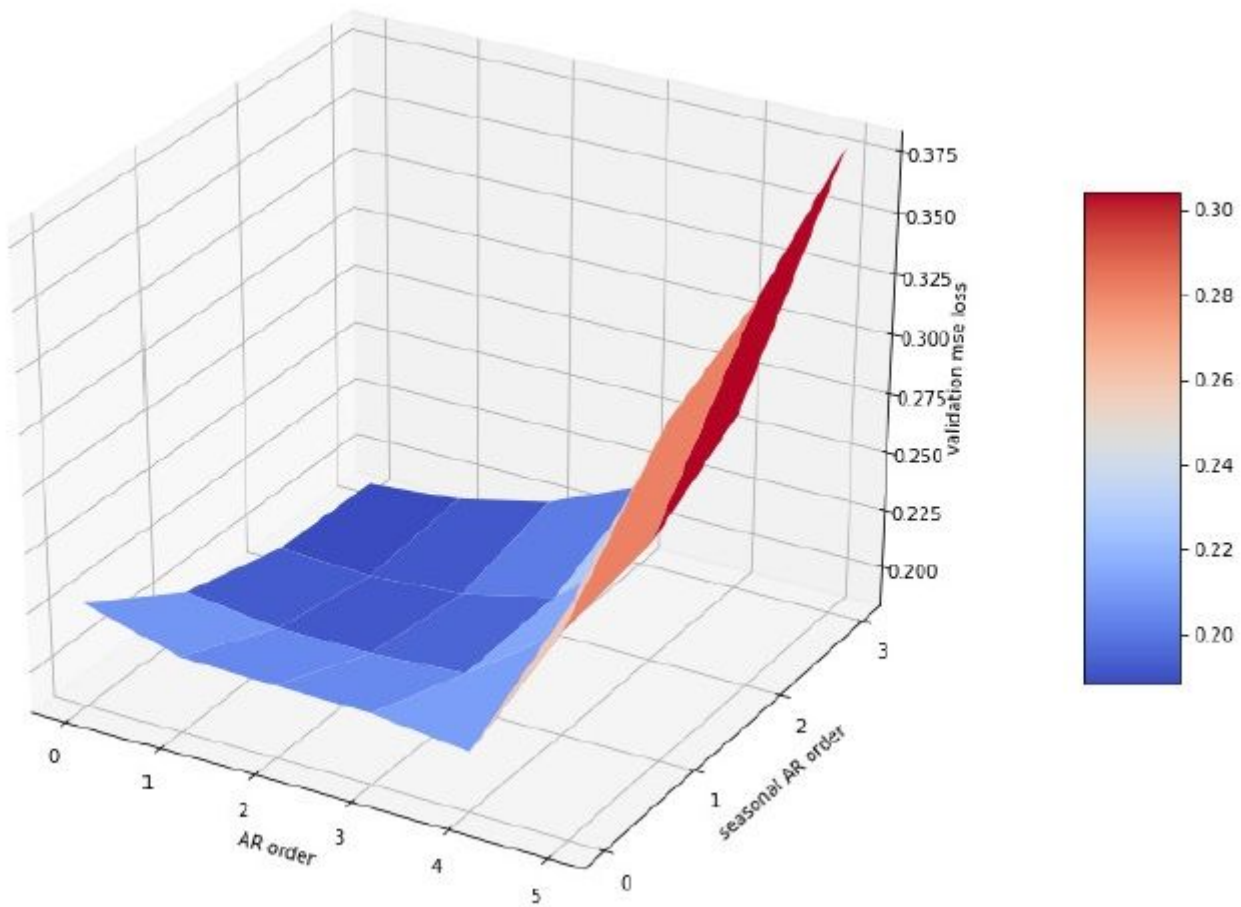
The results of the SARIMA-ONS model.

Figure 5

Optimization hyper parameters Auto-regression and seasonal auto-regression for SARIMA online model at 05 o'clock AM being on behalf on low electricity load consumption.The best hyper-parameters (p; P) = (1; 2).
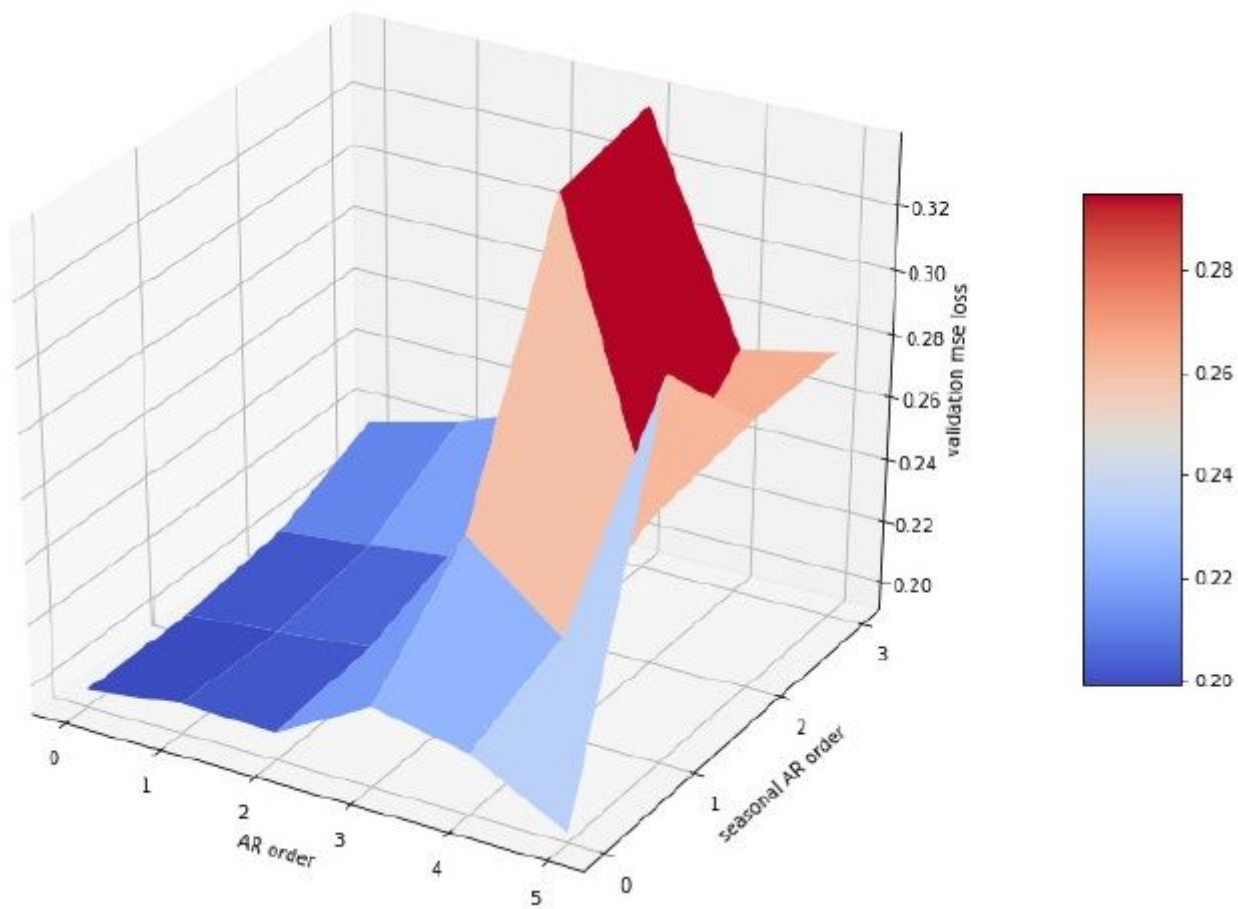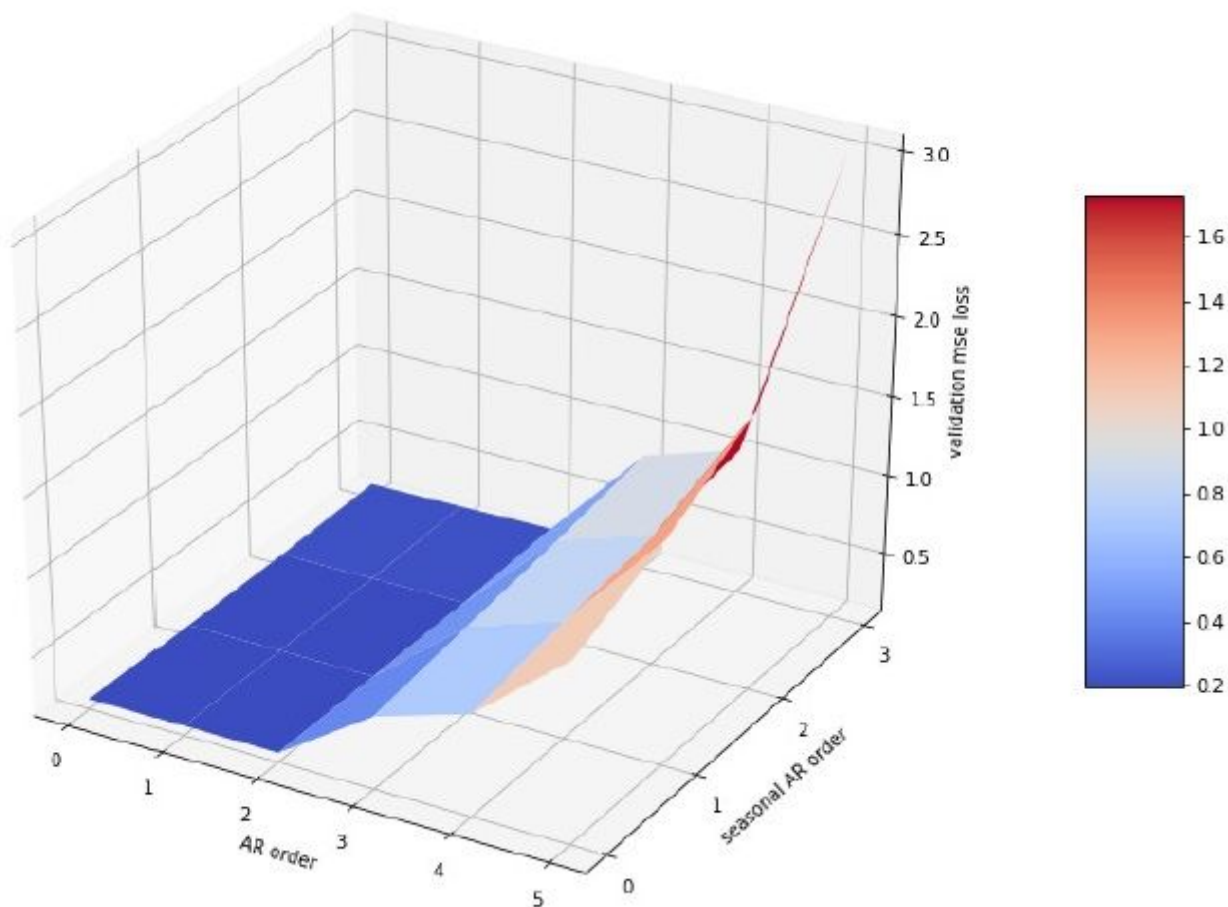
Figure 6

Optimization hyper parameters Auto-regression and seasonal auto-regression for SARIMA online model at 13 o'clock PM being on behalf on high electricity load consumption. The best hyper-parameters (p; P) = (0; 5).

Figure 7

Optimization hyper parameters Auto-regression and seasonal auto-regression for SARIMA online model at 21 o'clock PM being on behalf on medium electricity load consumption.The best hyper-parameters (p; P) = (2; 0).
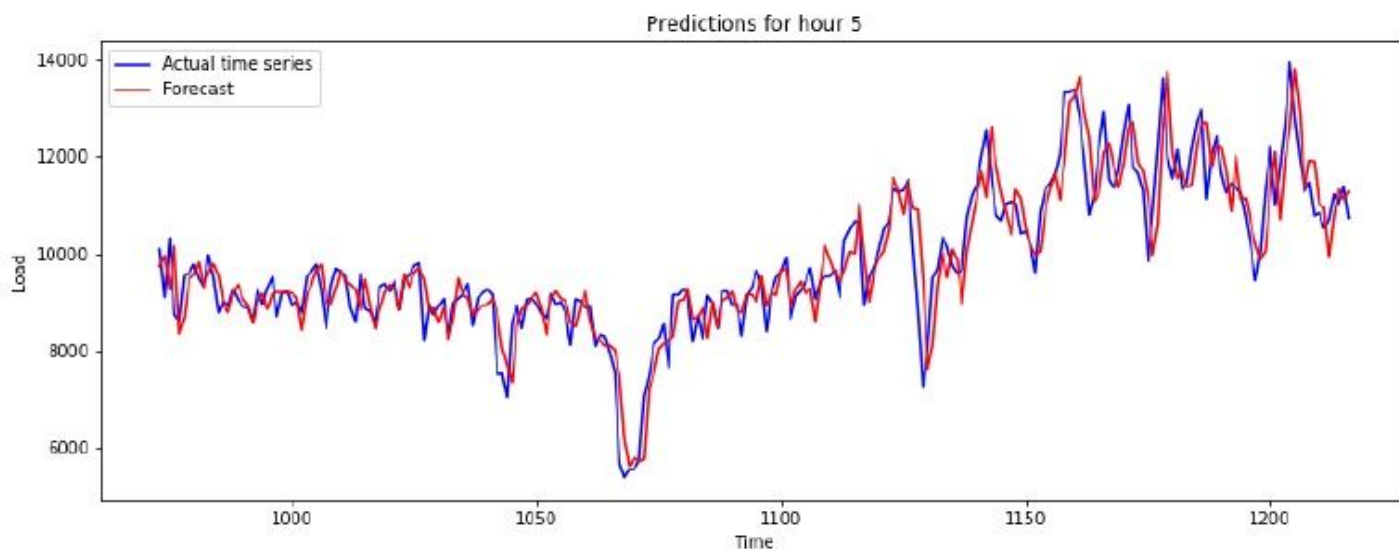
Figure 8

The results forecast at 05 o'clock AM of the SARIMA-ONS model for Electricity load data.
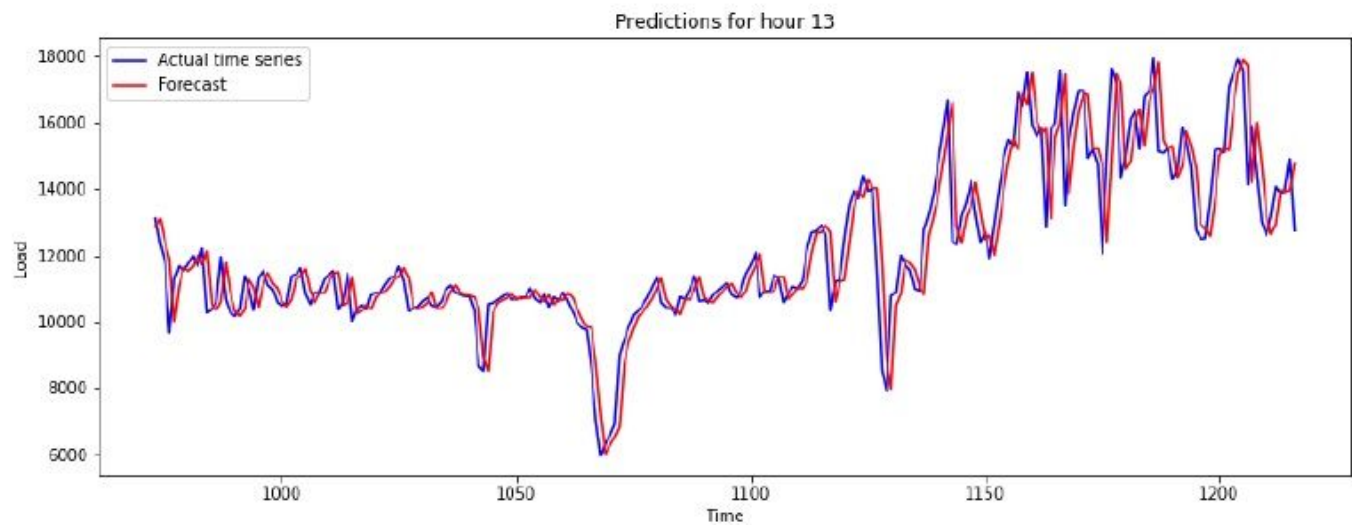

Figure 9

The results forecast at 13 o'clock PM of the SARIMA-ONS model for Electricity load data.
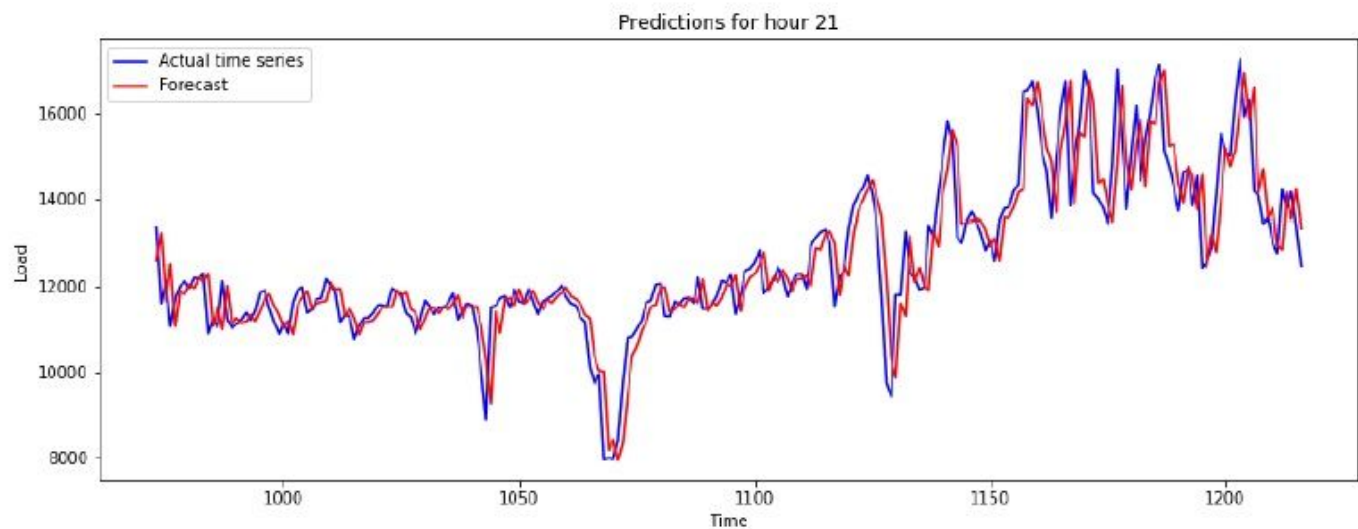

Figure 10

The results forecast at 21 o'clock PM of the SARIMA-ONS model for Electricity load data.