Journal of Mathematical Sciences and Optimization. e-ISSN: 3025-3926
Vol. 2, No. 2 (2025), pp. 144–160.

Exploring Multiple Seasonal MA Models for Short-Term Load Forecasting

SYALAM ALI WIRA DINATA SIMATUPANG[1*], MOH DANIL HENDRY GAMAL[2]

[1] STATISTICS STUDY PROGRAM, DEPARTMENT OF SCIENCE AND DATA ANALYTICS, FACULTY OF SCIENCE AND INFORMATION TECHNOLOGY (FSTI), KALIMANTAN INSTITUTE OF TECHNOLOGY, JL. SOEKARNO HATTA NO.KM 15, KARANG JOANG, BALIKPAPAN CITY, EAST KALIMANTAN 76127, INDONESIA.
*CORRESPONDING AUTHOR'S EMAIL: SALAM_ALI_WIRA_DINATA@LECTURER.ITK.AC.ID
[2] DEPARTMENT OF MATHEMATICS, FACULTY OF MATHEMATICS AND NATURAL SCIENCES, UNIVERSITY OF RIAU, PEKANBARU, INDONESIA. EMAIL: MDHGAMAL@UNRI.AC.ID

Abstract
This study explores the use of multiple seasonal Moving Average (MA) models for short-term load forecasting, focusing on identifying the most suitable model order, which may involve subset, multiplicative, or additive components. While many seasonal MA models for time series forecasting tend to assume non-multiplicative structures, often without performing statistical tests, this research introduces a new procedure to determine the most appropriate multiple MA order. The study includes a case analysis of short-term load forecasting in a specific country. The findings of the study indicate that incorporating multiple multiplicative parameters can significantly improve model accuracy.
Keywords : multiple seasonal, MA, subset, multiplicative, additive

1. INTRODUCTION

Box-Jenkins initially presented the IMA approach, which has since grown to be the most widely used model for predicting univariate time series data [1]. This model has been originated from the moving average model (MA) and the combination of the differencing and MA, or IMA models. To accommodate various real-world data scenarios, the moving average (MA) model and the integrated with MA (IMA) model can be modified to differing degrees. When two seasonal elements are included in the model, it is called the multiple or double seasonal MA model. If we define a nonseasonal moving average operator of order q by $\theta_q(B) = \left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)$, first seasonal moving average operator of order Q_1 by $\Theta_{Q_1}\left(B^{S_1}\right) = \left(1-\Theta_1 B-\Theta_2 B^{2S_1}-\cdots-\Theta_{Q_1} B^{Q_1 S_1}\right)$, second seasonal moving average operator of order Q_2 by $\Psi_{Q_2}\left(B^{S_2}\right) = \left(1-\Psi_1 B-\Psi_2 B^{2S_2}-\cdots-\Psi_Q B^{Q_2 S_2}\right)$. Economically, the nonseasonal and seasonal moving average model can be expressed as $\mathscr{L}_t = \theta_q(B)\Theta_{Q_1}\left(B^{S_1}\right)\Psi_{Q_2}\left(B^{S_2}\right)\epsilon_t$. it contain $q+2, Q_1+2, Q_2+2$ unknown parameters $\mu, \theta_1, \cdots, \theta_q \Theta_1, \cdots, \Theta_{Q_1}$ and $\Psi_1, \cdots, \Psi_{Q_2}, \sigma_\varepsilon^2$ which in practice have to be estimated from the data. In the case where seasonal components are included in this model, then the model is called as the Seasonal Integrated MA model.

Box-Jenkins procedure that contains three main stages to build an MA model, that is model identification, model estimation and model checking, is usually used for determining the best MA model for certain time series data. Seasonal IMA models $\nabla^d \nabla_{1440}^{D_1} \nabla_{10080}^{D_2} = (1-B)\left(1-B^{1440}\right)\left(1-B^{10080}\right)$ have been applied in many different forecasting domains up to this point. For load forecasting, a variety of forecasting methods have been employed, with varying degrees of success. In forecasting research, double SARIMA models are frequently utilized as benchmarks. e.g., [2] - [5].

Various forecasting methods have been applied in different sectors, with notable studies highlighting the use of SARIMA and machine learning approaches. Suhartono [6] explored time series

Received: 04-03-2025, accepted: 06-05-2025.

forecasting using seasonal autoregressive integrated moving average (SARIMA) models, Voronin [7] focused on electrical load and solar generation forecasting in hybrid energy systems, Bala [8] used machine learning and hybrid approaches to forecast energy consumption in the United Kingdom, Wang et al. [9] analyzed electricity price instability over time using time series methods, while Debnath and Mourshed [10] reviewed various forecasting methods in energy planning models. For instance, Kramar and Alchakov [11] evaluated machine learning techniques for forecasting seasonal time series. Wu et al. [12] employed a hybrid SARIMA-LSTM approach for forecasting daily tourist arrivals. Similarly, Dinata et al. [13] compared short-term load forecasting methods using data from Kalimantan. In the energy sector, Peshkov and Alsova [14] explored autoregressive models for forecasting active energy complexes. Durova et al. [15] applied ensemble algorithms for short-term electricity price forecasting. In another study, Durova et al. [16] also employed autoregressive models to forecast active energy complexes. Petrusevich [17] explored time series forecasting using high-order ARIMA functions for various applications, Sadaei et al. [18] introduced a STLF method based on fuzzy time series, seasonality, and long memory processes. Hachim [19] applied time series approaches to predict electricity production and demand in Russia. Lastly, Nigam et al. [20] used time series modeling for river runoff forecasting.

Although numerous studies have concentrated on model estimation, the identification stage is actually the most vital part in constructing MA models. Improper model identification can result in errors during the estimation phase and lead to higher costs for re-identification. In particular, for MA models, most previous research has generally relied on the multiplicative model without initially testing the significance of the multiplicative parameter. This assumption implies that the multiplicative MA model presumes a significant relationship between the non-seasonal and seasonal parameters. Additionally, widely used statistical software like MINITAB and SPSS often only offer tools for fitting the multiplicative model.

The main contributions of this study include the application of a fast-processing statistical model, the Double Seasonal IMA, which has not been previously used on actual electricity load data. The model is applied to a real dataset from a case study in Poland. The study demonstrates that the Double Seasonal IMA model is more effective in predicting electrical demand data with a double seasonal pattern and associated errors. By adhering to the principle of parsimony, the model offers a cost-effective solution for electrical data operators, making it highly suitable for rapid electricity load forecasting.

2. RESEARCH METHODS

2.1. Procedures of Modeling MA

Most previous studies typically used the regular multiplicative MA model directly when the ACF and PACF indicated that the data contained both non-seasonal and double seasonal orders. In this study, we develop and consider a more precise model identification step, particularly at lags 1440 and 10080, as an implication of double seasonal orders for additive, subset, and multiplicative models on load data. As an example, it may appear that using the option Q=(1 10080) is the solution, but the MA(1) process also affects past values; what is actually needed are the MA parameters at lags 1, 10080, and 10081. One option is to define a subset model with distinct parameters at these lags, or alternatively, a factored model can be specified, representing the model as the product of an MA(1) model and an MA(10080) model.

2.2. Sample Autocorrelations (ACF) and Partial ACF

The autocorrelation plot illustrates the correlation between the current values of the series and its past values. For the selected model, the parameters are estimated based on the data. These plots are referred to as autocorrelation functions because they illustrate the extent of correlation with previous values of the series, based on the number of periods in the past (or lag) at which the correlation is calculated.

The output of the sample autocorrelation function plot and sample partial autocorrelation function plot from the identify statement is displayed in Figure 2.

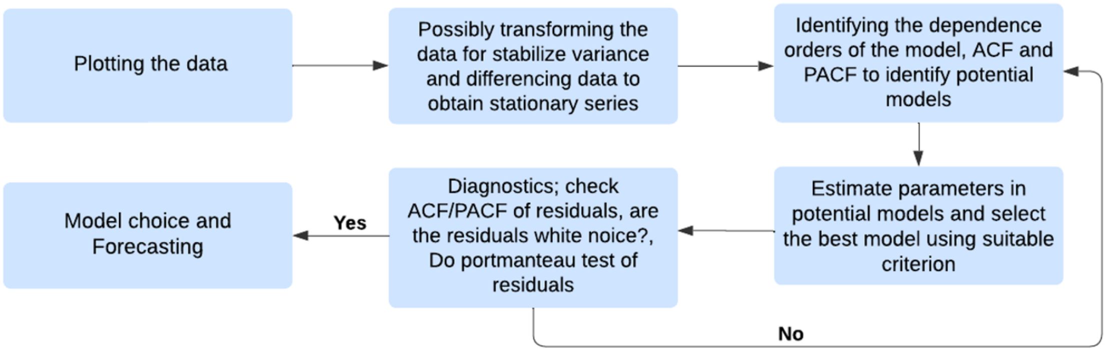


FIGURE 1. Box-Jenkins Methodology

TABLE 1. Behavior of the ACF and PACF for values at nonseasonal ARMA Models

	for Causal and Invertible ARMA Models		(p,q)
ARMA Model	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag(q)	Tails off
PACF	Cuts off after lag (p)	Tails off	Tails off

TABLE 2. Behavior of the ACF and PACF for values at seasonal ARMA Models

	for Causal and Invertible Seasonal ARMA Models		(P,Q)
ARMA Model	AR(P)s	MA(Q)s	ARMA(P,Q)
ACF	Tails off at lags ks, k=1,2,...,	Cuts off after lag Q_1s_1, Q_2s_2	Tails off at lags ks
PACF	Cuts off after lag P_1s_1, P_2s_2	Tails off at lags ks	Tails off at lags ks

3. RESULTS AND DISCUSSION

3.1. Data

The data used in this study was provided by the National Grid, which is the transmission company in Poland. The data consists of the 30 weeks of minute-by-minute observations for electricity demand in Poland from 1 January 2015 to 30 October 2020. The series consists of 201598 observations, and is shown in Fig. 3. This gave 201,600 minute-by-minute observations for estimation, and 100,800 for evaluation.

3.2. Identification

The Box-Jenkins method employs two graphical tools for modeling traditional statistical models: the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

LISTING 1. language=R

```
# Model1 for subset MA
# Model2 for multiplicative MA: use theta2
# Model3 for additive MA: use theta3

# Define the parameters for each model
theta1 <- c(-0.8, rep(0, 10), -0.9, 0.3)
theta2 <- c(-0.8, rep(0, 10), -0.9, 0.54)
theta3 <- c(-0.8, rep(0, 10), -0.9)
```

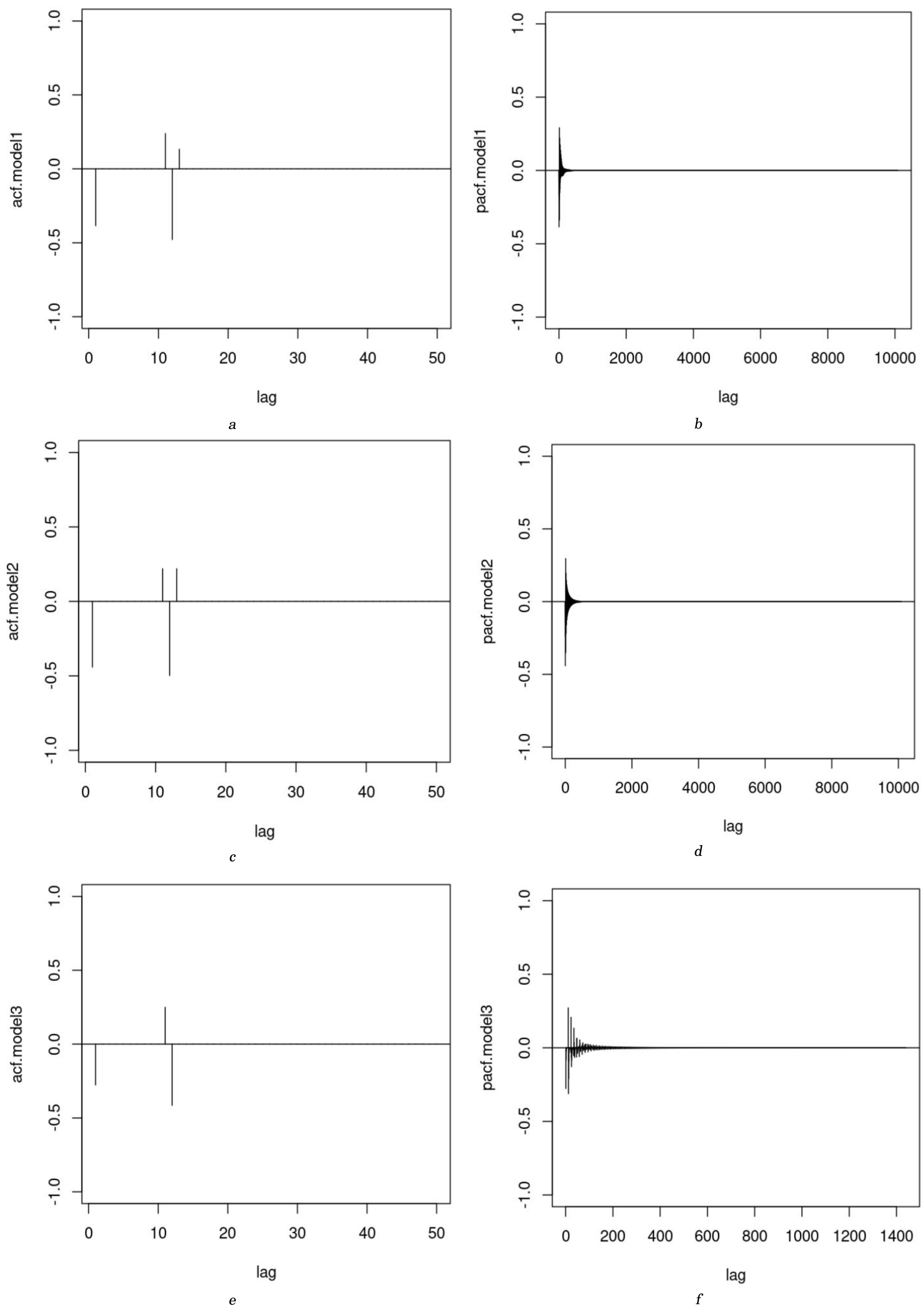
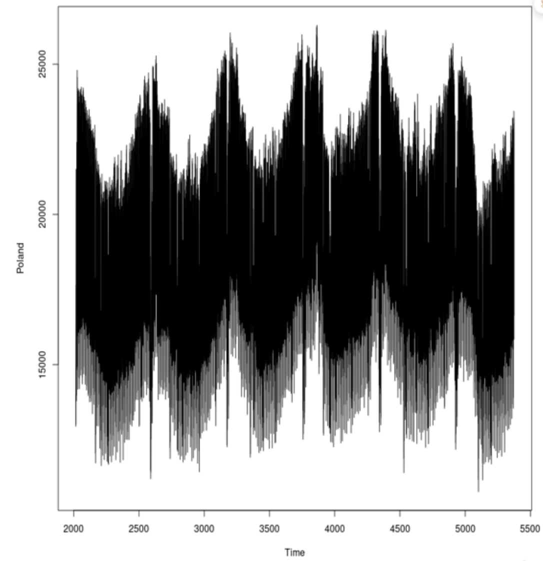
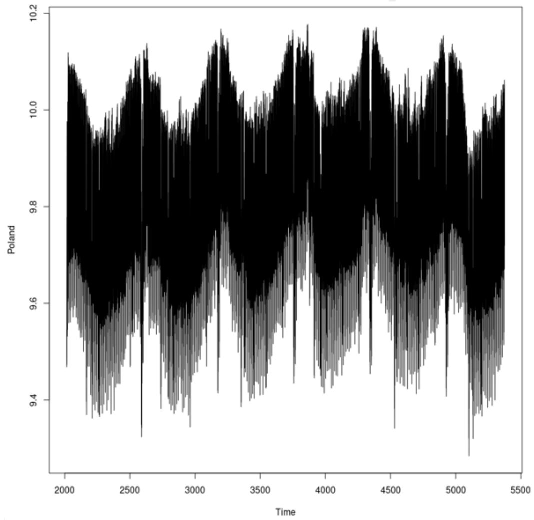


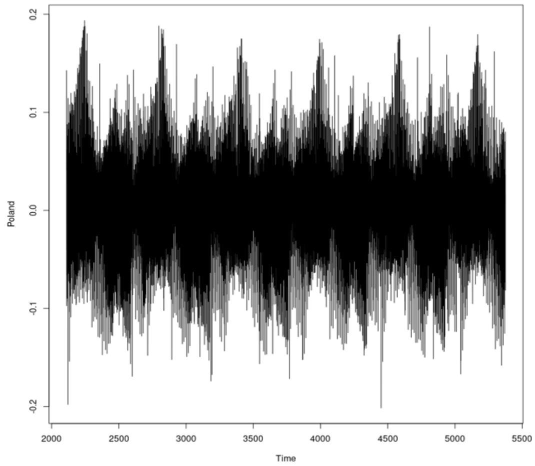
FIGURE 2. Previous ACF Model 3 for the subset with truncation after lag, ACF Model 1 for the subset with truncation after lag q (*a*), PACF Model 1 for the subset with exponential decay q (*b*), ACF Model 2 for the multiplicative model with a break after lag q (*c*), PACF Model 2 for the multiplicative model with exponential decay break (*d*), and PACF Model 2 for the additive model with the tail of exponential decay (*e*) PACF Model 2 for the additive model with tail exponential decay



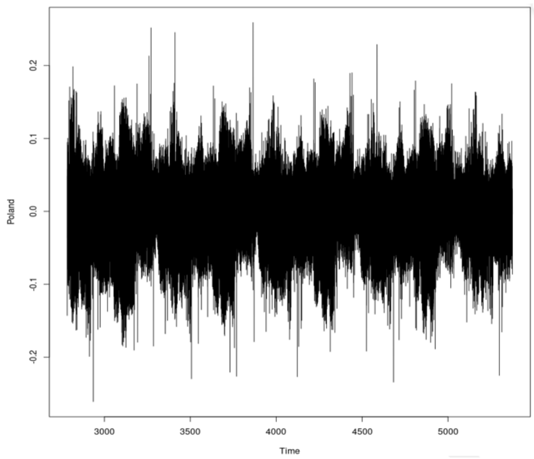
a



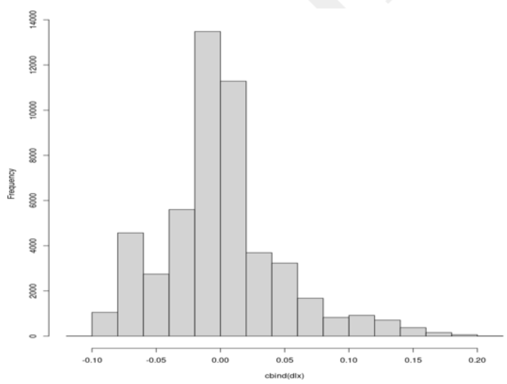
b



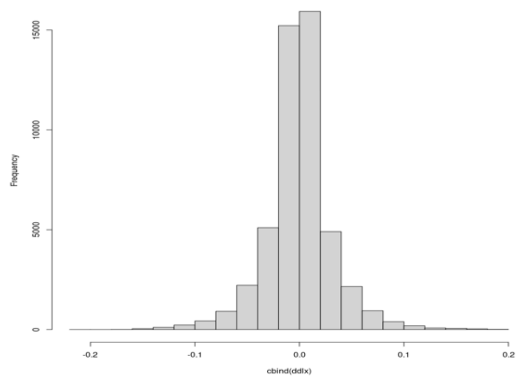
c



d



e



e

FIGURE 3. Load demand series (*a*), Load demand series after Ln (*b*), Load demand series before differencing with Ln (*c*), Load demand series after differencing with Ln (*d*), and Histogram plot series after Ln (*e*) Histogram plot series after Ln with Differencing

```
# Compute the ACF and PACF for model1 using theta1
acf.model1 <- ARMAacf(ar = 0, ma = theta1, 10080)
pacf.model1 <- ARMAacf(ar = 0, ma = theta1, 10080, pacf = TRUE)

# Trim the first value from ACF for model1 (ACF is zero at lag 0)
acf.model1 <- acf.model1[2:10080]

# Create the plots
win.graph() # Open a new graphical window
par(mfrow = c(1, 2)) # Set the layout to 1 row and 2 columns

# Plot the ACF
plot(acf.model1, type = "h", xlab = "Lag", ylim = c(-1, 1))
abline(h = 0) # Add a horizontal line at y = 0

# Plot the PACF
plot(pacf.model1, type = "h", xlab = "Lag", ylim = c(-1, 1))
abline(h = 0) # Add a horizontal line at y = 0
```

Plot autocorrelations and limits might model load as a combination of an MA (1) process reflecting short term dependencies and an MA (1440) model reflecting the seasonal pattern. It might seem that the way to do this is with the option Q = (1 1440 1441 10080 10081 11520 11521), but the MA (1) process also operates in past years, need MA parameters at lags 1 1440 1441 10080 10081 11520 and 11521. We can specify a subset model with separate parameters at these lags, or can specify a factored model that represents the model as the product of Q = (1 1440 1441 10080 10081 11520 11521) model and MA (1) model.

Identifying the dependence orders of the autocovariances generating function models for autocovariances generating function of a MA with $q = (1)\, Q = (1)\,(1440)\,(10080)$

$$\begin{aligned}\mathscr{L}_t = \sigma_\epsilon^2 (1-\theta_1\mathscr{L} - \theta_{1440}\mathscr{L}^{1440} + \theta_1\theta_{1440}\mathscr{L}^{1441} - \Theta_{10080}\mathscr{L}^{10080} + \theta_1\Theta_{10080}\mathscr{L}^{10081} + \theta_{1440}\Theta_{10080}\mathscr{L}^{11520}\\ -\theta_1\theta_{1440}\Theta_{10080}\mathscr{L}^{11521}\big)\big(1-\theta_1\mathscr{L}^{-1} - \theta_{1440}\mathscr{L}^{-1440} + \theta_1\theta_{1440}\mathscr{L}^{-14441} - \Theta_{10080}\mathscr{L}^{-10080}\\ +\theta_1\Theta_{10080}\mathscr{L}^{-10081} + \theta_{1440}\Theta_{10080}\mathscr{L}^{-11520} - \theta_1\theta_{1440}\Theta_{10080}\mathscr{L}^{-11521}\big)\end{aligned}$$

Identifying the dependence orders of the autocovariances generating function models for autocovariances generating function of MA $q = (1)$ Q = (1 1440 1441 10080 10081 11520 11521) is of the process is

$$\begin{aligned}\mathscr{L}_t = \sigma_\epsilon^2 (1-(\theta_1 + \Theta_1)\mathscr{L} + \theta_1\Theta_1\mathscr{L}^2 - \theta_{1440}\mathscr{L}^{1440} + (\theta_1\theta_{1440} + \theta_{1440}\Theta_1 + \theta_1\theta_{1440})\mathscr{L}^{1441} - \theta_1\theta_{1440}\Theta_1\mathscr{L}^{1442}\\ - \Theta_{10080}\mathscr{L}^{10080} + (\Theta_1\Theta_{10080} + \theta_1\Theta_{10080})\mathscr{L}^{10081} - \theta_1\Theta_1\Theta_{10080}\mathscr{L}^{10082} + \theta_{1440}\Theta_{10080}\mathscr{L}^{11520}\\ - (\theta_1\Theta_{1440}\Theta_{10080} + \theta_{1440}\Theta_1\Theta_{10080})\mathscr{L}^{11521} + \theta_1\theta_{1440}\Theta_1\Theta_{10080}\mathscr{L}^{11522}\big)\big(1-(\theta_1+\Theta_1)\mathscr{L}^{-1}\\ + \theta_1\Theta_1\mathscr{L}^{-2} - \theta_{1440}\mathscr{L}^{-1440} + (\theta_1\theta_{1440} + \theta_{1440}\Theta_1 + \theta_1\theta_{1440})\mathscr{L}^{-1441} - \theta_1\theta_{1440}\Theta_1\mathscr{L}^{-1442}\\ - \Theta_{10080}\mathscr{L}^{-10080} + (\Theta_1\Theta_{10080} + \theta_1\Theta_{10080})\mathscr{L}^{-10081} - \theta_1\Theta_1\Theta_{10080}\mathscr{L}^{-10082}\\ + \theta_{1440}\Theta_{10080}\mathscr{L}^{-11520} - (\theta_1\theta_{1440}\Theta_{10080} + \theta_{1440}\Theta_1\Theta_{10080})\mathscr{L}^{-11521}\\ + \theta_1\theta_{1440}\Theta_1\Theta_{10080}\mathscr{L}^{-11522}\big)\end{aligned}$$

Identifying the dependence orders of the autocovariances generating function models for autocovariances generating function of MA $q = (1)$ Q = (1 1440 1441 10080 10081) is of the process is

$$
\begin{aligned}
\mathscr{L}_t = &(1-(\Theta_1+\ \theta_1)\mathscr{L}^1-(\theta_{1440})\mathscr{L}^{1440}-\Theta_{10080}\mathscr{L}^{10080}+(\theta_1\Theta_1)\mathscr{L}^2+(\Theta_1\Theta_{10080}+\theta_1\Theta_{10080})\mathscr{L}^{10081}\\
&+(\Theta_1\Theta_{1440}+\theta_{1440}\theta_1)\mathscr{L}^{1441}-\Theta_{10080}\Theta_1\theta_1\mathscr{L}^{10082}+\theta_{1440}\Theta_{10080}\mathscr{L}^{11520}\\
&-(\Theta_1\theta_{1440}\Theta_{10080}+\Theta_{10080}\theta_{1440}\theta_1)\mathscr{L}^{11521}+\theta_1\Theta_{1440}\Theta_1\Theta_{10080}\mathscr{L}^{11522}\big)\left(1-(\Theta_1+\theta_1)\mathscr{L}^{-1}\right.\\
&+(\theta_1\Theta_1)\mathscr{L}^{-2}-(\theta_{1440})\mathscr{L}^{-1440}-\Theta_{10080}\epsilon_{t-10080}\mathscr{L}^{-10080}+(\Theta_1\Theta_{10080}+\Theta_1\Theta_{10080})\mathscr{L}^{-10081}\\
&+(\Theta_1\theta_{1440}+\theta_{1440}\theta_1)\mathscr{L}^{-1441}-\Theta_{10080}\Theta_1\theta_1\mathscr{L}^{-10082}+\theta_{1440}\Theta_{10080}\mathscr{L}^{-11520}\\
&\left.-(\Theta_1\theta_{1440}\Theta_{10080}+\Theta_{10080}\theta_{1440}\theta_1)\mathscr{L}^{-11521}+\theta_1\theta_{1440}\Theta_1\Theta_{10080}\mathscr{L}^{-11522}\right)
\end{aligned}
$$

The stationarity of the time series is tested as the first step in the modelling procedure. To acquire a fair estimate of stationarity, utilise the partial auto correlation function (PACF) and auto correlation function (ACF) plots of the time series.

a. Option $Q = (1\ 1440\ 1441)$; MA Operator $\left(1-\Theta_1 B-\Theta_{1440}B^{1440}-\Theta_{1441}B^{1441}\right)$, ACF cuts off after lag Qs and PACF tails off at lags $ks\,(k=1,1440,1441)$ and since $\{\epsilon_t\}$ is white noise the expected value of the MA process is $E(\mathscr{L}_t)=E\left(\mu+\epsilon_t-\Theta_1\epsilon_{t-1}-\Theta_{1440}\epsilon_{t-1440}+\Theta_{1441}\epsilon_{t-1441}\right)=\mu$, also its variance is $\mathrm{Var}(\mathscr{L}_t)=\gamma_{\mathscr{L}_t}(0)=\mathrm{Var}\left(\mu+\epsilon_t-\Theta_1\epsilon_{t-1}-\Theta_{1440}\epsilon_{t-1440}+\ \Theta_{1441}\epsilon_{t-1441}\right)=\sigma^2\left(1+\Theta_1^2+\Theta_{1440}^2+\Theta_{1441}^2\right)$,

b. Option $Q=(1,1440)$; MA Operator $\left(1-\Theta_1 B-\Theta_{1440}B^{1440}\right)$, ACF cuts off after lag Qs and PACF Tails off at lags $ks\,(k=1,1440)$ and $E(\mathscr{L}_t)=E\left(\mu+\epsilon_t-\Theta_1\epsilon_{t-1}-\ \Theta_{1440}\epsilon_{t-1440}\right)=\mu$, $\mathrm{Var}(\mathscr{L}_t)=\gamma_{\mathscr{L}_t}(0)=\mathrm{Var}\left(\mu+\epsilon_t-\Theta_1\epsilon_{t-1}-\ \Theta_{1440}\epsilon_{t-1440}\right)=\sigma^2\left(1+\Theta_1^2+\Theta_{1440}^2\right)$,

c. Option $Q=(1)(1440)$; MA Operator $(1-\Theta_1 B)\left(1-\Theta_{1440}B^{1440}\right)$, ACF cuts off after lag Qs and PACF Tails off at lags $ks\,(k=(1)(1440))$ and $E(\mathscr{L}_t)=E\left(\mu+\epsilon_t-\ \Theta_1\epsilon_{t-1}-\Theta_{1440}\epsilon_{t-1440}\right)=\mu$, $\mathrm{Var}(\mathscr{L}_t)=\gamma_{\mathscr{L}_t}(0)=\mathrm{Var}\left(\mu+\epsilon_t-\Theta_1\epsilon_{t-1}-\ \Theta_{1440}\epsilon_{t-1440}\right)=\sigma^2\left(1+\Theta_1^2+\Theta_{1440}^2+(\Theta_1\Theta_{1440})^2\right)$,

The ACF measures the correlation of a time series value with other values from the same time series at various delays. PACF evaluates the connection between a time series value and a value with a different lag. The ACF is derived from the linear correlation between each value of the time series $\mathscr{L}_t$ and the other values at different lags, such as $\mathscr{L}_{t-1},\mathscr{L}_{t-1440},\mathscr{L}_{t-10080},\mathscr{L}_{t-10081},\mathscr{L}_{t-11520},\mathscr{L}_{t-11521}$, and so on. On the other hand, the PACF yields a similar result but eliminates the influence of other values. For instance, in the ACF, the correlation between $\mathscr{L}_t$ and $\mathscr{L}_{t-2}$ is affected by $\mathscr{L}_{t-1}$, but the PACF removes this influence. Both the ACF and PACF have individual interpretations, as well as a combined interpretation when considered together. The types of MA models were chosen to verify the significance of a multiplicative parameter.

LISTING 2. Double Seasonal MA Model Selection Algorithm in R

```
# 1. Fit the subset double seasonal MA model
# 2. Test the double seasonal multiplicative parameters significance
if (multiplicative_param_significant) {
# 3. Test if double seasonal multiplicative parameters equal
    the product of non—seasonal and seasonal coefficients
    if (multiplicative_param == (non_seasonal_coef * seasonal_coef)) {
# 4. Check if double seasonality is present
        if (is_double_seasonal) {
# 5. Choose Double Seasonal Multiplicative MA as
            the appropriate model
            model <— "Double Seasonal Multiplicative MA"
        } else {
# 6. Choose double seasonal Multiplicative MA as
            the appropriate model
            model <— "Multiplicative MA"
        }
    } else {
```

```
# 7. Choose double seasonal Subset MA as the appropriate model
        model <- "Subset MA"
    }
} else {
# 8. Choose Additive MA as the appropriate model
    model <- "Additive MA"}
```

Initially, by focusing on the seasonal lags $s_1 = 1440$ and seasonal lags $s_2 = 10080$, the ACF and PACF characteristics of the series show a prominent peak at lag lags $k = 1s$ in the autocorrelation function, with smaller peaks at lags $k = 2s, 3s$, along with additional peaks at $k = 1s, 2s, 3s$, and $4s$ in the partial autocorrelation function.

It seems that either: (i) the ACF cuts off after lag 1s, while the PACF gradually decreases at the seasonal lags, (ii) the ACF cuts off after lag 3s, with the PACF tailing off at the seasonal lags, or (iii) both the ACF and PACF are gradually decreasing at the seasonal lags.

Referring to Table 2, this implies that either (i) additive, (ii) multiplicative, or (iii) subset. The theoretical discussion regarding ACF and PACF for these models concentrated on the orders of non-seasonal and seasonal moving averages, i.e.

$$\nabla^d \nabla_{1440}^{D_1} \mathscr{L}_t = \left(1 - \theta_1 B - \Theta_{1440} B^{1440} + \theta_1 \Theta_{1440} B^{1441}\right) \varepsilon_t \tag{1}$$

$$\nabla^d \nabla_{1440}^{D_1} \nabla_{10080}^{D_2} \mathscr{L}_t = \left(1 - \theta_1 B - \Theta_{1440} B^{1440} + \theta_1 \Theta_{1440} B^{1441}\right)\left(1 - \Psi_{10080} B^{10080}\right) \varepsilon_t \tag{2}$$

This section introduces several modifications to the IMA (Integrated Moving Average) model aimed at addressing seasonal and nonstationary patterns. Often, the influence of past values is most significant at multiples of an underlying seasonal lag s. The first step in the modeling process involves testing the stationarity of the time series. To obtain a reliable estimate of stationarity, it is recommended to use the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) plots of the time series.

Additive MA. The operator option $\hat{\mathscr{L}}_t = \nabla^d \nabla_{10080}^{D_1} \ln(\mathscr{L}_t - \mu)$ of the additive MA $q = (1, 10080)$, $Q = (1, 10080)$ as follows:

$$\hat{\mathscr{L}}_t = \left(1 - \theta_1 \mathscr{L}^1 - \theta_{10080} \mathscr{L}^{10080}\right)\left(1 - \Theta_1 \mathscr{L}^1 - \Theta_{10080} \mathscr{L}^{10080}\right) \varepsilon_t \tag{3}$$

Next, we collect like terms (i.e., terms with the same power of $\mathscr{L}$):

$$\zeta = \begin{cases} -(\Theta_1 + \theta_1), & k = 0, \\ -(\Theta_1 + \theta_1), & k = 1, \\ \theta_1 \Theta_1, & k = 2, \\ -(\Theta_{10080} + \theta_{10080}), & k = 10080, \\ -(\Theta_{10081} + \theta_{10081}), & k = 10081, \\ (\theta_1 \Theta_{10080} + \theta_{10080} \Theta_1), & k = 10081, \\ (\theta_1 \Theta_{10081} + \theta_{10081} \Theta_1), & k = 10082, \\ \theta_{10080} \Theta_{10080}, & k = 20160, \\ (\theta_{10080} \Theta_{10081} + \theta_{10081} \Theta_{10080}), & k = 20161, \\ \theta_{10081} \Theta_{10081}, & k = 20162, \end{cases} \tag{4}$$

$$\begin{aligned} \hat{\mathscr{L}}_t = \varepsilon_t - (\Theta_1 + \theta_1)\varepsilon_{t-1} + \theta_1 \Theta_1 \varepsilon_{t-2} - (\Theta_{10080} + \theta_{10080})\varepsilon_{t-10080} \\ - (\Theta_{10081} + \theta_{10081})\varepsilon_{t-10081} + (\theta_1 \Theta_{10080} + \theta_{10080} \Theta_1)\varepsilon_{t-10081} \\ + (\theta_1 \Theta_{10081} + \theta_{10081} \Theta_1)\varepsilon_{t-10082} \\ + \theta_{10080} \Theta_{10080} \varepsilon_{t-20160} \\ + (\theta_{10080} \Theta_{10081} + \theta_{10081} \Theta_{10080})\varepsilon_{t-20161} \\ + \theta_{10081} \Theta_{10081} \varepsilon_{t-20162} \end{aligned}$$

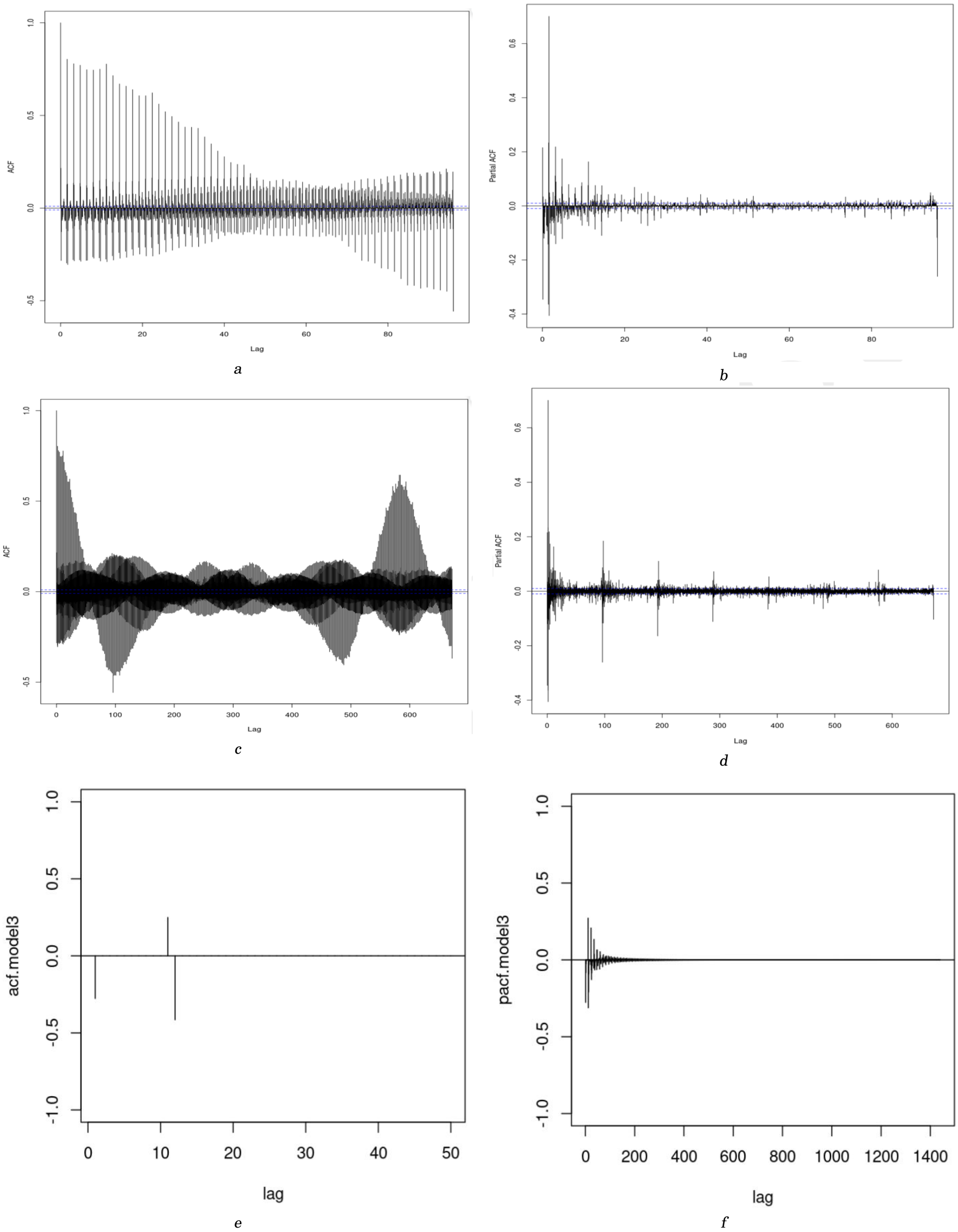


FIGURE 4. Previous ACF Model 3 for the subset with truncation after lag, ACF Model 1 for the subset with truncation after lag q (*a*), PACF Model 1 for the subset with exponential decay q (*b*), ACF Model 2 for the multiplicative model with a break after lag q (*c*), PACF Model 2 for the multiplicative model with exponential decay break (*d*), and PACF Model 2 for the additive model with the tail of exponential decay (*e*) PACF Model 2 for the additive model with tail exponential decay

Multiplicative MA. The option of the multiplicative operator $\hat{\mathscr{L}}_t = \nabla^d \nabla^{D_1}_{10080} \ln(\mathscr{L}_t - \mu)$ of the MA $q = (1)(10080)$, $Q = (1)(10080)$ as follows:

$$\hat{\mathscr{L}}_t = \left(1-\theta_1\mathscr{L}^1\right)\left(1-\theta_{10080}\mathscr{L}^{10080}\right)\left(1-\Theta_1\mathscr{L}^1\right)\left(1-\Theta_{10080}\mathscr{L}^{10080}\right)\varepsilon_t. \quad (5)$$

Now, we combine all these results:

$$\begin{aligned}\hat{\mathscr{L}}_t = \varepsilon_t - \Theta_{10080}\varepsilon_{t-10080} - \Theta_1\varepsilon_{t-1} + \Theta_1\Theta_{10080}\varepsilon_{t-10081} \\ -\theta_{10080}\varepsilon_{t-10080} + \theta_{10080}\Theta_{10080}\varepsilon_{t-20160} + \theta_{10080}\Theta_1\varepsilon_{t-10081} - \theta_{10080}\Theta_1\Theta_{10080}\varepsilon_{t-20161} \\ -\theta_1\varepsilon_{t-1} + \theta_1\Theta_{10080}\varepsilon_{t-10081} + \theta_1\Theta_1\varepsilon_{t-2} - \theta_1\Theta_1\Theta_{10080}\varepsilon_{t-10082} + \\ \theta_1\theta_{10080}\varepsilon_{t-10081} - \theta_1\theta_{10080}\Theta_{10080}\varepsilon_{t-20161} - \theta_1\theta_{10080}\Theta_1\varepsilon_{t-10082} + \theta_1\theta_{10080}\Theta_1\Theta_{10080}\varepsilon_{t-20162}\end{aligned}$$

Next, we collect like terms (i.e., terms with the same power of $\mathscr{L}$):

$$\zeta = \begin{cases} 1, & k=0, \\ -(\Theta_1+\theta_1), & k=1, \\ \theta_1\Theta_1, & k=2, \\ -(\Theta_{10080}+\theta_{10080}), & k=10080, \\ (\Theta_1\Theta_{10080}+\theta_{10080}\Theta_1+\theta_1\Theta_{10080}+\theta_1\theta_{10080}), & k=10081, \\ -(\theta_1\Theta_1\Theta_{10080}+\theta_1\theta_{10080}\Theta_1), & k=10082, \\ \theta_{10080}\Theta_{10080}, & k=20160, \\ -(\theta_{10080}\Theta_1\Theta_{10080}+\theta_1\theta_{10080}\Theta_{10080}), & k=20161, \\ \theta_1\theta_{10080}\Theta_1\Theta_{10080}, & k=20162, \end{cases} \quad (6)$$

Subset MA. The option of the subset MA operator $q = (1, 10080, 10081)$ and $Q = (1, 10080, 10081)$ in as follow:

$$\begin{aligned}\nabla^d \nabla^{D_1}_{10080} \nabla^{D_2}_{10081} \ln(\hat{\mathscr{L}}_t - \mu) = \left(1-\theta_1\mathscr{L}^1 - \theta_{10080}\mathscr{L}^{10080} - \theta_{10081}\mathscr{L}^{10081}\right) \\ \left(1-\Theta_1\mathscr{L}^1 - \Theta_{10080}\mathscr{L}^{10080} - \Theta_{10081}\mathscr{L}^{10081}\right)\varepsilon_t. \quad (7)\end{aligned}$$

Thus, the final expanded form of the polynomial product is:

$$\begin{aligned}\hat{\mathscr{L}}_t = \varepsilon_t - (\Theta_1+\theta_1)\varepsilon_{t-1} + \theta_1\Theta_1\varepsilon_{t-2} - (\Theta_{10080}+\theta_{10080})\varepsilon_{t-10080} - (\Theta_{10081}+\theta_{10081}) + (\theta_1\Theta_{10080}+\theta_{10080}\Theta_1)\varepsilon_{t-10081} \\ + (\theta_1\Theta_{10081}+\theta_{10081}\Theta_1)\varepsilon_{t-10082} + \theta_{10080}\Theta_{10080}\varepsilon_{t-20160} \\ + (\theta_{10080}\Theta_{10081}+\theta_{10081}\Theta_{10080})\varepsilon_{t-20161} + \theta_{10081}\Theta_{10081}\varepsilon_{t-20162}\end{aligned}$$

Next, we collect like terms (i.e., terms with the same power of $\mathscr{L}$):

$$\zeta = \begin{cases} 1, & k=0, \\ -(\Theta_1+\theta_1), & k=1, \\ \theta_1\Theta_1, & k=2, \\ -(\Theta_{10080}+\theta_{10080}), & k=10080, \\ -(\Theta_{10081}+\theta_{10081}) + (\theta_1\Theta_{10080}+\theta_{10080}\Theta_1), & k=10081, \\ (\theta_1\Theta_{10081}+\theta_{10081}\Theta_1), & k=10082, \\ \theta_{10080}\Theta_{10080}, & k=20160, \\ (\theta_{10080}\Theta_{10081}+\theta_{10081}\Theta_{10080}), & k=20161, \\ \theta_{10081}\Theta_{10081}, & k=20162, \end{cases} \quad (8)$$

which are polynomial functions of orders P_1, P_2, and P_3 respectively. These delay polynomials allow AR to model daily, weekly, and annual cycles. These are polynomial functions of orders P_1, P_2, and P_3 respectively. These delay polynomials allow the modeling of daily, weekly, and annual cycles, as well as the d-th and D-th non-seasonal periods, and the seasonal difference $\nabla^d = (1-\mathscr{L})^d$, $\nabla^{D_1}_{S_1} = (1-\mathscr{L}^{1440})^{D_1}$, $\nabla^{D_2}_{S_2} = (1-\mathscr{L}^{10080})^{D_2}$.

3.3. Estimation

SAS is utilized in this step, and the following code provides an example of the program used to estimate subset, multiplicative, and additive MA models. The ESTIMATE statement is used to define the MA model to be applied to the variable identified in the previous IDENTIFY statement and to estimate the model's parameters. Additionally, the ESTIMATE statement generates diagnostic statistics to assist in evaluating the model suitability.

```
• proc arima data=vstl;
  identify var=Load(1,1440)
  nlag=1440;
  run;
• /*** for subset AR ***/
  estimate p=(1,1440,1441)
  noconstant
  method=uls;
  run;
• /*** for multiplicative AR ***/
  estimate p=(1)(1440) noconstant
  method=uls;
  run;
• /*** for additive AR ***/
  estimate p=(1,1440) noconstant
  method=uls;
  run;
```

```
• proc arima data=vstl;
  identify
  var=Load(1,10080,10081)
  nlag=10080 noprint;
  run;
• /*** Subset AR ***/
  estimate
  p=(1)
  P=(1,10080,10081)
  noconstant method=uls;
  run;
• /*** Multiplicative AR ***/
  estimate
  p=(1)
  P=(1)(10080)(10081)
  noconstant method=uls;
  run;
• /*** Additive AR ***/
  estimate
  p=(1)
  P=(1,10080)
  noconstant method=uls;
  run;
```

```
• proc arima data=vstl;
  identify var=Load(1,1440)
  nlag=1440;
  run;
• /*** Subset MA ***/
  estimate q=(1,1440,1441)
  noconstant
  method=uls;
  run;
• /*** Multiplicative MA ***/
  estimate q=(1)(1440) noconstant
  method=uls;
  run;
• /*** Additive MA ***/
  estimate q=(1,1440) noconstant
  method=uls;
  run;
```

```
• proc arima data=vstl;
  identify
  var=Load(1,10080,10081)
  nlag=10080 noprint;
  run;
• /*** Subset MA ***/
  estimate
  q=(1)
  Q=(1,10080,10081)
  noconstant method=uls;
  run;
• /*** Multiplicative MA ***/
  estimate
  q=(1)
  Q=(1)(10080)(10081)
  noconstant method=uls;
  run;
• /*** Additive MA ***/
  estimate
  q=(1)
  Q=(1,10080)
  noconstant method=uls;
  run;
```

3.4. Model Checking

Summary output from Table 3 estimation for subset, Table 4 estimation for multiplicative, Table 5 estimation for additive. Using the Ljung-Box test, it is confirmed that all models meet the assumption that the residuals are white noise.

TABLE 3. Unconditional Least Squares Estimation for Subset

Unconditional Least Squares Estimation					
Parameter	Estimate	St Error	t Value	Pr > |t|	Lag
MA1,1	-0.048	0.00416	-11.69	$< 1{,}00 \times 10^{-4}$	1
MA2,1	-0.432	0.00318	-135.92	$< 1{,}00 \times 10^{-4}$	1
MA2,2	**0.787**	0.0030	259.67	$< 1{,}00 \times 10^{-4}$	**10080**
MA2,3	0.645	0.00323	199.72	$< 1{,}00 \times 10^{-4}$	10081

TABLE 4. Unconditional Least Squares Estimation for Multiplicative

Unconditional Least Squares Estimation					
Parameter	Estimate	St Error	t Value	Pr > |t|	Lag
MA1,1	-0.174	0.0686	-2.55	0.0109	1
MA2,1	-0.207	0.0680	-3.04	0.0023	1
MA2,2	**0.999**	0.0046	213.10	<.0001	**10080**
Correlations of Parameter Estimates					
	Parameter		MA1,1	MA2,1	MA2,2
	MA1,1		1.000	-0.999	0.226
	MA2,1		-0.999	1.000	-0.227
	MA2,2		0.226	-0.227	1.000

TABLE 5. Unconditional Least Squares Estimation for Additive

Unconditional Least Squares Estimation					
Parameter	Estimate	St Error	t Value	Pr > |t|	Lag
MA1,1	-0.352	0.00246	-143.11	<.0001	1
MA2,1	-0.1003	0.00254	-39.46	<.0001	1
MA2,2	**0.899**	0.00309	290.23	<.0001	**10080**
Correlations of Parameter Estimates					
	Parameter		MA1,1	MA2,1	MA2,2
	MA1,1		1.000	-0.486	-0.168
	MA2,1		-0.486	1.000	-0.157
	MA2,2		-0.168	-0.157	1.000

Then, the process continues to calculate the forecasting values based on the subset with eq. (9), multiplicative with eq. (10), and additive with eq. (11) models for performance comparison and evaluation.

$$\hat{\mathscr{L}}_t^* = (1+0.04863B)(1+0.43289B-0.7876B^{10080}-0.64521B^{10081})\varepsilon_t \tag{9}$$

where $d_1 = D_1 = D_2 = 1$ and $\hat{\mathscr{L}}_t^* = \nabla^d \nabla_{1440}^{D_1} \nabla_{10080}^{D_2} \ln(\mathscr{L}_t - \mu)$ and $\mathscr{L}_t$ is the actual data, i.e. the very short term load.

$$\hat{\mathscr{L}}_t^* = (1+0.17476B)(1+0.20723B)(1-0.999B^{10080})\varepsilon_t \tag{10}$$

where $d_1 = D_1 = D_2 = 1$ and $\hat{\mathscr{L}}_t^* = \nabla^d \nabla_{1440}^{D_1} \nabla_{10080}^{D_2} \ln(\hat{\mathscr{L}}_t - \mu)$ and $\mathscr{L}_t$ is the actual data, i.e. the very short term load.

$$\hat{\mathscr{L}}_t^* = (1-0.35269B)(1+0.10035B-0.89965B^{10080})\varepsilon_t \tag{11}$$

where $d_1 = D_1 = D_2 = 1$ and $\hat{\mathscr{L}}_t^* = \nabla^d \nabla_{1440}^{D_1} \nabla_{10080}^{D_2} \ln(\hat{\mathscr{L}}_t - \mu)$ and $\mathscr{L}_t$ is the actual data, i.e. the very short term load.

Additive Model. In this model, all parameters are significant at the 0.1 alpha significance level, with white noise residuals as confirmed by the Ljung-Box test up to lag 10080. Table 6 presents 10 extreme residual values. However, the residuals of the model in Table 7 do not follow a normal distribution due to the presence of outliers in the data.

TABLE 6. Extreme Observations for Additive Model

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-1200.480	176864	1162.90	160053
-1147250	175620	1205.25	169108
-1062.530	173406	1233.03	176586
-1000.685	169107	1163.91	173407
-824.213	174488	1581.99	176865

TABLE 7. Test for Normality for Additive Model

Test for Normality				
Test	Statistics		P Value	
Kolmogorov-Smirnov	D	0.045354	Pr >D	<0.0100
Cramer-von Mises	W-Sq	165.4151	Pr >W-Sq	<0.0050
Anderson-Darling	A-Sq	1094.89	Pr >A-Sq	<0.0050

Subset Model. In this model, all the parameters are significant at alpha 0.1 significance level with white noise residuals based on the Ljung-Box test until lag 10080. Table 8 from this model give also 10 extreme residual values. Table 9 with the model's residuals however do not satisfy the normal distribution because of the presence of outliers in the data.

TABLE 8. Extreme Observations for Subset Model

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-1173.776	176864	1177.49	160053
-1154.947	175620	1261.15	169108
-1041.418	173406	1278.06	173407
-985.862	169107	1452.62	176586
-908.523	176585	1662.62	176865

TABLE 9. Test for Normality for Subset Model

Test for Normality				
Test	Statistics		P Value	
Kolmogorov-Smirnov	D	0.034702	Pr >D	<0.0100
Cramer-von Mises	W-Sq	98.70502	Pr >W-Sq	<0.0050
Anderson-Darling	A-Sq	657.2371	Pr >A-Sq	<0.0050

Multiplicative Model. In this model, all the parameters are significant at alpha 0.1 significance level with white noise residuals based on Ljung-Box test until lag 10080. Table 10 give 10 extreme residual values. Table 11 model's residuals how ever do not satisfy the normal distribution because of the presence of outliers in the data.

TABLE 10. Extreme Observations for Multiplicative Model

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-1199.605	176864	1151.45	160053
-1152.882	175620	1153.26	169108
-1064.021	173406	1161.01	176586
-1002.799	169107	1196.74	173407
-822.757	174488	1509.94	176865

TABLE 11. Test for Normality for Multiplicative Model

Test for Normality				
Test	Statistics		P Value	
Kolmogorov-Smirnov	D	0.047206	Pr >D	<0.0100
Cramer-von Mises	W-Sq	176.3649	Pr >W-Sq	<0.0050
Anderson-Darling	A-Sq	1162.973	Pr >A-Sq	<0.0050

3.5. Performance Comparison

The model identification step for the MA models showed that there were different ACF and PACF between the subset, multiplicative and additive models, particularly in the lag order, as the results of the performance evaluation show that the multiplicative model yields a better forecast in the sample data set, or less SBC and AIC, than the additive and subset MA model, also the estimated parameter of multiplicative effect was significant. The comparison results between the additive, subset and multiplicative models are summarized in Table 12.

TABLE 12. AIC, SBC for MA Model

	Subset	Additive	Multiplicative
Variance Est.	1589.84	1839.802	1178.873
Std Error Est.	39.8728	42.89291	34.33472
AIC	2030623	2059872	1977496
SBC	2030664	2059902	1977537
Number of Res.	198716	198716	198716

4. CONCLUSIONS

This paper discusses three types of seasonal MA models, namely the subset, multiplicative, and additive models, including the theoretical aspects of ACF and PACF, how to model these values using the R program, and how to test them using the SAS program when estimating the model using the Box-Jenkins method. Most previous studies simply used the multiplicative MA model directly, without extensively defining ACF or PACF with lags as the order of multiplication and without testing whether the multiplicative parameter is significant. Overall, these empirical results show that when determining the orders in the subset, multiplicative, or additive models, the MA model should account for the subset, multiplicative, or additive order.

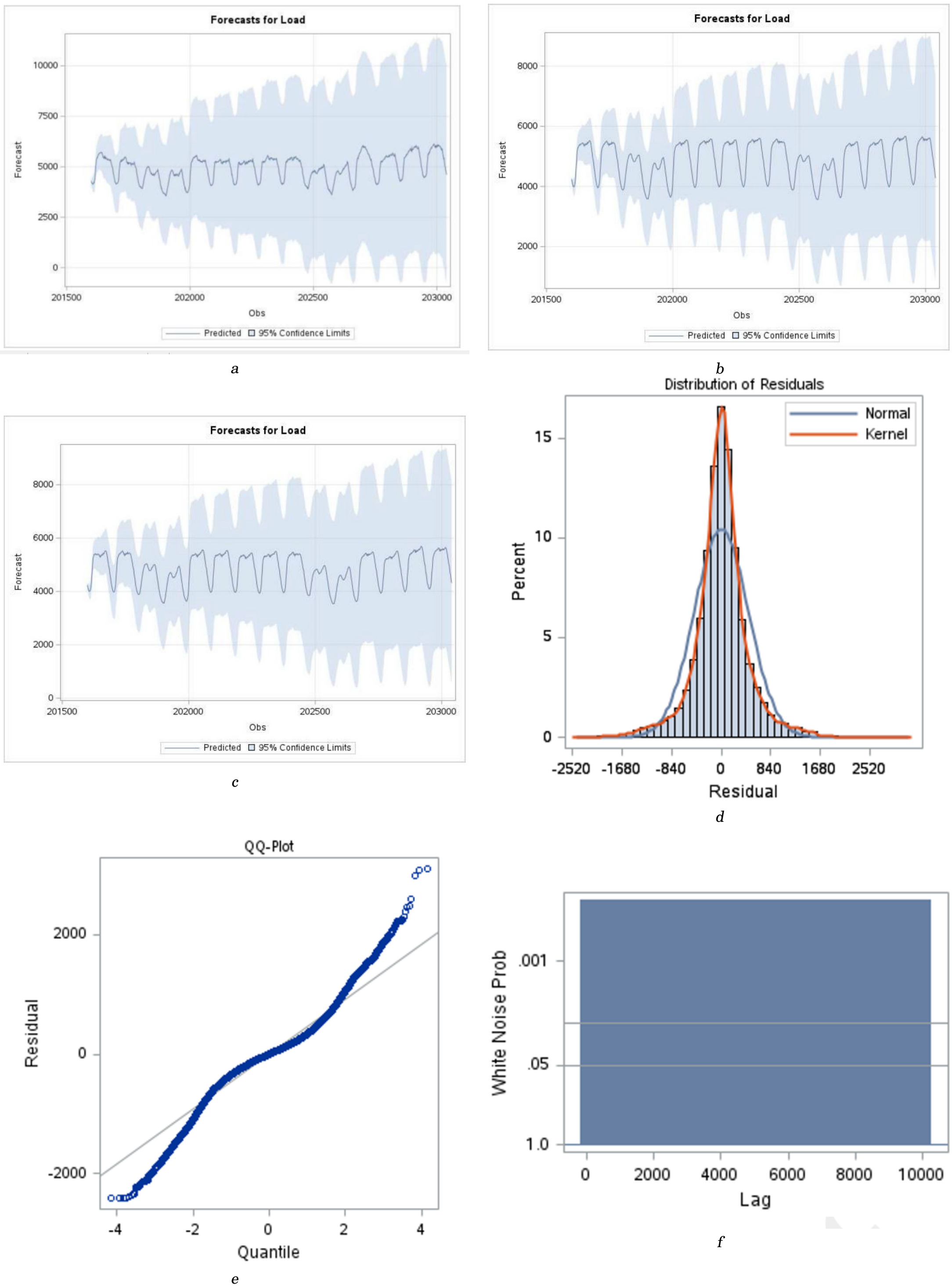


FIGURE 5. Model subset (*a*), Model multiplicative (*b*), Model additive (*c*), Distribution of residual normality diagnostics for load (*d*), and QQ-plot residual normality diagnostics for load (*e*) White noise prob residual correlation diagnostics for load.

Acknowledgment

We are indebted to Dr. Suhartono (Institut Teknologi Sepuluh Nopember), who generously provided the R codes he used with his students when teaching through online Zoom sessions.

REFERENCE

[1] Lee, Muhammad Hisyam, and others, 2011, Forecasting of tourist arrivals using subset, multiplicative or additive seasonal ARIMA model, *Matematika*, **27**(2), pp. 169–182.

[2] Rosadi, Dedi, and others, 2012, New procedure for determining order of subset autoregressive integrated moving average (ARIMA) based on over-fitting concept, in *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, pp. 1–5, IEEE.

[3] Dinata, Syalam Ali Wira, Azka, Muhammad, Faisal, Muliady, Suhartono, Suhartono, Yendra, Rado, Gamal, Moh Danil Hendry, 2020, Short-Term load forecasting double seasonal ARIMA methods: An evaluation based on Mahakam-East Kalimantan data, in *AIP Conference Proceedings*, **2268**(1), pp. 1–5, AIP Publishing.

[4] Giovani, Muhammad, Anggriani, Indira, Simatupang, Syalam Ali Wira Dinata, 2022, Comparison in predicting the short-term using the SARIMA, DSARIMA, and TSARIMA methods, *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, **16**(4), pp. 1487–1496.

[5] Azizah, Nur, Nugraheni, Kartika, Dinata, Syalam Ali Wira, Triple-Seasonal ARIMA untuk peramalan data konsumsi beban listrik.

[6] Suhartono, 2011, Time series forecasting by using seasonal autoregressive integrated moving average: Subset, multiplicative or additive model, *J. Math. Stat.*, **7**(1), pp. 20–27.

[7] Voronin, Vyacheslav A., and Nepsha, Fedor S., 2022, Electrical Load and Solar Generation Forecasting in Hybrid Energy Systems, *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, pp. 2040–2044. IEEE,

[8] Bala, Dahiru A., and Shuaibu, Mohammed, 2022, Forecasting United Kingdom's energy consumption using machine learning and hybrid approaches, *Energy & Environment*, SAGE Publications Sage UK: London, England,

[9] Wang, Diankai, Gryshova, Inna, Kyzym, Mykola, Salashenko, Tetiana, Khaustova, Viktoriia, and Shcherbata, Maryna, 2022, Electricity price instability over time: Time series analysis and forecasting, *Sustainability*, **14**(15), pp. 9081. MDPI.

[10] Debnath, Kumar Biswajit, and Mourshed, Monjur, 2018, Forecasting methods in energy planning models, *Renewable and Sustainable Energy Reviews*, **88**, pp. 297–325. Elsevier.

[11] Kramar, Vadim, and Alchakov, Vasiliy, 2023, Evaluation of the Machine Learning Techniques for Forecasting the Seasonal Time Series, *2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, pp. 262–265. IEEE.

[12] Wu, Don Chi Wai, Ji, Lei, He, Kaijian, and Tso, Kwok Fai Geoffrey, 2021, Forecasting tourist daily arrivals with a hybrid Sarima–Lstm approach, *Journal of Hospitality & Tourism Research*, **45**(1), pp. 52–67. SAGE Publications Sage CA: Los Angeles, CA.

[13] Dinata, Syalam Ali Wira, Azka, Muhammad, Hasanah, Primadina, Suhartono, Suhartono, and Gamal, Moh Danil Hendry, 2021, Comparison of Short-Term Load Forecasting Based on Kalimantan Data, *Indonesian Journal of Statistics and Its Applications*, **5**(2), pp. 243–259.

[14] Peshkov, AV, and Alsova, OK, 2020, Autoregressive Models for Solving the Problem of Forecasting Active Energy Complexes, *Journal of Physics: Conference Series*, **1661**(2), pp. 012070,

[15] Durova, M. A., Zein, A. N., Borisova, S. V., Mishin, A. A., Kan, D., and Osipov, S. K., 2023, Short-term forecasting of the time series of electricity prices with ensemble algorithms, *2023 5th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE)*, **1661**(2), pp. 1–6.

[16] Durova, M. A., Zein, A. N., Borisova, S. V., Mishin, A. A., Kan, D., and Osipov, S. K., 2023, Autoregressive Models for Solving the Problem of Forecasting Active Energy Complexes, *2023 5th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE)*, **5**, pp. 1–6. IEEE,

[17] Petrusevich, Denis, 2019, Time series forecasting using high order ARIMA functions, *International Multidisciplinary Scientific GeoConference: SGEM*, **19**(2.1), pp. 673–679. Surveying Geology & Mining Ecology Management (SGEM).

[18] Sadaei, Hossein Javedani, Guimaraes, Frederico Gadelha, da Silva, Cidiney Jose, Lee, Muhammad Hisyam, and Eslami, Tayyebeh, 2017, Short-term load forecasting method based on fuzzy time series, seasonality and long memory process, *International Journal of Approximate Reasoning*, **83**, pp. 196–217. Elsevier,

[19] Hachim, Mohammed Qandeel Hachim, 2021, Prediction of electricity production and demand in Russia by time series approaches, *International Journal of Approximate Reasoning*, **83**, pp. 196–217. Elsevier,

[20] Nigam, R., Nigam, SMSK, and Mittal, SK, 2014, The river runoff forecast based on the modeling of time series, *Russian Meteorology and Hydrology*, **39**, pp. 750–761. Springer,