

PERFORMANCE COMPARISON OF SARIMA INTERVENTION AND PROPHET MODELS FOR FORECASTING THE NUMBER OF AIRLINE PASSENGER AT SOEKARNO-HATTA INTERNATIONAL AIRPORT

Vivin Nur Aziza^{1*}, Fatma Hilali Moh'd², Firda Aulia Maghfiroh³,
Khairil Anwar Notodiputro⁴, Yenni Angraini⁵

^{1,2,3,4,5}Departement of Statistics, Faculty of Mathematics and Natural Sciences, IPB University
Meranti St at Dramaga, Bogor, 16680, Indoensia

Corresponding author's e-mail: * vivin.nuraziza@apps.ipb.ac.id

ABSTRACT

Article History:

Received: 26th June 2023

Revised: 28th September 2023

Accepted: 29th October 2023

Keywords:

Covid-19 pandemic;
intervention model; FB
Prophet; airline passenger;
time series data.

The impact of the COVID-19 pandemic on the air transportation sector, particularly Soekarno-Hatta (Soetta) International Airport, has been quite significant. The number of passengers at Soetta Airport has decreased due to the COVID-19 pandemic, but flight activities are still ongoing to this day. An accurate forecasting model is needed to predict the number of airline passengers at Soetta Airport with the presence of the COVID-19 pandemic as an intervention. In this study we discuss performance comparison of two models namely SARIMA intervention and Prophet in forecasting the number of domestic passengers at Soetta Airport. The research results showed that the best SARIMA intervention model was SARIMA $(0,1,1)(1,0,0)^{12}$ $b = 0$, $s = 20$, $r = 0$, with a Mean Absolute Percentage Error (MAPE) of 28% and Root Mean Square Error (RMSE) of 433473. On the other hand, the Prophet model yielded a MAPE of 37% and an RMSE of 497154. In terms of MAPE and RMSE, the SARIMA intervention method provides better results than the Prophet model in forecasting the number of domestic passengers at Soetta Airport.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite this article:

V. N. Aziza, F. H. Moh'd, F. A. Maghfiroh, K. A. Notodiputro and Y. Angraini., "PERFORMANCE COMPARISON OF SARIMA INTERVENTION AND PROPHET MODELS FOR FORECASTING THE NUMBER OF AIRLINE PASSENGER AT SOEKARNO-HATTA INTERNATIONAL AIRPORT," *BAREKENG: J. Math. & App.*, vol. 17, iss. 4, pp. 2107-2120, December, 2023.

Copyright © 2023 Author(s)

Journal homepage: <https://ojs3.unpatti.ac.id/index.php/barekeng/>

Journal e-mail: barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id

Research Article • Open Access

1. INTRODUCTION

Time series data can be affected by external events, such as natural disasters, government policies, national holidays, financial crises, and regime changes. These events are referred to as interventions in time series data analysis. If time series data analysis is conducted without considering the impact of interventions, it will result in significant model errors. The larger the error value, the less accurate the generated model will be, and the less capable the model will be of depicting the observed data. Therefore, it is important to consider interventions in time series data modeling to ensure more accurate and relevant analysis results.

Interventions in time series data, as mentioned, can have a profound impact on the accuracy of the models employed. In particular, the COVID-19 pandemic serves as a significant example of such an intervention, disrupting various aspects of daily life, including the aviation industry. This context sets the stage for understanding how interventions can lead to substantial changes in time series data, as we will explore in the case of airline passenger at Soekarno-Hatta (Soetta) International Airport.

Soekarno-Hatta (Soetta) International Airport is Indonesia's largest and main airport, serving a busy traffic of domestic and international flights to various destinations. In 2018, it ranked 18th in the world and 1st in Southeast Asia as the busiest airport, serving thousands of passengers daily [1]. However, the COVID-19 pandemic has significantly impacted the air transportation sector, including Soetta Airport. The COVID-19 pandemic began in Indonesia in March 2020. The government implemented a series of measures to contain the spread of the virus, including Large-Scale Social Restrictions (PSBB) and Community Activity Restrictions (PPKM). These measures had a significant impact on the number of airline passengers, which decreased significantly during several periods. Additionally, the requirement to provide expensive rapid test results and the fear of contracting the virus further discouraged people from flying. A study found that at least 34 of Indonesia's 187 airports suspended passenger flights during a ban on homecoming travel from April 24 to June 1, 2020. This was also reflected in the revenue decline of PT Garuda Indonesia, one of Indonesia's major airlines, in the first quarter of 2020 compared to the previous year. Additionally, there was no increase in passenger numbers during high-traffic periods such as Eid homecoming, school holidays, and the Hajj and Umrah pilgrimages in 2020, resulting in a significant decline in revenue [2]. The COVID-19 pandemic is considered an intervention that disrupted the time series data of passenger numbers at Soetta Airport.

Time series forecasting is the process of predicting future values in a time series based on existing patterns and trends in the data. The goal is to provide accurate and useful estimates of how the time series will develop in the future. Forecasting often combines mathematical models with the insights and considerations of a manager [3]. In a previous study, [4] compared forecasting methods between SARIMA and SARIMA Intervention for the number of domestic passengers at Soekarno-Hatta Airport. They concluded that the SARIMA Intervention method was the best. Another study conducted by [5] compared ARIMA, ARIMA with intervention, and Extreme Learning Machine for forecasting the number of passengers at Domine Eduard Osok Sorong Airport and concluded that ARIMA with intervention reduces MAPE by up to 45%. [6] Conducted a comparison between the SARIMA method and the step function intervention model to forecast the number of visitors to the Londa tourist attraction, with the conclusion that the best model is SARIMA with intervention. Expanding on the previous research, this study will compare the SARIMA intervention model with a relatively new machine learning model called Prophet, which is claimed to be a powerful forecasting tool. The goal is to predict the number of passengers at Soekarno-Hatta Airport while considering the impact of Covid-19 as an intervention.

SARIMA, or Seasonal Autoregressive Integrated Moving Average, is a statistical model used to analyze and predict data with seasonal patterns and trends. The SARIMA model is an extension of the Autoregressive Integrated Moving Average (ARIMA) model, with the addition of seasonal components [7]. Seasonal models have observation series that depend on previous and repetitive observations, forming cyclic patterns. The general notation for SARIMA is $ARIMA(p,d,q)(P,D,Q)^s$, where (p,d,q) represents the non-seasonal order of the model, (P,D,Q) represents the seasonal order, and s represents the number of periods per season [8].

Interventions are assumed to occur at a known point in time and cause a change in the pattern of time series data. Intervention analysis is conducted to determine the extent to which the intervention effect influences the pattern of the time series data. An ARIMA time series that exhibits seasonal patterns and includes interventions is referred to as a SARIMA Intervention model [6]. Intervention events can result in two different types of impacts: temporary intervention events that occur within a specific period (pulse function) and intervention events that have long-term impacts (step function) [5].

The Prophet model, also known as FB Prophet, is a forecasting tool developed by Facebook's data science team. It is designed for forecasting time series data with a simple and user-friendly approach. The Prophet model consists of three main components: seasonality, trend, and holidays. Prophet incorporates time series interventions within the holiday component, which are referred to as special events [9]. Prophet has several advantages over classical time series models like ARIMA. It can handle missing data, work with various seasonal patterns, and easily incorporate new components such as special events. Prophet is also easy to use, even for analysts who may not have an in-depth understanding of time series forecasting techniques [10].

This research aims to find the best model between SARIMA intervention and Prophet in forecasting seasonal time series data, such as the number of airline passengers at Soekarno-Hatta International Airport, considering the intervention of the Covid-19 pandemic. It is expected that the results of this research can be used to forecast seasonal time series data with interventions more accurately.

2. RESEARCH METHODS

2.1 Data

The data used is the number of domestic airline passengers at Soetta International Airport obtained from the website of the Biro Pusat Statistik (BPS) [11]. The data covers a monthly period starting from January 2010 to March 2023. There is an intervention period or special event due to the COVID-19 pandemic that affects the time series data, which started in March 2020 and lasted until December 2021.

2.2 Stages of Data Analysis

The stages of data analysis carried out in this study are as follows:

1. Exploring data through time series plots helps to understand the characteristics or patterns of the data.
2. Splitting the data into two sets: the training data, which spans from January 2010 to December 2021, and the testing data, which spans from January 2022 to March 2023. Next, the training data (Z_t) is further divided into two subsets: the pre-intervention training data (N_t) and the intervention training data (I_t). The intervention period begins in March 2020, so the intervention training data is defined as the period from March 2020 to December 2021.
3. Develop an intervention SARIMA model, with the following steps:
 - a. develop a SARIMA model for the pre-intervention training data (N_t)
 - i. Checking for stationarity in mean and variance. If the data is not stationary in the mean, differencing is performed. The examination of stationarity in variance is done using the Box-Cox plot. The transformation criterion is based on the λ value obtained from the plot. If $\lambda=1$, no Box-Cox transformation is necessary [12].
 - ii. After making the time series stationary, we can use the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots to identify the appropriate SARIMA model [13].

Estimates the parameters of the model using the Maximum Likelihood Estimation (MLE) method to estimate the parameters of the model. The hypothesis used in the MLE method is H_0 : The parameter estimate is equal to 0 (non-significant parameter), and H_1 : The parameter estimate is not equal to 0 (significant parameter) at a significance level of 5%. A good model is indicated when all parameter estimates are significant.

Perform a diagnostic check of the model to assess whether the residuals of the formed model exhibit the properties of white noise or not. Diagnostic tests on the model include tests for independence and normality of residuals. The independence of residuals is tested using the Ljung-Box test statistic with hypothesis H_0 : there is no serial autocorrelation in time series data, and H_1 : there is serial autocorrelation in time series data. While the normality is tested using the Kolmogorov-Smirnov test with hypothesis H_0 : the data follows a normal distribution, and H_1 : the data does not follow a normal distribution. The model meets the assumptions if the p-value $> \alpha$ (5%) [14].

- iii. To obtain alternative models, perform model overfitting by gradually increasing the orders of p , q , P , and Q . Overfitting involves progressively increasing the orders of the model parameters. The selected overfitting model is the one in which all coefficient estimates are statistically significant.
- iv. Evaluating the model from the value of the Akaike Information Criterion (AIC) with the formula [15]:

$$AIC = -2\ell + 2k \quad (1)$$

where ℓ is the value of the log-likelihood model, and k is the number of parameters. The best SARIMA model is the model with the smallest AIC value. This model will have a good fit to the data and will be less likely to overfit the data.

- b. Evaluate the model with training data during the intervention period obtain from step 3a(vi).
 - c. Make a graph of the residual SARIMA data model pre-intervention to identify the intervention order that exceeds the $\pm 3\hat{\sigma}$ limit.
 - d. Identify the intervention response that is the order b , s , and r based on the residual graph. The parameters b , s , and r represent the start time of the intervention, the duration of the intervention until stability is restored, and the pattern of the intervention effect, respectively.
 - e. Estimate the parameters of the intervention model and check if they are statistically significant. If they are not, return to step 3d.
 - f. Perform a diagnostic check of the model, as described in step 3a(iv).
 - g. Do overfitting by trying several different values for b , s , and r , then select the model order that produces the best fit for the data.
 - h. Select the best intervention model from a set of candidate models by comparing their AIC values. The model with the smallest AIC values is the best intervention model.
4. Create the Prophet model
- a. Determine the training data during the intervention period as mentioned in step 2 as the Special Event.
 - b. Create a Prophet model using the training data by adding the Special Event component.
 - c. Perform diagnostic checks on the model, including tests for independence and normality of residual.
5. Forecasting as much test data using the best SARIMA intervention and Prophet model to calculate the MAPE and RMSE values in each model. The equations for MAPE and RMSE are as follows [16]:

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{(A_t - F_t)}{A_t} \right|}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2} \quad (3)$$

where A is actual value, F is the forecast value, and n is the number of data points.

6. The model with the smallest RMSE and MSE value is selected as the best forecast model [17].

The flowchart of the research method is illustrated in **Figure 1** below.

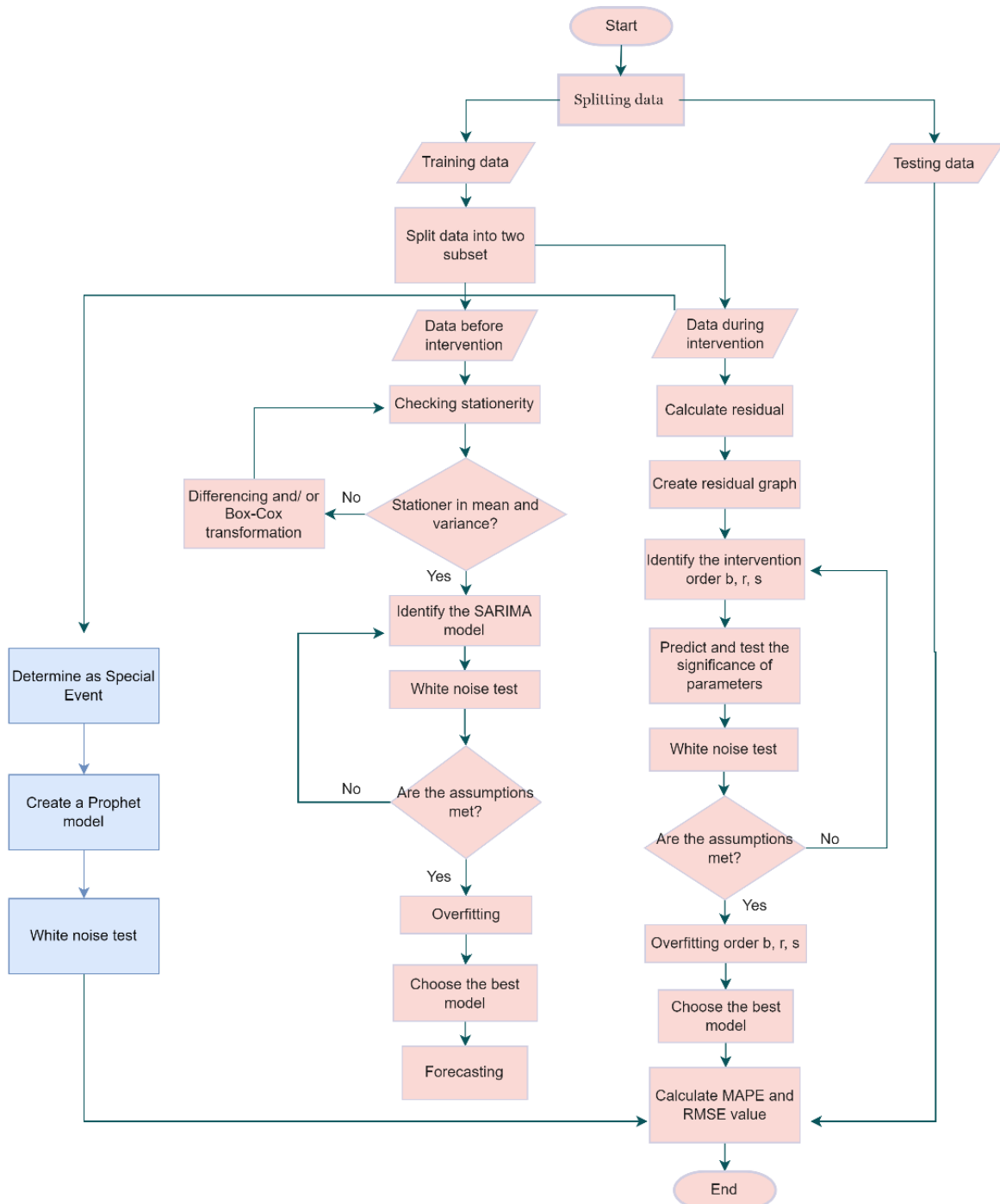


Figure 1. Flowchart of SARIMA Intervention and Prophet method

3. RESULTS AND DISCUSSION

3.1 Data Exploration

In **Figure 2**, the plot of the domestic passenger count time series data at Soekarno-Hatta Airport is presented. It can be observed that the data exhibit seasonal patterns and an increasing trend over time.

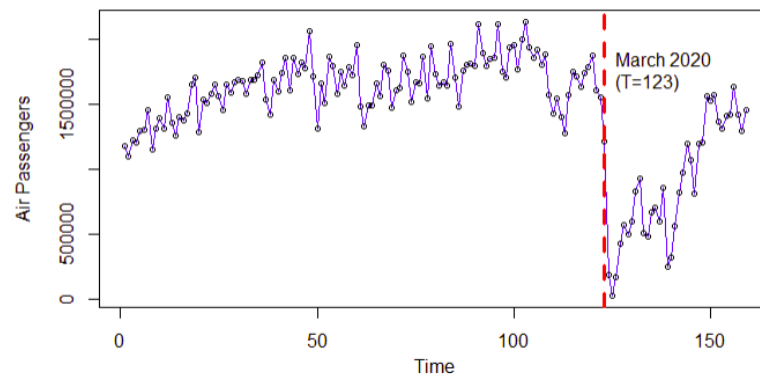
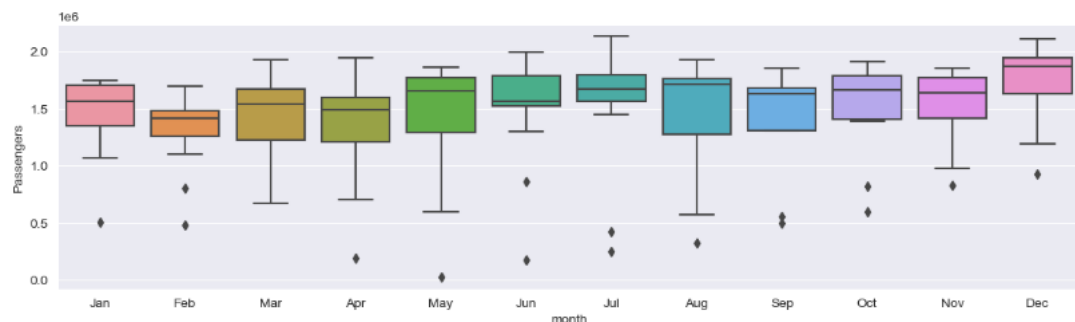


Figure 2. A plot of Time Series Data For Domestic Passengers at Soetta Airport



(a)



(b)

Figure 3. Box plot of domestic Passenger Count at Soetta Airport
(a) Monthly and (b) Yearly

The Box-Plot in **Figure 3** illustrates that the data reaches peaks in July and December within one seasonal cycle. The highest passenger count occurred in the year 2018. However, when the COVID-19 pandemic hit and the PPKM policy was implemented, a significant decrease in passenger count can be observed starting from March 2020 ($T=123$). This decline continued for some time until the passenger count returned to normal levels before the COVID-19 period. Therefore, the training data is divided into two parts: before intervention (January 2010 - February 2020) and after intervention (March 2020 - December 2021). The data from January 2022 to March 2023 is used as the test data for cross-validation.

3.2 SARIMA Intervention Model

3.2.1 SARIMA Modeling Before Intervention

Checking Stationarity in Mean

The SARIMA model was first tested for stationarity on the training data before the intervention consisting of 122 observations (N_t). The ACF plot (**Figure 4**) showed a slow decrease, indicating that the data was not stationary. This was confirmed by the ADF test, which had a p-value of 0.1987, which is greater than the significance level of 0.05. This means that the null hypothesis cannot be rejected, and the data can be concluded to be non-stationary.

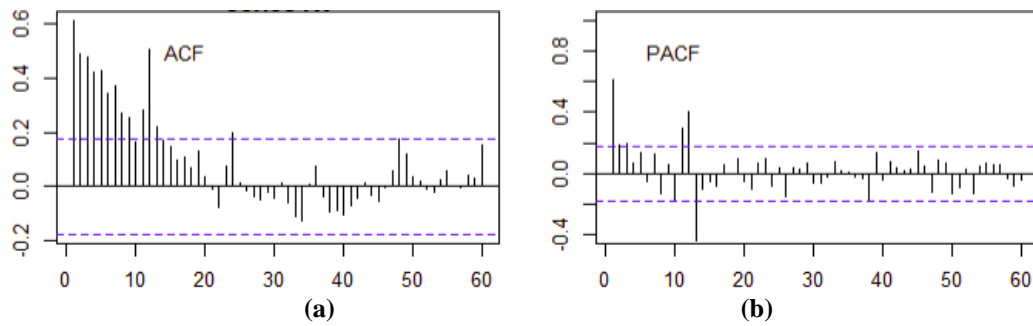


Figure 4. (a) ACF and (b) PACF plots of the N_t time series

After performing first-order differencing ($d=1$) and conducting the ADF test, the obtained p-value ($0.01 < \alpha (0.05)$), indicates that the differenced data is now stationary. The plot of the differenced data is shown in **Figure 5**.

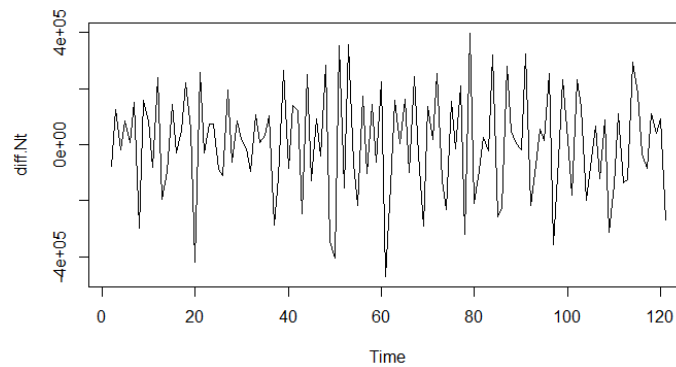


Figure 5. A plot of Differenced Time Series N_t ($d=1$)

Checking Stationerity in Variance

Based on the Box-Cox plot as shown in **Figure 6**, the obtained λ value is close to or exceeds 1, suggesting that the data is already stationary in variance. Therefore, there is no need to perform the Box-Cox transformation.

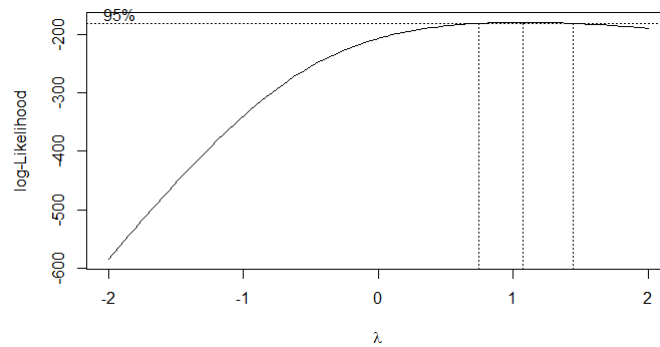


Figure 6. A box-Cox plot of the differenced N_t

Identification of Model

The SARIMA model is suitable for the given data of the number of airline passengers because the data exhibit seasonality. This means that the data has a regular pattern of ups and downs that repeats itself every 12 months. To identify the appropriate SARIMA model, we examine the ACF and PACF plots of the stationary data as shown in **Figure 7**. The ACF plot shows a cut-off at lag 1 and gradually decreases at seasonal lags (12, 24, 36, and so on). This indicates that there is a significant correlation between the data and its lagged values, but this correlation decreases as the lag increases. The PACF plot gradually decreases at non-seasonal lags and cuts off at the first seasonal lag (lag 12). This indicates that there is no significant correlation between the data and its lagged values beyond the seasonal lag. Based on this information, the tentative models obtained are SARIMA(0,1,1)(1,0,0)¹² and SARIMA(2,1,0)(1,0,0)¹².

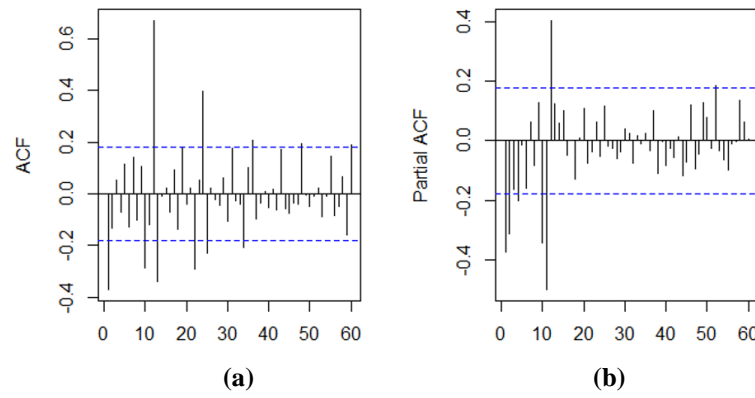


Figure 7. (a) ACF and (b) PACF plots of the differenced Nt

Parameter Estimation

After identifying the two tentative models, the next step is to estimate the parameters of the models. The estimated parameters for both models are shown in Table 1. The p-values for all the parameters in both models are less than α (0,05), which indicates that all the parameter estimates are significant.

Table 1. Estimated Parameter Values for SARIMA Tentative Models

Model	Parameter	Estimate	p-value	Significance	AIC
SARIMA(0,1,1)(1,0,0) ¹²	MA(1)	-0,5146	< 0,05	yes	3174,30
	SAR(1)	0,6980	< 0,05	yes	
SARIMA(2,1,0)(1,0,0) ¹²	AR(1)	-0,5103	< 0,05	yes	3177,45
	AR(2)	-0,1817	< 0,05	yes	
	SAR(1)	0,7109	< 0,05	yes	

Model Diagnostic

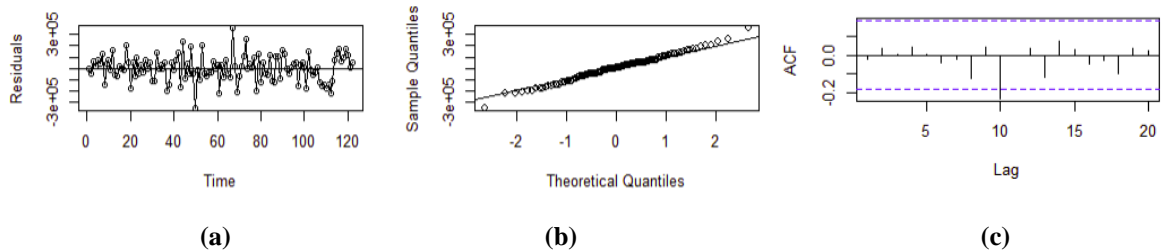


Figure 8. Plot model diagnostic SARIMA(0,1,1)(1,0,0)¹²
 (a) residuals plot, (b) Q-Q plot, and (c) ACF's residuals plot

The diagnostic tests presented in Figures 8 above and Figure 9 below show that both models do not exhibit autocorrelation in their residuals. This is indicated by the fact that the ACF plots of the residuals remain within significant bounds. Additionally, the residual plots and Q-Q plots show that the residuals in both models are centered around the mean value. This is further supported by the Kolmogorov-Smirnov test, which yields a p-value greater than α , indicating that the residuals are normally distributed. Based on these results, it can be inferred that the residuals of both models meet the assumption of white noise. However, based on the criterion of the smallest AIC, the best model for Nt is SARIMA(0,1,1)(1,0,0)¹².

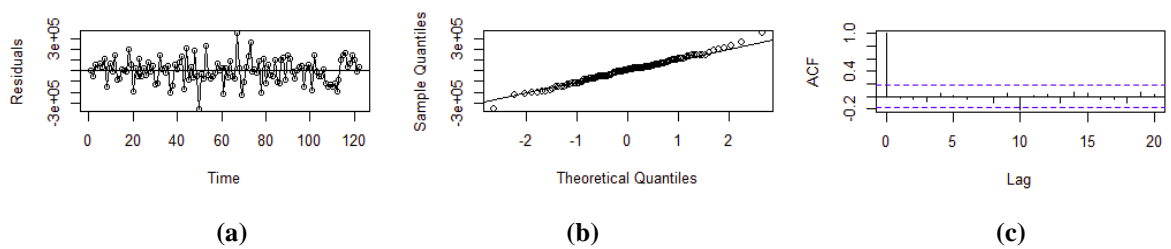


Figure 9. Plot Model Diagnostic SARIMA(2,1,0)(1,0,0)¹²
 (a) residuals plot, (b) Q-Q plot, and (c) ACF's residuals plot

Overfitting

The pre-intervention SARIMA model's overfitting results are presented in the following **Table 2**.

Table 2. Estimating values of the SARIMA model parameters resulting from overfitting

Model	Parameter	Estimate	p-value	Significance	AIC
SARIMA(0,1,2)(1,0,0) ¹²	MA(1)	-0.5329	< 0,05	yes	3176.13
	MA(2)	0.0401	> 0,05	no	
	SAR(1)	0.7033	< 0,05	yes	
SARIMA(1,1,1)(1,0,0) ¹²	AR(1)	-0.0765	> 0,05	no	3176.13
	MA(1)	-0.4562	< 0,05	yes	
	SAR(1)	0.7032	< 0,05	yes	
SARIMA(0,1,1)(2,0,0) ¹²	MA(1)	-0.5146	< 0,05	yes	3176.30
	SAR(1)	0.6979	< 0,05	yes	
	SAR(2)	-0.0002	> 0,05	no	

From the overfitting results in **Table 2**, it can be observed that none of the models have all their parameter estimates significantly at the 5% significance level. Therefore, the best model for Nt is SARIMA(0,1,1)(1,0,0)¹².

3.2.2 SARIMA Modeling During Intervention Period

Identification of Intervention Response

The initial step to identify the intervention order is by observing the difference between the forecasted values of Nt for $t \geq T$ and the observed values Zt. The forecasting is performed for the duration of the intervention, which is from March 2020 to December 2021 (22 observations). As illustrated in **Figure 10**, the comparison between Zt and the forecasted values of Nt shows that the SARIMA(0,1,1)(1,0,0)¹² model closely matches the actual data. However, the intervention of COVID-19 starting in March 2020 has significantly impacted the comparison between the forecasted values and the actual data.

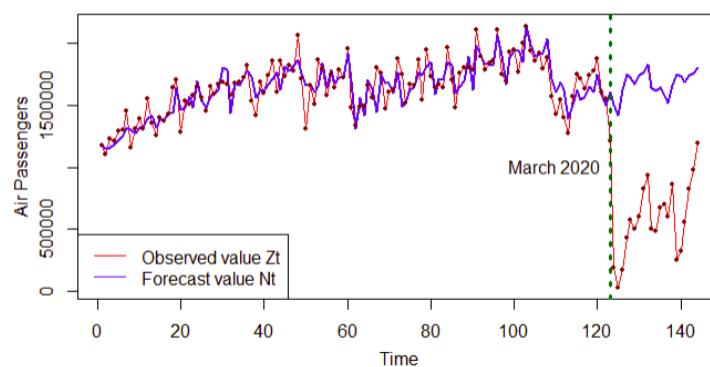


Figure 10. Plot data between forecast value of Nt and observed value of Zt

The outlier effect plot in **Figure 11** shows that the intervention occurs abruptly and has a temporary impact. This indicates that the appropriate intervention pattern is an abrupt temporary intervention.

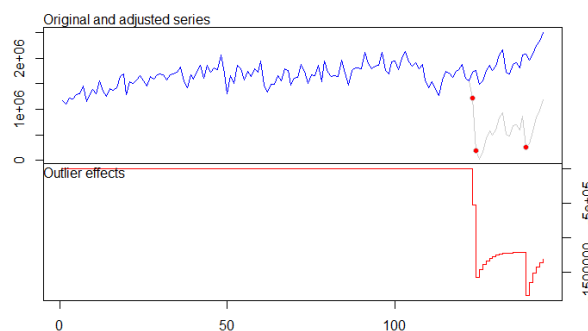


Figure 11. The plot of the intervention pattern

The order of intervention can be identified by observing the residual plot that exceeds the $\pm 3\hat{\sigma}$ limit. The residual plot in **Figure 12** shows that the residual first crosses the $\pm 3\hat{\sigma}$ limit in March 2023. This indicates that the intervention begins at that time, suggesting an order of b of 0. The order of s is determined by how long the intervention lasts until the data returns to normal. The initial assumption for the order of s is

20. The order of r is determined by the presence or absence of a clear pattern in the residual plot. The value of r is 0 if the residual pattern is not clear, and 1 if there is a clear pattern. The tentative model for parameter estimation will have $b = 0$, $s = 20$, and either $r = 0$ or $r = 1$.

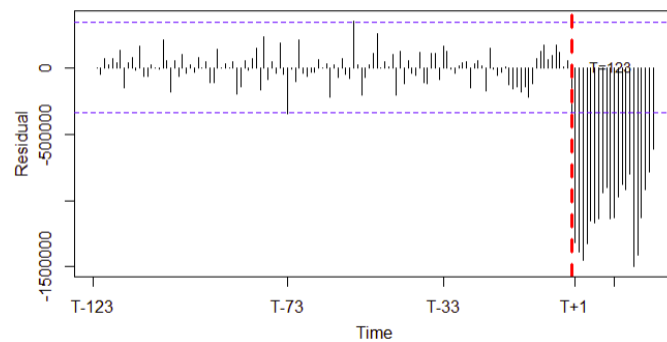


Figure 12. Residual graph of SARIMA(0,1,1)(1,0,0)¹²

The estimated parameter values for the intervention model are presented in the following **Table 3**.

Table 3. Estimating the value of intervention model parameters

Model	Parameter	Estimate	p-value	Significance	AIC
SARIMA(0,1,1)(1,0,0) ¹² $b=0, s=20, r=0$	MA(1)	0.7008	< 0,05	yes	3323.51
	SAR(1)	0.4718	< 0,05	yes	
	ω_0	-1053237.7	< 0,05	yes	
	ω_{20}	-293367.4	< 0,05	yes	
SARIMA(0,1,1)(1,0,0) ¹² $b=0, s=20, r=1$	MA(1)	0.5417	< 0,05	yes	3300.39
	SAR(1)	0.5600	< 0,05	yes	
	ω_0	-810450.1	< 0,05	yes	
	ω_{20}	-201776.2	< 0,05	no	
	δ_0	0.43277	< 0,05	yes	

The intervention model with all its estimated parameters being statistically significant at a significance level of 5% is $b=0$, $s=20$, and $r=0$. This means that the intervention had a significant impact on the data, and that the effect of the intervention was visible for 20 time periods after the intervention. The diagnostic tests of the model, examining the independence and normality of the residuals, are shown in **Figure 13**. The diagnostic tests show that the residuals are independent and normally distributed. This means that the model is a good fit for the data, and that the residuals are not influenced by any systematic patterns.

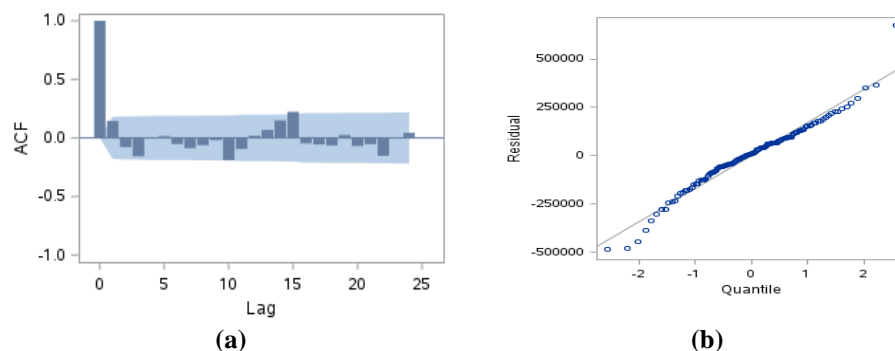


Figure 13. Model diagnostic of intervention model parameters $b=0, s=20, r=0$
(a) ACF residual plot, (b) Q-Q plot of residual

The ACF plot of the residuals shows that no lag exceeds the significance boundary, indicating no significant autocorrelation. The Q-Q plot also reveals that the residuals are closely scattered around the mean value. The Ljung-Box test and Kolmogorov-Smirnov test yield p-values greater than the significance level, indicating that the residuals of the intervention model with $b=0$, $s=20$, and $r=0$ satisfy the assumptions of independence and normality.

Overfitting the Intervention Order

The next step is to try different values for the intervention order to see if any of them improve the model. The results of this are shown in **Table 4**. None of the models with intervention order $b=0, s=21, r=0$

or $b=0, s=21, r=1$ have all significant parameter estimates. This means that these models are not reliable and cannot be used. The best SARIMA intervention model for the number of domestic passengers at Soetta Airport is SARIMA(0,1,1)(1,0,0)¹² with intervention order $b=0, s=20$, and $r=0$.

Table 4. Estimating The Value Of Intervention Model Parameters From Overfitting

Model	Parameter	Estimate	p-value	Significance	AIC
ARIMA(0,1,1)(1,0,0) ¹² $b=0, s=21, r=0$	MA(1)	0.6880	< 0,05	yes	3298.42
	SAR(1)	0.4846	< 0,05	yes	
	ω_0	-1049693.6	< 0,05	yes	
	ω_{21}	-310840.1	> 0,05	no	
ARIMA(0,1,1)(1,0,0) ¹² $b=0, s=21, r=1$	MA(1)	0.5298	< 0,05	yes	3275.48
	SAR(1)	0.5724	< 0,05	yes	
	ω_0	-809925.3	< 0,05	yes	
	ω_{21}	-237936.7	> 0,05	no	
	δ_0	0.4364	< 0,05	yes	

3.3 Prophet Model

The Prophet method is a forecasting method that is designed to be simple and easy to use. It is implemented in the Prophet library in Python software. Prophet does not require differencing or Box-Cox transformation of non-stationary data, unlike the ARIMA method. The first step in using Prophet is to define the COVID-19 intervention period, which is from March 2020 to December 2021, as a special event. This will exclude the intervention period from the time series in the resulting model. This can be seen from the holidays component in Prophet, as shown in **Figure 14**.

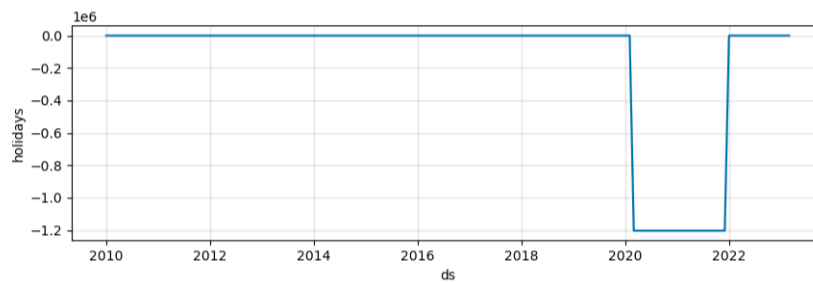


Figure 14. Holidays Component of The Prophet Model

Figure 15 shows the plot of the fitted Prophet model compared to the actual data. It can be observed that the Prophet modeling results are in line with the actual data before the intervention period. During the intervention period, the Prophet model captures a similar pattern to the actual data.

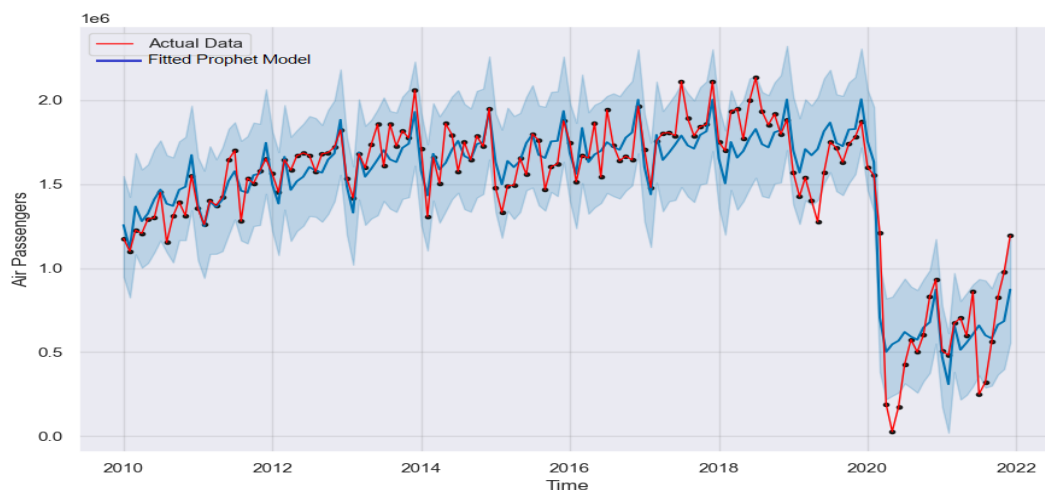


Figure 15. Plot Data Between the Fitted Prophet Model and Actual Data

Next, the model is diagnosed by examining the autocorrelation and distribution of residuals, as shown in **Figure 16** below.

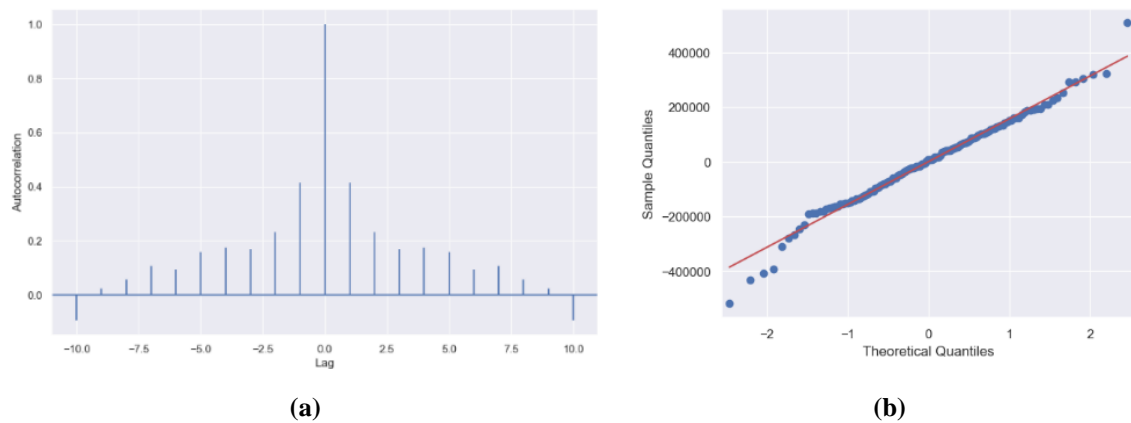


Figure 16. The Diagnostic Results Of The Prophet Model
(a) ACF residual plot, (b) Q-Q plot of residual

The residuals of the Prophet model were analyzed to check if they met the assumption of white noise. The ACF plot showed no significant lags, and the Ljung-Box test resulted in a p-value greater than the significance level. This indicates that the residuals are not autocorrelated. The Q-Q plot showed that the distribution of the residuals is tightly clustered around the middle value of 0. However, the Kolmogorov-Smirnov test resulted in a p-value less than the significance level, indicating that the residuals are not normally distributed.

The Central Limit Theorem states that a distribution can be approximated by a normal distribution when the sample size is equal to or greater than 30 [18]. In this study, a sample size of 144 is used, so it can be said that the assumption of normality is satisfied. Therefore, the residuals of the Prophet model have met the assumption of white noise.

3.4 Determine The Best Model

After building the SARIMA intervention and Prophet models, the next step is to use them to forecast the test data (which has 15 data points). Figure 17 below shows a plot that compares the cross-validated forecasts of both models with the actual test data.

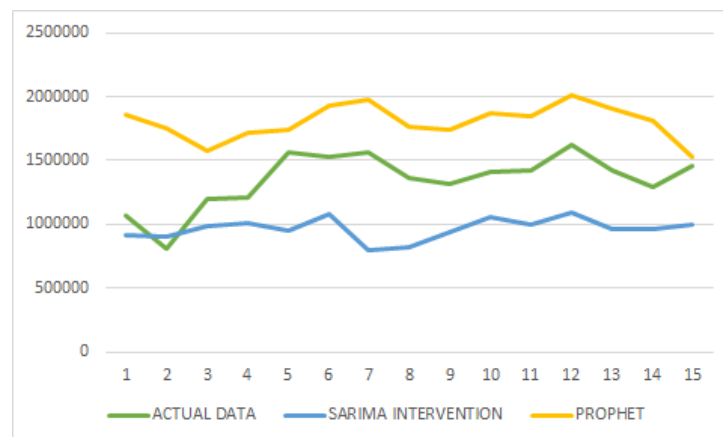


Figure 17. A plot of Test Data and Forecasting Data Using the SARIMA Intervention and Prophet Model

The best model between SARIMA intervention and Prophet is selected by comparing their MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Square Error) scores. MAPE and RMSE are two common metrics used to evaluate the performance of forecasting models. MAPE stands for Mean Absolute Percentage Error, and it is calculated by taking the average of the absolute percentage errors between the predicted and actual values. RMSE stands for Root Mean Square Error, and it is calculated by taking the square root of the average of the squared errors between the predicted and actual values. Figure 18 below shows a bar chart that compares the MAPE and RMSE scores of both models.

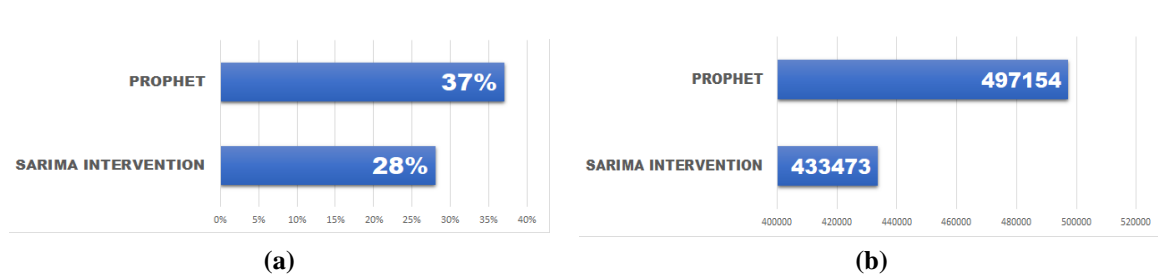


Figure 18. (a) MAPE and (b) RMSE value of SARIMA Intervention and Prophet Model

The SARIMA intervention model outperforms Prophet in terms of forecasting accuracy, as indicated by its lower MAPE and RMSE values. Several factors may contribute to this. Firstly, SARIMA is a more established model than Prophet. SARIMA has been around for decades and has been used to forecast a variety of different types of data [19]. Prophet is a newer model that was developed specifically for forecasting social and economic data. As a result, SARIMA may have been better able to capture the unique characteristics of airline passenger data during the COVID-19 period. Secondly, SARIMA allows for the use of intervention terms. Intervention terms are used to account for sudden changes in the data that are not captured by the model has underlying trend and seasonality components. The COVID-19 pandemic caused a sudden and significant decrease in airline passenger demand. SARIMA can account for this change may have led to more accurate forecasts than Prophet, which does not support intervention terms. Thirdly, Prophet is not as well-suited for forecasting short-term trends. SARIMA models can effectively capture and predict patterns and fluctuations in the data over a relatively short period [20]. Prophet is a good model for forecasting long-term trends, but it is not as well-suited for forecasting short-term trends. The COVID-19 pandemic caused a sharp decline in airline passenger demand that lasted for several months. Prophet may have struggled to accurately forecast this short-term trend, while SARIMA may have been better able to do so.

In this case, the SARIMA intervention model has a MAPE of 28% and an RMSE of 433473. This means that the model's forecasts are, on average, 28% off the actual values, and the values are spread out by 433473. While this is not perfect accuracy, it is still reasonably good, especially considering the impact of COVID-19 on the airline industry.

4. CONCLUSIONS

The best model for forecasting the number of domestic passengers at Soetta Airport is a time series model called SARIMA $(0,1,1)(1,0,0)^{12}$ with intervention orders $b=0$, $s=20$, and $r=0$. This model achieved a forecast accuracy of 28% measured by MAPE and 433.474 measured by RMSE. These results were obtained from a short-term forecast of 15 data points. Future research should expand the scope of the intervention period until December 2022, coinciding with the end of the PPKM measures, and obtain a sufficiently long test data set. This is expected to improve the accuracy of the forecasts. However, please note that the relevance of these recommendations may vary depending on the evolving conditions in the aviation industry and the global health situation. Continuous monitoring and periodic updates to the forecasting model will be essential to adapt to changing circumstances.

REFERENCES

- [1] F. Rianda, "Pemodelan Intervensi Untuk Menganalisis dan Meramalkan Jumlah Penumpang Pesawat di Bandara Soekarno-Hatta Akibat Pandemi Covid-19," *Seminar Nasional Official Statistics*, vol. 2021, no. 1, pp. 283–292, Nov. 2021, doi: 10.34123/semnasoffstat.v2021i1.857.
- [2] D. Andaka, "DAMPAK PELARANGAN MUDIK AKIBAT PANDEMI COVID19 TERHADAP BISNIS ANGKUTAN UDARA DI INDONESIA," 2020.
- [3] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 1st ed. Melbourne: OTexts, 2021.
- [4] A. Iswari, Y. Angraini, and M. Masjkur, "Comparison of The SARIMA Model and Intervention in Forecasting The Number of Domestic Passengers at Soekarno-Hatta International Airport," *Indonesian Journal of Statistics and Its Applications*, vol. 6, no. 1, pp. 132–146, May 2022, doi: 10.29244/ijsa.v6i1p132-146.
- [5] N. Faizin, A. Fauzan, and A. H. Primandari, "IMPLEMENTATION OF THE STEP FUNCTION INTERVENTION AND EXTREME LEARNING MACHINE FOR FORECASTING THE PASSENGER'S AIRPORT IN SORONG," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 1, pp. 0535–0544, Apr. 2023, doi: 10.30598/barekengvol17iss1pp0535-0544.

- [6] G. Christie, D. Hatidja, and R. Tumilaar, "Penerapan Metode SARIMA dalam Model Intervensi Fungsi Step untuk Memprediksi Jumlah Pegunjung Objek Wisata Londa (Application of the SARIMA Method in the Step Function Intervention to Predict the Number of Visitors at Londa Tourism Object)," *JURNAL ILMIAH SAINS*, vol. 22, no. 2, p. 96, Aug. 2022, doi: 10.35799/jis.v22i2.40961.
- [7] V. P. Ariyanti and Tristyanti Yusnitasari, "Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 405–413, Mar. 2023, doi: 10.29207/resti.v7i2.4895.
- [8] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*. Cham, Switzerland: Springer, 2017.
- [9] S. J. Taylor and B. Letham, "Forecasting at Scale," *Am Stat*, vol. 72, no. 1, pp. 37–45, Jan. 2018, doi: 10.1080/00031305.2017.1380080.
- [10] C. B. Aditya Satrio, W. Darmawan, B. U. Nadia, and N. Hanafiah, "Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 524–532. doi: 10.1016/j.procs.2021.01.036.
- [11] Badan Pusat Statistik, "Jumlah penumpang pesawat di bandara utama," bps.go.id. [Online]. Available: <https://www.bps.go.id/indicator/17/66/1/jumlah-penumpang-pesawat-di-bandara-utama.html>
- [12] J. I. Vélez, J. C. Correa, and F. Marmolejo-Ramos, "A new approach to the Box–Cox transformation," *Front Appl Math Stat*, vol. 1, Oct. 2015, doi: 10.3389/fams.2015.00012.
- [13] S. W. Astuti and Jamaludin, "Forecasting Surabaya - Jakarta Train Passengers with SARIMA model," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Sep. 2018. doi: 10.1088/1757-899X/407/1/012105.
- [14] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. New York: Springer, 2016.
- [15] M. J. Brewer, A. Butler, and S. L. Cooksley, "The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity," *Methods Ecol Evol*, vol. 7, no. 6, pp. 679–692, Jun. 2016, doi: 10.1111/2041-210X.12541.
- [16] A. N. A. Mahmad Azan, N. F. A. Mohd Zulkifly Mototo, and P. J. W. Mah, "The Comparison between ARIMA and ARFIMA Model to Forecast Kijang Emas (Gold) Prices in Malaysia using MAE, RMSE and MAPE," *Journal of Computing Research and Innovation*, vol. 6, no. 3, pp. 22–33, Sep. 2021, doi: 10.24191/jcrinn.v6i3.225.
- [17] S. Bouhaddour, C. Saadi, I. Bouabdallaoui, F. Guerouate, and M. Sbihi, "Tourism in Singapore, prediction model using SARIMA and PROPHET," in *VII INTERNATIONAL CONFERENCE "SAFETY PROBLEMS OF CIVIL ENGINEERING CRITICAL INFRASTRUCTURES" (SPCECI2021)*, AIP Publishing, Mar. 2023, p. 020042. doi: 10.1063/5.0131288.
- [18] J. Avigad, J. Hölzl, and L. Serafin, "A Formally Verified Proof of the Central Limit Theorem," *J Autom Reason*, vol. 59, no. 4, pp. 389–423, Dec. 2017, doi: 10.1007/s10817-017-9404-x.
- [19] J. Sanchez, *Time Series for Data Scientists*. Los Angeles: Cambridge University Press, 2023. doi: 10.1017/9781108942812.
- [20] H. Levenbach, *Change & Chance Embraced: Achieving Agility with Demand Forecasting in the Supply Chain*. Atascadero: Createspace, 2017.