Dengue Spread Model using Climate Variables

by

Ashaari Muhammad Bin Ayob
16159

Dissertation submitted in partial fulfilment of
the requirements for the
Degree of Study (Hons)
(Electrical & Electronics)

MAY 2016

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Dengue Spread Model using Climate Variables

By

Ashaari Muhammad Bin Ayob
16159

A project dissertation submitted to the

Electrical & Electronics Engineering Programme

Universiti Teknologi PETRONAS

in partial fulfilment of the requirement for the

BACHELOR OF ENGINEERING (Hons)

(ELECTRICAL & ELECTRONICS)

Approved by,

____

(Dr Vijanth Sagayan Asirvadam)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

MAY 2015

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgement, and that original work contained herein have not been undertaken or done by unspecified sources or persons.

(ASHAARI MUHAMMAD BIN AYOB)

ABSTRACT

Many cases related to dengue reported each year, threatening the life of certain victims. The number of victims increase annually and become biggest issue in early 21^{st} century until today. Therefore, the project takes initiative to develop model that can predicts dengue outbreak. This project aims to deliver the early warning system for possible dengue outbreak, thus enhance the efficiency of primary dengue surveillance system. The forecasting model using Autoregressive Integrated Moving Average (ARIMA) is generated using the data acquired from Ministry of Health Malaysia which consisted of data from January 2010 to December 2011 in Selangor. Selangor is chosen because of the rising in case of dengue and has the most population compared to other state. The model is then validated using the data from January 2012 to June 2012. The relationship between dengue incidence with climate variable is examined using cross correlations to reducing the error forecasted. The result of this study revealed that ARIMA $(1,1,2)(1,1,0)_{20}$, ARIMA $(1,0,1)(1,1,0)_{20}$, ARIMA $(0,1,1)(1,1,1)_{20}$, ARIMA $(1,0,1)(1,1,1)_{20}$, ARIMA $(1,0,2)(0,1,1)_{20}$, ARIMA$(1,0,2)(0,1,1)_{20}$ and ARIMA$(1,0,1)(0,1,1)_{20}$ is respectively the result for Petaling, Gombak, Klang, Kuala Selangor, Hulu Langat, Hulu Selangor and Sepang. The cross-correlation of mean temperature, humidity and rainfall shows positive correlation for most district at different lag. However, humidity does not correlate with dengue cases at Petaling, Gombak, Hulu Langat and Sepang. Forecasted warning based on the data can be applied in real situation that could assists community in improving vector control, public awareness and personal preservation.

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude to the following persons who have support and help me to complete my Final Year Project successfully. First and foremost, I would like to extent my heartfelt gratitude to my parents Khoriah Binti Abdul Razak and Ayob Bin Mat Isa for always supporting and give positive advices for me towards completing this project. Moreover, I also would like to extent my gratitude to my supervisors, Dr Vijanth Sagayan Asirvadam for giving me the most thorough guidance and constant supervision in order to complete this project.

Although my major is in power, I am able to expand my scope into other areas for my FYP 2 thus, gain a little more experience. This is all possible thanks to my supervisor for providing a title with a statistical based subject. In addition to my major knowledge obtained in lectures, I also acquire some understanding in statistical area which could be useful in the future. Without his never ending support, the project would have been hard to complete.

Special thanks to the FYP coordinator, Dr Norasyikin Binti Yahya as she always guide the electrical & electronics engineering's student by providing us with information and reminders. Moreover, thanks to the Ministry of Health for providing valuable data of dengue cases recorded in Selangor for this project. Not to forget to all of the examiners who help me a lot to improve my project through my mistakes. The knowledge they shared to us had opened our mind to make the project more successful.

Last but not least, I am highly in debt to the lab technicians, lab supervisors, friends and those who had helped directly or indirectly throughout the period of finishing the project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background Study

Dengue virus (DENV) is the most common disease widely spread throughout the world [1] especially subtropical and tropical areas of the world which having suitable climate for mosquito breeding. Coming from just one type of mosquito that is Aedes mosquito, dengue is spread among human systemically [2, 3]. Previous research done by Lounibos, L. Philip stated that Aedes aegypti is closely related to human habitats which is usually found in domestic area all around the world except Africa [3]. This somehow proves that human habitats can be one of the suitable breeding factors of this species. DENV acknowledged by many researchers to be very dangerous and fatal to human[4].

People commonly used many instruments and equipment to control the DENV problem at homes. The most popular method is using insecticide sprays. This method not only causes harm to mosquito but it can also do serious damage to humans. Similar method also applied in a large scale such as fogging. Most government including Malaysia consider fogging is the best method for controlling the mosquito breeding in the outdoor space of residential area especially the drain and enclosed space.[5]

Some control measure used may not be suitable at some situation and may cause some negative issues. For instance, usage of insecticide and fogging may cause harm to human as well as the environment because of the chemical contents. Thus, the world needs a prediction of the forecast model which can predict hot zone of high potential for DENV to spread. The early warning system for possible dengue incidence can create awareness to the public in the locality and thus will reduce dengue fever and dengue haemorrhage fever rate. This study is focusing on the modelling dengue prediction using climate as the variables to highlight their risks.

1.2 Problem Statement

In order to decrease the victims of dengue disease, major control measure should be assisted with predictive model to know the proper period of outbreak will happen. Many cases related to dengue reported throughout the world. Some cases resulting in death, some are not. The disease is spreads commonly influenced by precipitation, temperature and unplanned rapid urbanization. Recently, news reported that there are 75 per cent rise in dengue cases in Selangor which recorded 6,686 cases over in January 2015 compared to 3,813 cases during January 2014[6]. The rapid increase in dengue case inspired this paper to successfully create a model which can predict any dengue outbreak.

1.3 Objectives and Scope of Study

The major objective of this study is to create a program for a vector borne disease management system and to evaluate predictive models for early warning dengue incidence using forecasting models. The variables which helps determined the possible risk is climate. The scope of study for this project is to identify the suitable model based on database of climate provided. In order to achieve the main objective, some other objective needs to be highlighted:

1. To analyse data and generate a predictive model from the data that reliable as early prediction of DENV.
2. To design a system that can identify and classify the potential level of DENV transmission in Selangor.

1.4 Relevancy and Feasibility of the Project

Student is expected to implement ARIMA model to predict dengue outbreak for several state in Selangor within two semesters. In reference to the task planned in methodology section, the project is considered achievable and practical for final year student. However, the knowledge and skill of the author is limited as there are no specific course on statistical method such as ARIMA. Despite of that, knowledge can be gain through research and continuous effort since basic statistical tools has already been taught in academic structure. Moreover, the time period allocated is sufficient to complete the project.

CHAPTER 2
LITERATURE REVIEW & THEORY

Dengue is transmitted by the mosquito that carry the dengue virus from other people infected with dengue. Dengue only transmitted by mosquito with species ‘aedes aegypti’, that only breed under clean water. The dengue carrier among the species is female type that not only feed on fruit but also blood whether human or not. Dengue is hazardous illness that affects infants, children and adults with the symptoms after 3-14 days bitten. Until today, vaccine is still not found to treat dengue, except preventive measure and reducing the fever before it comes to final stages.

Inclusion of climate condition as one of the variable affect the rate of dengue cases is vague as not many research and study is done that simply relate the two. However, climate is said to has complex influence with the dengue and aedes aegypti breeding factor. For this study, the amount of time for dengue transmission is important to know and guess correctly the outbreak period. Thus, knowing the life cycle of an aedes aegypti is a must so that the model predicted is relatable and did not contradict with the theory.

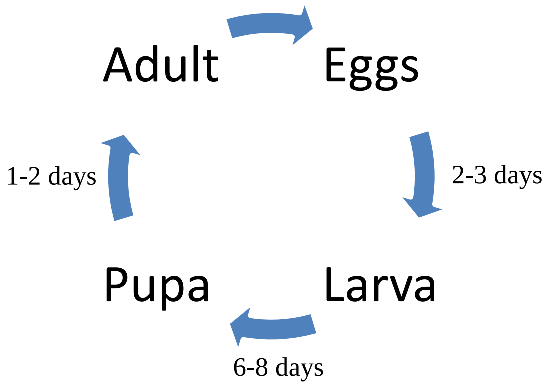


FIGURE 2.1. Life Cycle of an Aedes Aegypti with time period

Generally, forecast use previous statistics to determine the direction of future trends. It is a decision making tool that attempts to deal with the ambiguity of the future which will benefit the users either directly or indirectly. Forecasting method might refer to formal statistical method employing one or more technique such as time-series, Delphi method, moving average, exponential smoothing, regression analysis, trend projection, cross-sectional and longitudinal data. Every prediction has its own advantages and disadvantages which contributed to error percentage that need to be proved through research and study. Sensitivity technique analysis is always implemented after every forecasting which selects a range of possible values. This section will discuss some selected method which is possible to be implemented in the project.

2.2 Regression Analysis

Regression analysis is an analytical tool for modelling and interprets the connection between independent and dependent variables. This analysis frequently uses to estimates the dependant variable values when the independent variables are constant. Regression analysis also interest to identify the range of dependent variable which relates to probability distribution. There is much evidence that has proven the reliability of this analysis technique which provides useful forecasts. However, in some cases where the regression-based analysis is used, may lead to overconfidence and contribute to the error [7].

2.2.1 Poisson Regression Model

Earnest et al. [8] used Poisson regression model to discover the affiliation between El-Nino Southern Oscillation (ENSO) indices, climate variables and DENV. The authors found that climate variable and ENSO were contradicting with dengue cases. They conclude that any climate variable considered will be having identical predictive ability since the climate variable used which is temperature, humidity and precipitation is surpassing the others.

Poisson regression using generalized additive model (GAM) model have been practiced in examining the association between rainfall and dengue by a study done by Chen et al. [9]. This model allows the evaluation of the multiple-lag effects of layered rainfall levels on particular illness. GAM is used to minimize the error percentage of a dependent variable from the predictor variables which let a Poisson regression suitable as a total of nonparametric smooth functions of predictor variable. They found that with increasing the risks of dengue the differential lag behaviour is observed based on rainfall level. Because of that, whole model was adapted for assessing between type of severe rainfall and diseases for the multiple-lag effects of temperature, month and area.

2.2.2 Linear Regression Model

Linear regression technique is an approach to examine the association between independent and dependent variable. There are two types of linear regression which is simple and multiple linear regressions. Single independent variable is used to predict the value of a dependent variable for simple linear regression while for two or more independent used is called multiple linear regressions. The only difference between the two types is the number of independent variable used. Colon-Gonzalez et al. [10] used multiple linear regression models in their study to relate the changes in climate variability with dengue incidence reported. Their results prove that cool and dry season is the optimum dengue incidence happen in Mexico. The outbreak is highly associated with the strength of El-Nino. Linear regression often used in practical application due to simple statistical calculation and because of the statistical properties which is easier to determine.

Previous study by by Hii et al. [11] has proved the reliability of using climate as the variable to forecasting dengue outbreak. In their study, apart from only using climate variable which is temperature, humidity and rainfall, the study also included autoregression, seasonality and trend in order to determine the final model of the forecasting DF. They have successfully developed a time series Poisson multivariate regression model based on the independent variable which combined together to form the model.

There are three main process involved in his study: model development and training using data from 2010-2011, model validation by predicting cases in 2012. During the model construction process, Hii et al. formulated bivariate equation using quasi Poisson regression for each of the element considered before combined them to form the multivariate model. The model then applied during the validation process to predict the dengue cases from week 1 of 2011 until week 16 of 2012 using only climate data. The model is predicting correctly during training with standard deviation errors of 0.3 and validation period with errors of 0.32 of reported cases.

Despite concentrating only on climate data, many research support the model by adding some other independent variable such as population, socio-demographic factors, Nazri et al. [12] reported in his study that land use and housing type has much influenced in dengue epidemic outbreak apart from climate changes. Comparing dengue outbreak cases with each climate parameters prove that humidity has low contributing factor influencing dengue comparing to temperature and rainfall level. This proves that not all climate variables are reliable enough to produce a perfect model without considering other factors contributing to dengue outbreak in Selangor.

2.3 Time Series Models

The forecasting model in this project will be done using time series model which is autoregressive integrated moving average **(**ARIMA). An ARIMA model estimate a value as a direct combination in a time series of former values, errors and present and past values of other time series. This model evaluates and predicts a variable time series data, intervention data and transfer function using ARIMA technique.

Time series analysis has been applied broadly in determining the effect of climate variable on DENV. As an example, a study by Hu et al. [13] utilize a seasonal autoregressive integrated moving average (SARIMA) model to test the relationship between El-Nino and dengue related cases from 1993-2005. They found that a lower Southern Oscillation Index (SOI) was associated to positive dengue outbreak in Queensland, Australia. Similarly, Gharbi et al. [14] fitted the SARIMA model approach to relate the climate variable (temperature, rainfall and humidity) with the DENV in French West Indies. They suggested that temperature reduced the limitation of the model considerably to forecast dengue cases. 5 weeks lag time with minimum temperature shows the most excellent pattern for dengue prediction. This past research shows that using ARIMA model is feasible and reliable as prediction tools.

ARIMA is said to be more reliable as statistical modelling approach compared to trend fitting approach even though ARIMA model is said having model specification error[15]. Because of this, Box and Jenkins has efficiently created a guidelines using the related information and thorough analysis on ARIMA model for the understanding among public[16]. Moreover, the guidelines making it possible for non-statistical people to apply this type of model for prediction. The accuracy of ARIMA models in modelling temporal structure especially for seasonal disease have been proved compared to other statistical tools[15, 17, 18]. There is some disease that has been successfully forecasted using ARIMA models. Some of the includes pneumonia deaths[17], malaria and hepatitis A[18].

Introducing climate variable as external regressive that influence the development of dengue outbreak can increase the accuracy of the predicted models[19]. Increase rainfall rate is positively associated with the dengue incidence[20-22]. Moreover, some study has also justified that temperature has some correlation with dengue especially in tropical country[20, 22]. The relationship among climate variable with dengue has been studied widely in the past several decades, but how much influence it brings still need to identified and explored.

CHAPTER 3
METHODOLOGY/ PROJECT WORK

This section will introduce the procedure and design approach used in selection of the model. The model will then estimate few model and looking at the characteristic checking to improve the model. The original meaning of modelling is to categorize the pattern of the data, identify suitable type of model that is created and influencing that pattern. The model that will be built should be compared with the history that can be projected into the future. There also several variable that should be considered after selecting the suitable model which is residuals. A residual is the difference between the actual values and fitted values. The values of residual should be around zero indicating that the model has already captured the pattern.

3.1 ARIMA model

ARIMA models consists of three different parameters namely: Autoregressive parameters (AR), Moving average parameters (MA) and differencing parameters (I). AR parameters is the lags of the stationary series while MA parameters model is the lags of the forecasts error. The differencing parameters is the amount of differencing applied to be made stationary. The usage of each parameter reflect the type of ARIMA models to be used which can vary from AR models, MA models or even ARMA models. The model will be referred as an AR model when only autoregressive model involves. When it only involves moving average terms, it may be called as MA model. Lastly, the presence of both terms without applying any differencing is called as ARMA model. This paper will be applying Box-Jenkins Methodology as the main reference.

3.1.1 Autoregressive (AR) model

Autoregressive (AR) model are considering the value of variable in one period together with the values in previous periods. Usually denoted by AR(p) which means the autoregressive model with p lags.

$$y_t = \mu + \sum_{i=1}^{p} \gamma_{t-i} + \epsilon_t \; ; \tag{1}$$

where μ is a constant and γ_p is the coefficient for the lagged variable in time $t - p$.

For example, AR(1) is expressed as: $y_t = \mu + \gamma_{t-1} + \epsilon_t$;

3.1.2 Moving Average (MA) models

Moving Average (MA) models relate the connection between a variable and the residual of past period. It is commonly denoted as MA(q) which means moving average model with q lags.

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \; ; \tag{2}$$

Where θ_q is the coefficient for the lagged error term in time $t - q$.

For example, MA(1) is expressed as: $y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$;

3.1.3 Autoregressive moving average (ARMA) models

This models combine both autoregressive terms (p) and moving average terms (q), also denoted as ARMA(p,q). It's considering the dependent values with the lag p and the residual error with the lag q.

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}; \tag{3}$$

Where its include both the dependent term and residuals term in the formula.

Each type of model has their own pros and cons depends on the application used. Besides that, selection of suitable model can reduce the amount of parameters used, that can give lowest error percentage. The types of ARIMA models is summarize in the Table 1:

TABLE 3-1. Summary of Type of ARIMA Model

	AR	**I**	**MA**
Definition	• Autoregressive • Lags of the stationaries series	• Integrated • A series which needs to be differenced to be made stationary	• Moving average • Lags of the forecast errors
Parameters	p, P	d, D	q, Q
AR(p)	✓		
MA(q)			✓
ARMA(p,q)	✓		✓
ARIMA(p,d,q)	✓	✓	✓

3.2 Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Selection of suitable ARIMA model is done through autocorrelation. A good way to distinguish between signal and noise is ACF (AutoCorrelation Function). This is developed by finding correlation between a series with its lagged values. Comparing the autocorrelation function (ACF) and partial autocorrelation function (PACF) pattern enable us to estimate some possible ARIMA model. The correlation statistics is used to identify the stochastic pattern in the data. The correlation between consecutive months is considered ACF of lag 1. Consider a set of data for last year, the correlation would be called the ACF of lag 12. Meanwhile, PACF controlling the values of its own lagged values together with all shorter lags. For example, a regression using a lag of 12 is not only focuses purely on that lag, but also consider all the lags before from 1 to 11 too. By knowing the autocorrelation of previous data, we can examine their relationship and improvise by adding related parameters to the model. The autocorrelation commonly arranged together for different lags visualize either as a bar chart or line chart. The chart is known as correlogram showing certain amount of confidence interval.

3.2.1 Autocorrelation Function (ACF)

ACF is the proportion of the autocovariance of y_t and y_{t-k} to the variance of a dependent variable y_t

$$ACF(k) = \rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} \tag{5}$$

Very slow decay of ACF indicated the non-stationary.

3.2.2 Partial Autocorrelation Function (PACF)

PACF is the simple correlation between y_t and y_{t-k} minus the part explained by the intervening lags

$$\rho_k^* = Corr[y_t - E^*(y_t|y_{t-1}, \dots, y_{t-k+1}), y_{t-k})] \tag{6}$$

Where $E^*(y_t|y_{t-1}, \dots, y_{t-k+1})$ is the minimum mean-squared error predictor of y_t by $y_{t-1}, \dots, y_{t-k+1}$.

3.2.3 Behaviour of ACF and PACF Properties for Estimating ARIMA Models

TABLE 3-2. ACF and PACF Properties

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts of after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

3.3 Box-Jenkins Methodology for ARIMA Model Selection

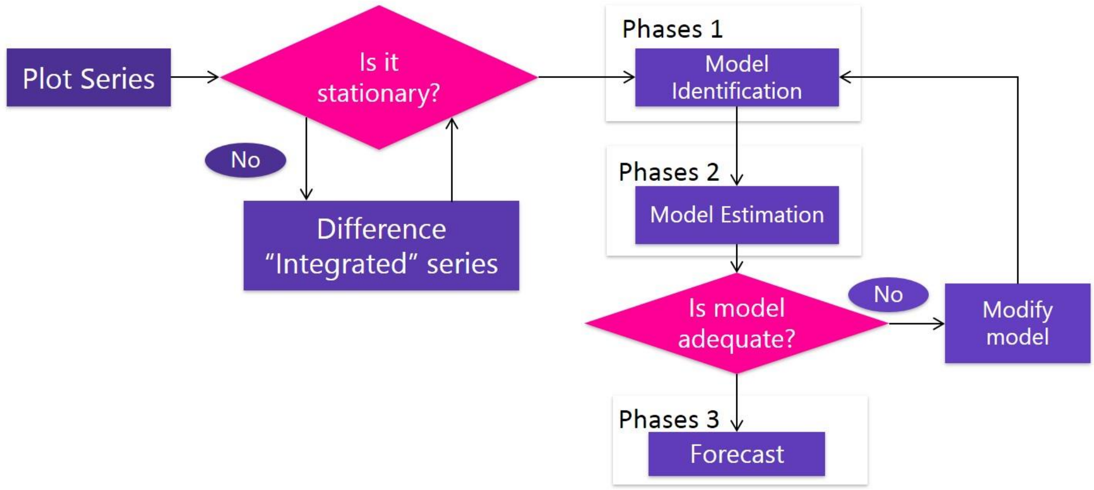


FIGURE 3.1. Box-Jenkins Model Building Process

Box-Jenkins model or methodology is divided into three steps which is model identification, model estimation and diagnostic checking step before the model can be forecasted.

3.3.1 Identification Step

In this step, the time plot of the series is examined to identify outliers, missing values and structural breaks in the data. The pattern is observed for the stationarity and transform using logs, differencing or detrending if not stationary. Differencing the data can remove trends and ease to classify the pattern of the data. However, over-differencing may introduce dependence when none exists. There is also case where seasonal difference has to be applied. Seasonal difference only valid for estimating Seasonal ARIMA model. Any pattern shows seasonality properties has to apply seasonal difference for the P,D, and Q parameters. Next is to examine the autocorrelation function (ACF) and partial autocorrelation function (PACF) behaviour. The ACF and PACF sample is observed based on the behaviour of ACF and PACF in Table 2 to estimate plausible models and select appropriate p,d, and q.

3.3.2 Estimation Step

This step introduces ARMA models estimating by examining various coefficients. The goal is to select stationary and parsimonious model that has significant coefficient and a good fit. There are several test can be done to check for stationarity and one of them is using Dickey-Fuller test. This step introduces several goodness of fit test. The model is examined for goodness of fit using Akaike Information Criteria (AIC). AIC measure the trade-off between model fit and complexity of the model

$$AIC = -2\ln(L) + 2k \tag{7}$$

The model with most parsimonious and lowest AIC is selected.

3.3.3 Diagnostic Checking Step

Residuals is considered in this step to increase the accuracy of the model. If the model fits well, then the residuals from the model should resemble a white noise process. Here, the residuals are checked for normality looking at the histogram and check for independence by examining ACF and PACF of the residuals. Besides, Ljung-Box-Pierce statistics can also be performed to check the residuals from the time series model is white noise.

3.4 Project Methodology

This project is a statistical analysis project which using R, a statistical software for model fitting and forecasting. The forecasted result obtained from R will be presented in form of table for better understanding and analysing.

3.4.1 Research for the literature

Each forecasting model has its own advantage and disadvantage, the purpose of the study from another paper is to justify the works and observe the limitation and some other possible methodology. Lastly, each methodology studied can be compared and the most accurate model can be made as reference. There are many literature discusses on the forecasting model mainly on Linear Regression Model, [10] Time Series Model, [13, 14, 23-26] Poisson Regression Model [8, 9] ,Bayesian Model [27] and Non-linear model [28]. The validation of each model is also studied to ensure the most precise model with least error. Some validation technique is Standardized Root Mean Square Errors (SRMSE) [11, 14] and receiver operating characteristics curve (ROC) technique. [29]

3.4.2 R Statistical Computation

Modelling a forecast data using R takes two main tasks. First main tasks are for R retrieve the data and analyse it to create a graph of dengue outbreak over climate change. In order to improve the predictive power of the model, external climate variable is introduced. Three climate variables are considered in this project: mean temperature, humidity and rainfall. R will be using each data to plot three different graphs for the three variables. After the time series model of the forecast is constructed, the plotted graphs will be analysed and examine using cross-correlation to find the optimum lag between climate with dengue cases. The optimum lag offset is determined in this stage to predict the dengue cases. It is not necessarily the longer the lag offset the better the model is where it depends entirely on the error percentage for each lag offset.

3.4.3 Documentation and Report

The results obtain later will be documented and the trending will be observed and analysed. Error percentage will be evaluated and improved throughout the study period in order to arrive with an appropriate conclusion.

3.4.4 Data Usage

The data of registered dengue cases is obtained from Ministry of Health Malaysia. The data recorded the dengue outbreak in each Selangor district (Petaling, Gombak, Klang, Kuala Selangor, Hulu Langat, Hulu Selangor and Sepang) from 2010 to 2012. The data also include 3 common climate variable which is rainfall, mean temperature, humidity. All variable is categorized in weekly data throughout 3 years. The data will be divided into two roles: training data and validation data. Training data will be collected from data of 2010 until 2011 while validation data will include only the data of first half of 2012 (25 weeks). Training data will be used to create and determine the suitable ARIMA model and forecast. Validation data is useful to validate the accuracy of forecasted value and thus, improving the model if any increasing in error occur.

3.4.5 Process Flow

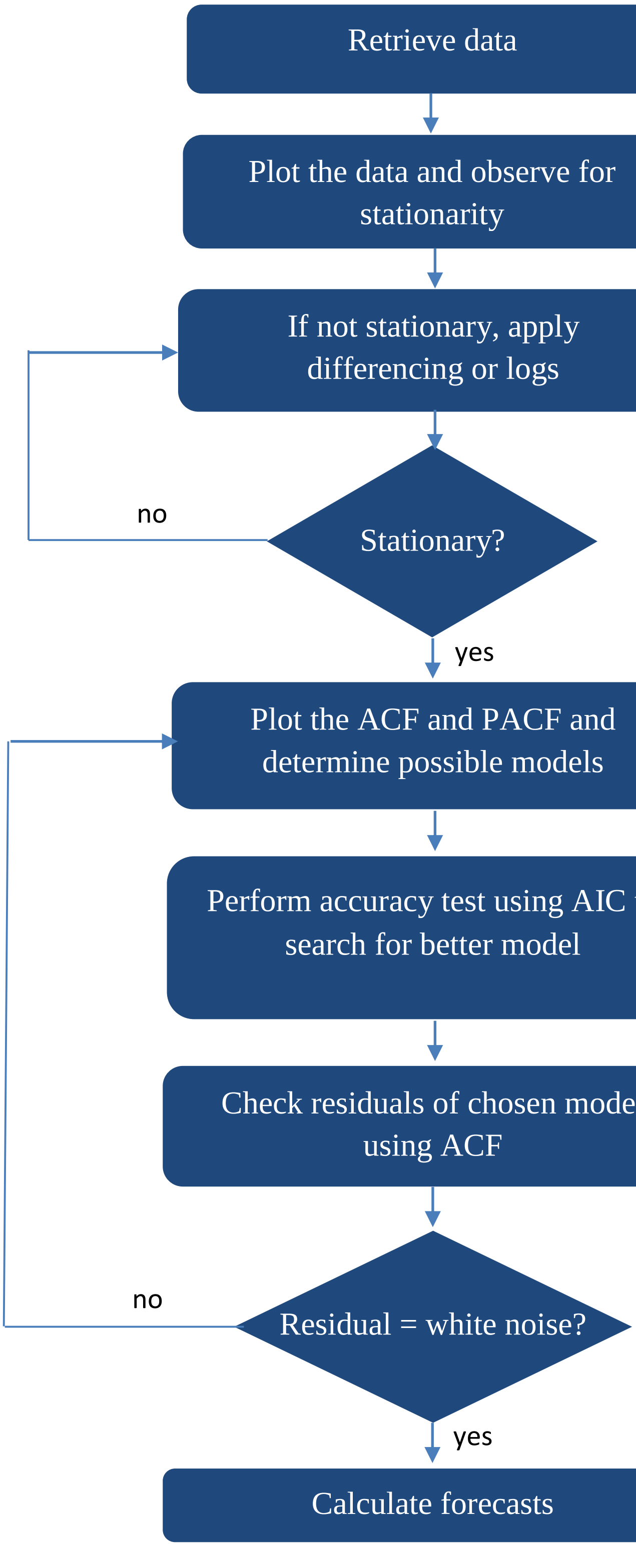


FIGURE 3.2. Power Flow Diagram

3.4.6 Gantt Chart and Key Milestone

TABLE 3-3. Final Year Project 1 Gantt Chart and Key Milestone

No	Detailed Work	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Project Selection	■													
2	Literature Review (research)		■	■	■	■									
3	Basic design methodology						■								
4	Research on ARIMA model							■	■	■					
6	Stationarity test on dengue outbreak data										■	■			
7	Examine ACF and PACF										■	■	■		
8	Comparing lowest AIC												■	■	■

TABLE 3-4. Final Year Project 2 Gantt Chart and Key Milestone

No	Detailed Work	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Analysis of residual ACF and PACF	■	■	■												
2	Selection of ARIMA model				●											
3	Extrapolate predicted pattern					■	■	■								
4	Progress Report Submission								●							
6	Cross correlations of climatic variables with dengue incidence								■	■	■					
7	Poster Presentation											●				
8	Model Comparison											■	■	■		
9	Submission of Final Report														●	
10	Viva															●

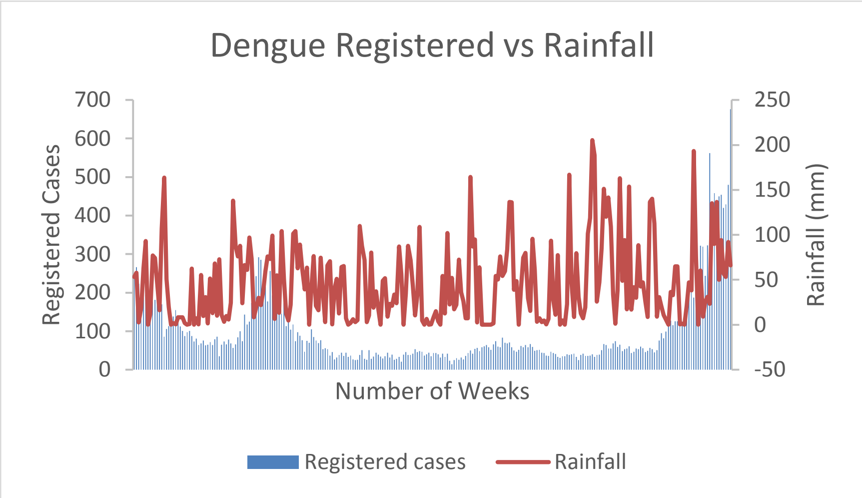
CHAPTER 4
RESULTS AND DISCUSSION

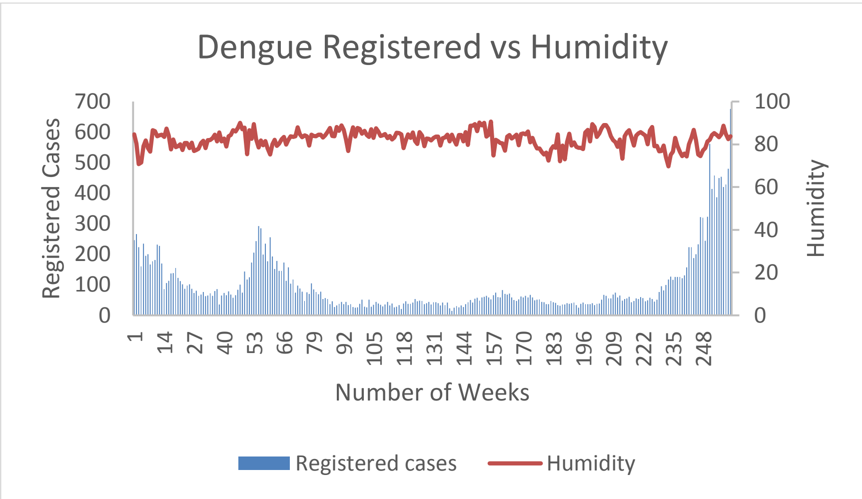
A total of 31991 dengue cases registered by Ministry of Health Malaysia from 2010 to 2012. The annual number of dengue cases is assessed and recorded in Table 5. The table shows that the dengue cases increase and decrease each year with 15689 cases is recorded on 2010. The difference of cases recorded of year 2010 with 2011 is bigger compared to difference between the other two years. It proves that dengue cases are decreases greatly from 15689 to 7498 cases only. However, each state has different number of dengue cases mainly because of different area of population. Petaling has recorded highest dengue cases in 2010 with 5147 cases.

TABLE 4-1. Number of Recorded Cases of Dengue Between 2010 and 2012 in Selangor

State	2010	2011	2012	Total No of Case each State
Petaling	5147	2066	2551	9764
Gombak	3107	1458	970	5535
Klang	1752	1371	2294	5417
Kuala Selangor	310	258	272	840
Hulu Langat	4852	1995	2242	9089
Hulu Selangor	300	261	261	822
Sepang	221	89	214	524
Total no of case each year	15689	7498	8804	
TOTAL				31991

Since the study emphasize the relationship of dengue cases with climate variables. Both of the data need to be assess for trend and seasonality pattern in order to identified their influence that may affect the overall final results on seasonal and long term time trends variation. The plot of weekly humidity and weekly mean temperature showed constant trend throughout the year, however the plot of weekly mean temperature display consistent with seasonal pattern. Precipitation plot against recorded cases is inconsistent and showed increasing trend with increasing year from 2010 until 2012 (Figure 4.1).





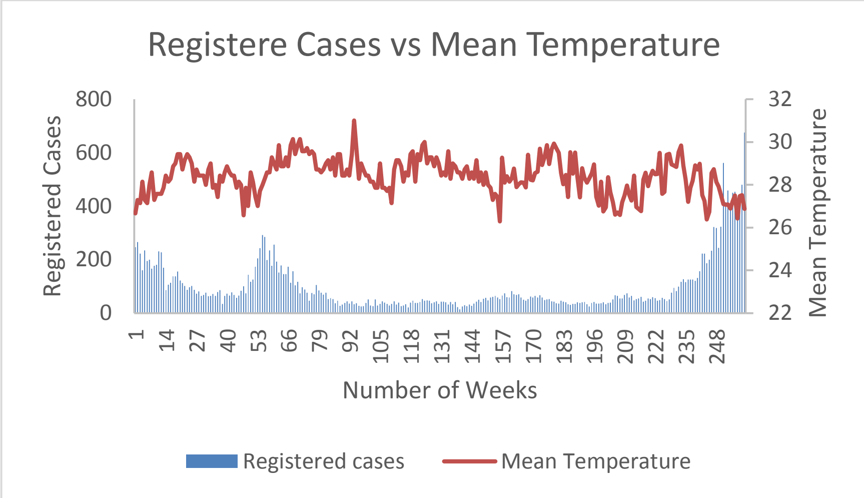


FIGURE 4.1. A: Weekly rainfall and reported cases, B: Humidity and reported cases, C: Mean temperature and reported cases in year 2010-2012

4.1 Petaling

4.1.1 Model Identification Step

In order to analysed the weekly data of recorded cases to perform ARIMA model, the pattern is observed for stationarity. Based on Figure 4.2A, it can be concluded that the data is not stationary. The stationarity is observed through the unstable variance, mean and autocorrelation. It is relatively simple to predict a stationaries series as its statistical properties will remain the same either for the past or future series. Apart from that, a time series with stationarity properties is capable to relate their means, variances and correlations with other variables. Since variable of dengue cases is not stationary, a common solution is by using differencing, either first order, second order or more. The number of differencing order is projected through the d parameter in ARIMA model. Figure 4.2 shows the data recorded before and after first differencing is applied. Finally, Figure 4.3B shows the stationary data of recorded cases in Petaling.

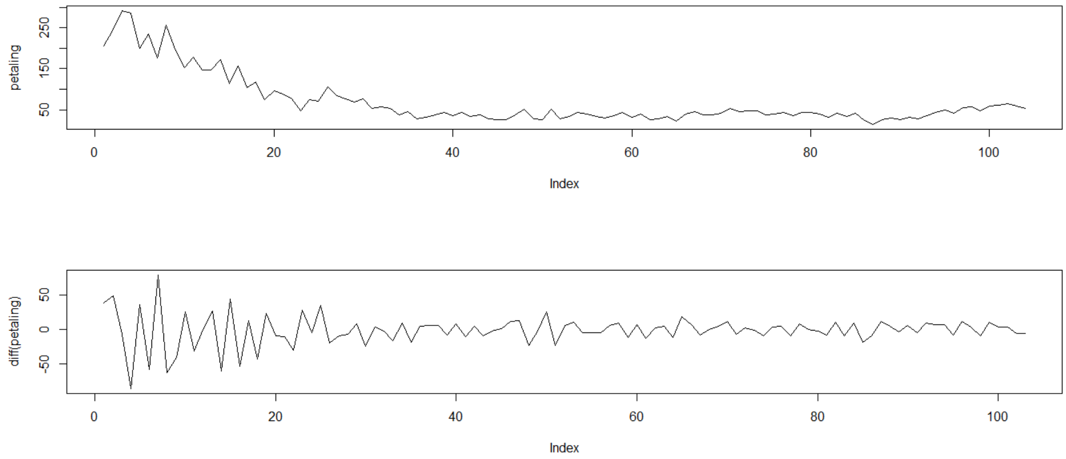


FIGURE 4.2. A: Plot of Original Dengue Cases, B: Plot of Dengue Cases after first differencing

The plot of original data shows insignificant trend for dengue incidence throughout the year. The differencing data with log transform has better pattern with constant mean. The order of differencing however may vary with different data. In this case, first order of differencing(d=1) is sufficient to prove their reliability.

However, in this data there are indefinite seasonality pattern and thus the pattern is assumed to be every 20 weeks. Hence, seasonality difference is applied in order to apply Seasonal ARIMA model.

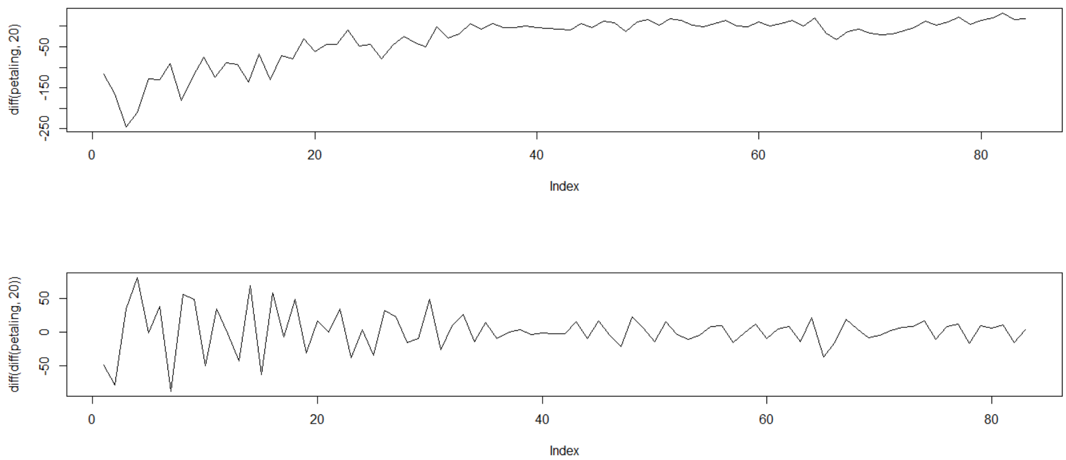


FIGURE 4.3. A: Plot of Dengue Cases after first seasonal difference, B: Plot of Dengue Cases after first differencing and first seasonal difference

Next is to determine the autocorrelation function (ACF) and partial autocorrelation function (PACF) to know which ARIMA model is most probably accurate. There are three possible ARIMA model as mention in the methodology which is AR, MA and ARMA. The properties of ACF and PACF is observed and analyse based on Table 2. The plot of ACF and PACF of original data (Figure 4.4A) shows that it is not stationary through their gradually slow decaying pattern. Thus, first differences data is preferable to perform ARIMA model.

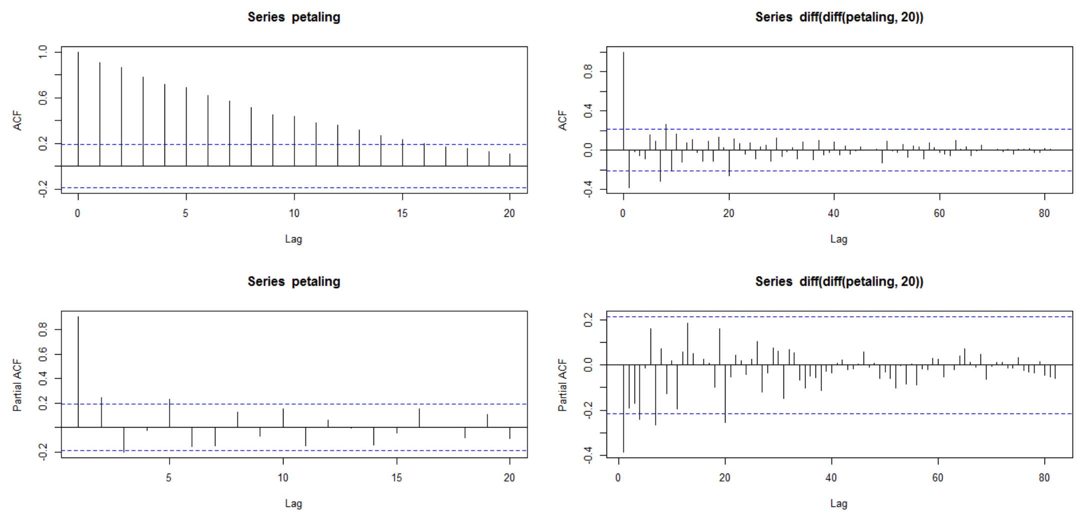


FIGURE 4.4. A: ACF and PACF plot of original dengue case recorded, B: ACF and PACF plot of original dengue cases recorded after first differencing

Before differencing, the ACF seems tails off after several lags and this shows the properties of non-stationary time series. After non-seasonality differencing and seasonal differencing applied to original data, a significant cut off is observed at one-week lag of ACF plot (Figure 4.4B). PACF also shows cuts off properties which is shown at lag one of the graph. According to ACF and PACF properties, cut-off ACF together with cuts-off PACF has the possibility of both AR and MA signature. The analysis from the correlograms suggests that p values should be 0 or 1 and q value should be 0, 1 or 2. Meanwhile, on the same ACF and PACF plot, the seasonality traits can be extract through observation in lag 20. The only lag is at lag 20 and not the 20 after, so the value of P and Q should both be 0, or 1. There are 15 possible models that can be considered from maximum value of p, P, q, and Q. The list of possible model is in Table 4.2.

4.1.2 Model Estimation Step

In order to confirm their reliability, several model may need to be considered. All possible ARIMA model is evaluate and examine using the Akaike Information Criteria (AIC). The most parsimonious model with the lowest value of AIC is most preferred.

TABLE 4-2. AIC values for different ARIMA model

Models	AIC	Models	AIC
$\mathrm{ARIMA}(0,1,1)(1,1,1)_{20}$	770.17	$\mathrm{ARIMA}(1,1,1)(0,1,1)_{20}$	772.19
$\mathrm{ARIMA}(0,1,2)(1,1,1)_{20}$	772	$\mathrm{ARIMA}(1,1,2)(0,1,1)_{20}$	768.7
$\mathrm{ARIMA}(1,1,0)(1,1,1)_{20}$	771.18	$\mathrm{ARIMA}(0,1,1)(1,1,0)_{20}$	768.3
$\mathrm{ARIMA}(1,1,1)(1,1,1)_{20}$	771.98	$\mathrm{ARIMA}(0,1,2)(1,1,0)_{20}$	770.19
$\mathrm{ARIMA}(1,1,2)(1,1,1)_{20}$	768.52	$\mathrm{ARIMA}(1,1,0)(1,1,0)_{20}$	770.4
$\mathrm{ARIMA}(0,1,1)(0,1,1)_{20}$	770.31	$\mathrm{ARIMA}(1,1,1)(1,1,0)_{20}$	770.18
$\mathrm{ARIMA}(0,1,2)(0,1,1)_{20}$	772.19	$\mathrm{ARIMA}(1,1,2)(1,1,0)_{20}$	766.97
$\mathrm{ARIMA}(1,1,0)(0,1,1)_{20}$	772.44		

*$\mathrm{ARIMA}(1,1,2)(1,1,0)_{20}$ is selected for having lowest AIC

Table 6 summarize the AIC values for each possible model. After all the possible model is tested, $\mathrm{ARIMA}(1,1,2)(1,1,0)_{20}$ seems to be reasonable as it have smaller AIC values of 766.97 compared to the other models. In addition, in order to get the most suitable model, the data can have additional confirmation with another accuracy tools such as BIC and mean absolute percentage error (MAPE).

4.1.3 Model Validation Step

In this section, the residual of selected model has to be white noise which is supposed to not having any spike out of the significance limits. Since there are several outside the region, the model may not be the most perfect. However, in some cases, the residual may not within the boundaries even with different closest model. The spike may contain some valuable information but comes in small quantity and can be neglected. Hence, the ARIMA$(1,1,2)(1,1,0)_{20}$ model is considered fine.

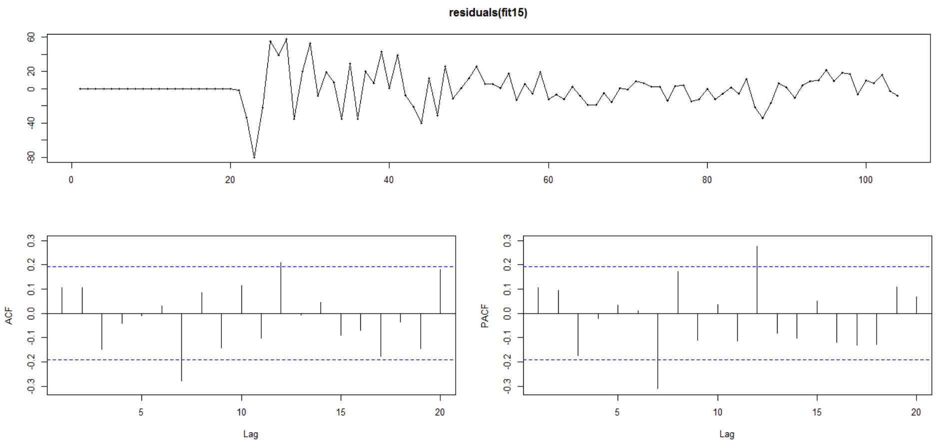


FIGURE 4.5. Residuals from the ARIMA$(1,1,2)(1,1,0)_{20}$ model

4.1.4 Influence of Climate Variable

Dengue cases is closely related to climate of the surrounding. Study always point that suitable habitual for dengue breeding is depends on the temperature, humidity and rainfall rate. Because of that, this study will prove if the three variable can influence the final result of forecasted data and their relationship.

a. Temperature

Cross-correlation original dengue cases data with temperature data does have any positive correlation with negative lag of -2 until -20 with peak correlation at lag -14. The correlation means that temperature lead the dengue cases by 14 weeks ahead. Applying the temperature data with 14 weeks lag as external regressors however does not improve the model at all. The cross-correlation function of dengue data with mean temperature is shown in Figure 4.6.

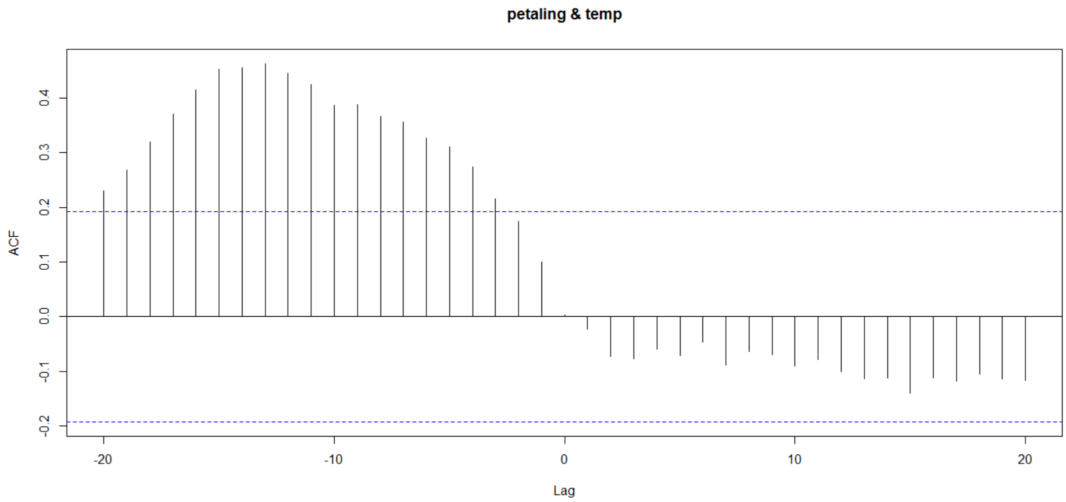


FIGURE 4.6. Cross-correlation between dengue cases recorded and mean temperature

b. Humidity

No cross-correlation between humidity with dengue cases present.

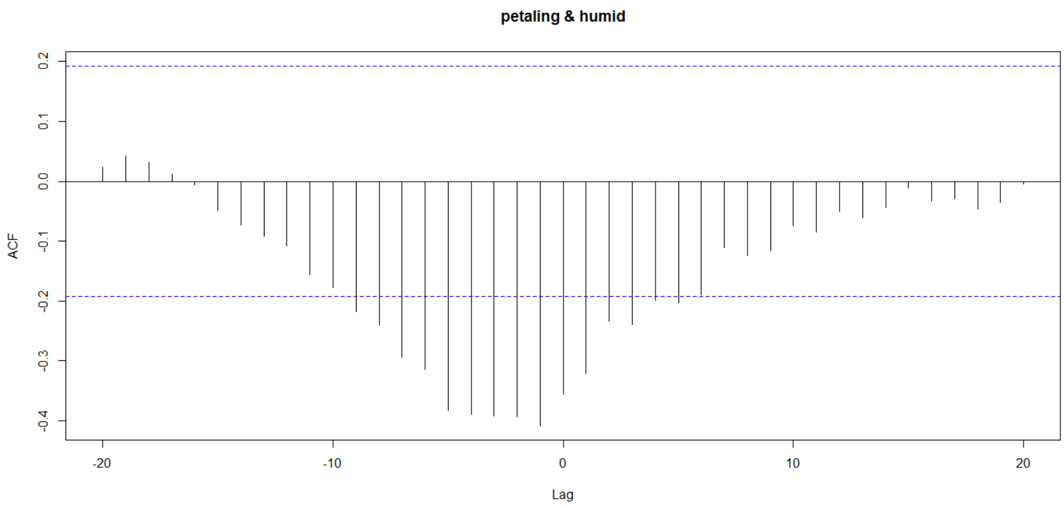


FIGURE 4.7. Cross-correlation between dengue cases recorded and humidity

c. Rainfall

The cross-correlation of rainfall and dengue cases is on negative lag of -5, -6, -10, -15 and -16 however the lag can be said as every -5 lag. Rainfall may lead the dengue cases by every 5 weeks as a suitable breeding factor to the dengue mosquito. With the breeding period of about 12 days, rainfall may contribute to increase in dengue cases every 5 weeks.

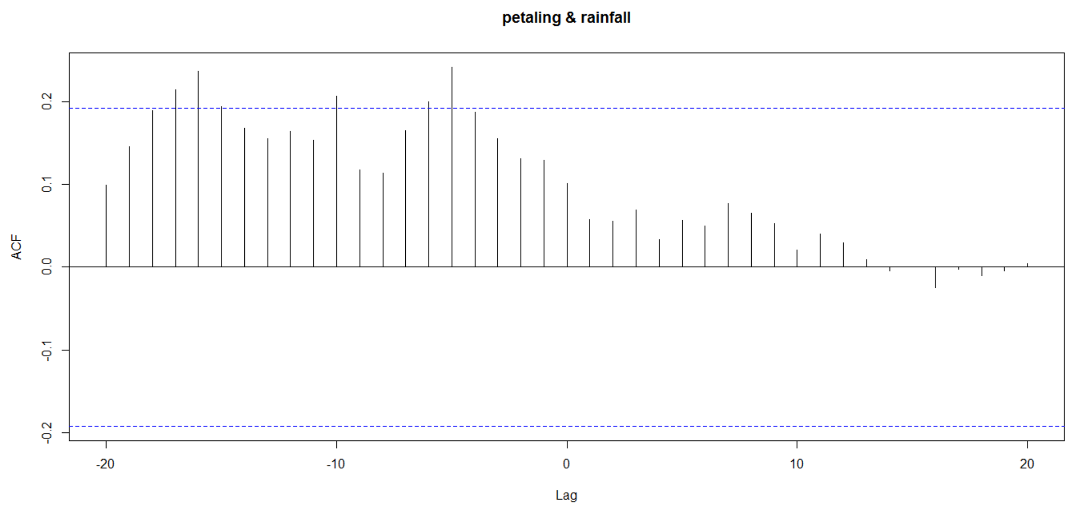


FIGURE 4.8. Cross-correlation between dengue cases recorded and rainfall

4.1.5 Model Forecasting Step (Univariate)

Forecasting 25 weeks ahead from ARIMA$(1,1,2)(1,1,0)_{20}$ model is plotted and shown in Figure 4.9.

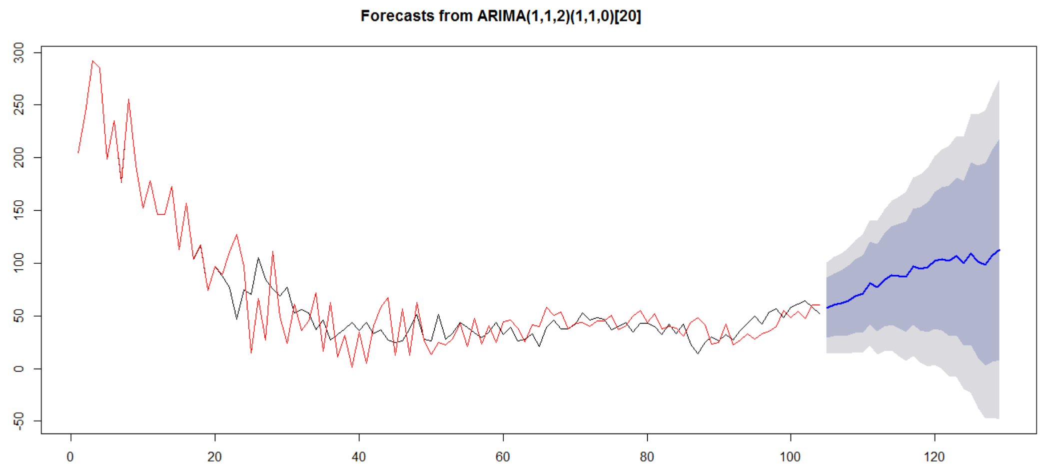


FIGURE 4.9. Forecasts from the ARIMA$(1,1,2)(1,1,0)_{20}$ model applied to the dengue recorded case data.

Based on Figure 4.9, it has predicted the range of the future data. The data forecasted has shown several pattern as referred to previous data. The forecasted value shows highest value at 129th weeks of the data. The fitted line (red line) is observed to fit well with ARIMA selected in later weeks. This is because the 'fitted' command is fitting the value using ARIMA produced from the data itself and refer 20 weeks before of its own value.

4.1.6 Model Forecasting Step (Multivariate with Temperature)

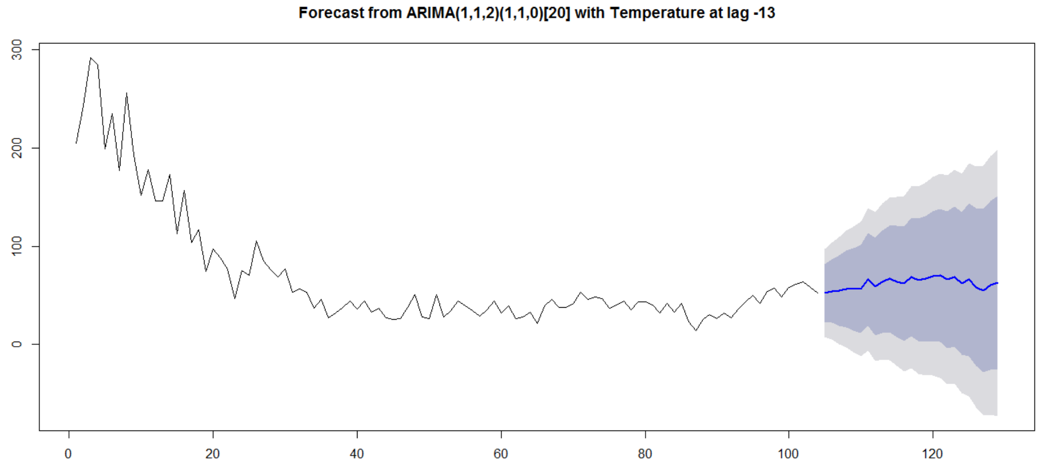


FIGURE 4.10. Forecasts from the ARIMA$(1,1,2)(1,1,0)_{20}$ model with temperature variable of lag -13 applied to the dengue recorded case data.

4.1.7 Model Forecasting Step (Multivariate with Rainfall)

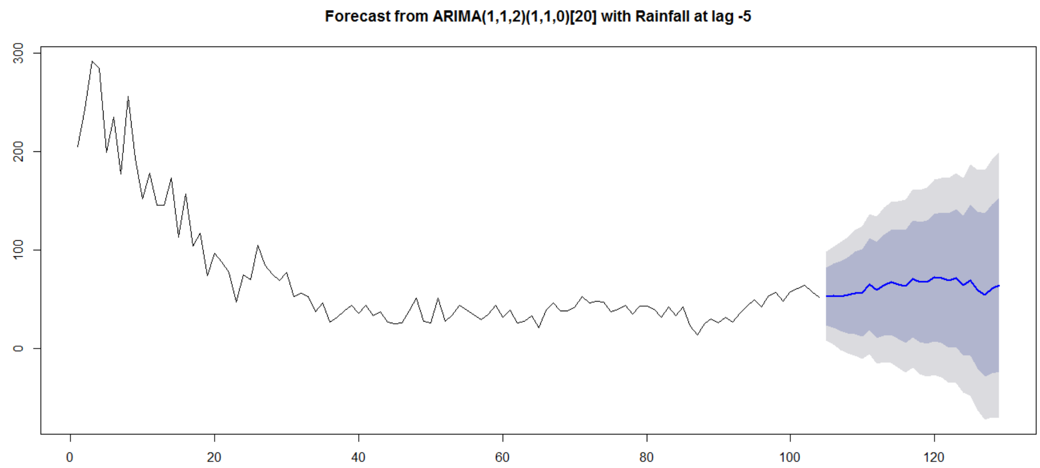


FIGURE 4.11. Forecasts from the ARIMA$(1,1,2)(1,1,0)_{20}$ model with rainfall variable of lag -5 applied to the dengue recorded case data.

4.1.8 Model comparison

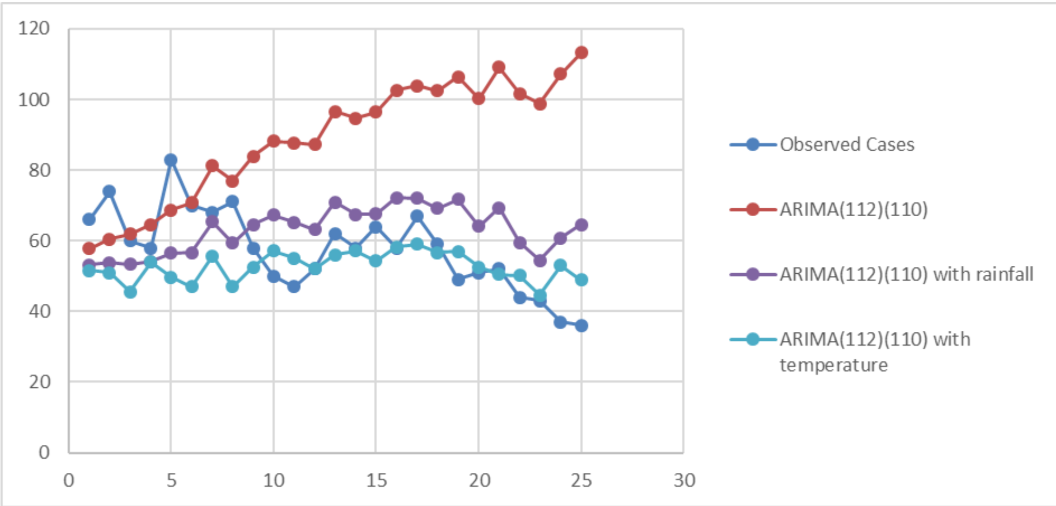


FIGURE 4.12. Model comparison between original dengue cases, forecasted cases using Univariate ARIMA and Multivariate ARIMA

4.1.9 Final Model

TABLE 4-3 Coefficient value of the final ARIMA model.

Model	AR1	MA1	SAR1	AIC	RMSE
ARIMA$(1,1,2)(1,1,0)_{20}$	-0.0703	-0.4521	-0.5353	770.18	20.43623
ARIMA$(1,1,2)(1,1,0)_{20}$ with Temperature	-0.9085	-0.4200	-0.4524	771.98	20.30672
ARIMA$(1,1,2)(1,1,0)_{20}$ with Rainfall	-0.0729	-0.4544	-0.5092	772.19	20.78052

The reliability of the forecasted value is tested through Root Mean Square Error which recorded on the last column of the table. Comparing the three model, ARIMA $(1,1,2)(1,1,0)_{20}$ with Temperature has the lowest error while ARIMA$(1,1,2)(1,1,0)_{20}$ with Rainfall recorded the highest error. Results shows that ARIMA model with rainfall variable fail to improve the model compared to ARIMA model with temperature variable which successfully improve the RMSE value of the model.

4.2 Gombak

4.2.1 Model identification step

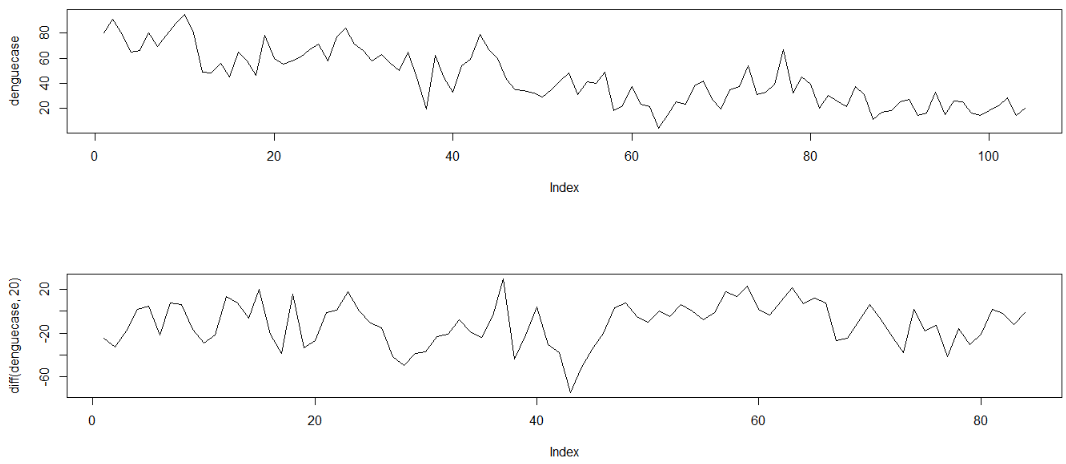


FIGURE 4.13. A. Plot of original dengue cases B. Plot of dengue cases after first seasonal difference

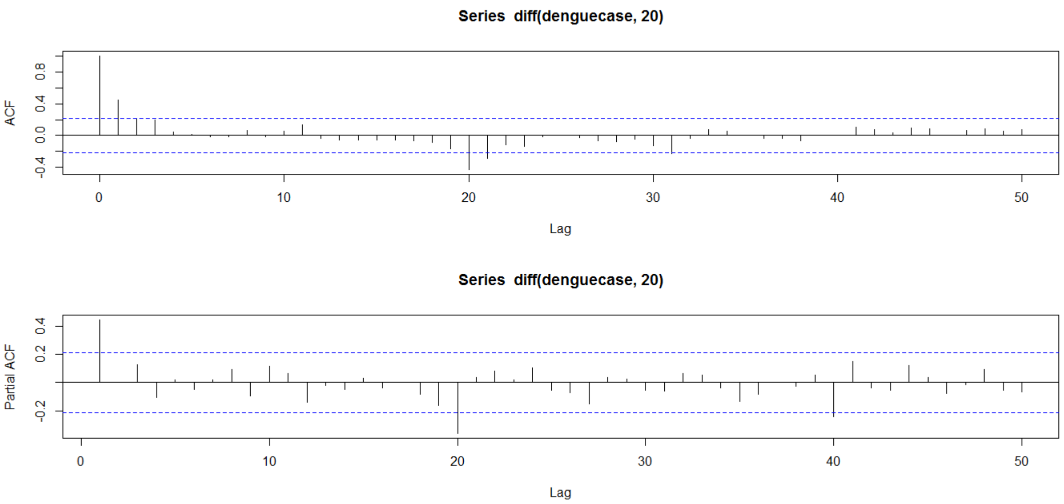


FIGURE 4.14. Plot of ACF and PACF of dengue cases after first seasonal difference

Based on plot of PACF, maximum p value is 1, P parameter is 2. Meanwhile ACF plot shows that maximum q value is 2 with Q parameter of 1 maximum. Hence numbers of ARIMA is identified and assessed.

4.2.2 Model Estimation Step

TABLE 4-4. AIC values for different ARIMA model

Models	AIC	Models	AIC
$ARIMA(0,0,1)(1,1,1)_{20}$	741.9	$ARIMA(1,0,1)(1,1,0)_{20}$	722.38
$ARIMA(0,0,2)(1,1,1)_{20}$	739.85	$ARIMA(1,0,2)(1,1,0)_{20}$	720.77
$ARIMA(1,0,0)(1,1,1)_{20}$	722.44	$ARIMA(0,0,1)(2,1,0)_{20}$	741.89
$ARIMA(1,0,1)(1,1,1)_{20}$	716.47	$ARIMA(0,0,2)(2,1,0)_{20}$	739.83
$ARIMA(1,0,2)(1,1,1)_{20}$	712.11	$ARIMA(1,0,0)(2,1,0)_{20}$	724.94
$ARIMA(0,0,1)(0,1,1)_{20}$	740.15	$ARIMA(1,0,1)(2,1,0)_{20}$	719.85
$ARIMA(0,0,2)(0,1,1)_{20}$	738.01	$ARIMA(1,0,2)(2,1,0)_{20}$	716.42
$ARIMA(1,0,0)(0,1,1)_{20}$	720.75	$ARIMA(0,0,1)(2,1,1)_{20}$	743.9
$ARIMA(1,0,1)(0,1,1)_{20}$	714.47	$ARIMA(0,0,2)(2,1,1)_{20}$	741.79
$ARIMA(1,0,2)(0,1,1)_{20}$	715.73	$ARIMA(1,0,0)(2,1,1)_{20}$	724.38
$ARIMA(0,0,1)(1,1,0)_{20}$	739.9	$ARIMA(1,0,1)(2,1,1)_{20}$	718.1
$ARIMA(0,0,2)(1,1,0)_{20}$	738	$ARIMA(1,0,2)(2,1,1)_{20}$	713.47
$ARIMA(1,0,0)(1,1,0)_{20}$	725.94		

*$ARIMA(1,0,2)(1,1,1)_{20}$ is selected for having lowest AIC

4.2.3 Model Validation Step

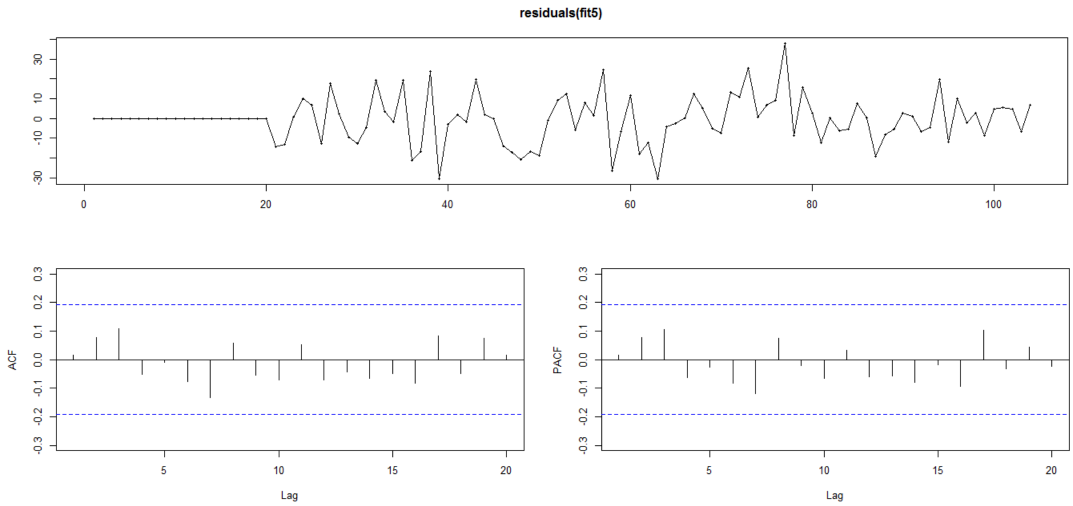


FIGURE 4.15. Residuals from the $ARIMA(1,0,2)(1,1,1)_{20}$ model

ACF of residual after applying model is within significance limits shows that no other information can be extracted and it is equal to white noise

4.2.4 Influence of Climate Variable

a. Temperature

Maximum lag of correlation between temperature with dengue cases recorded is at lag -14. The dengue cases recorded is higher after 14 weeks of higher temperature.

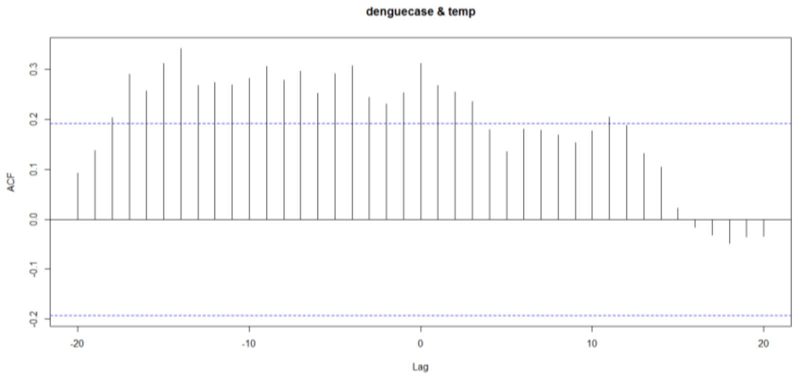


FIGURE 4.16. Cross-correlation between dengue cases recorded and mean temperature in Gombak

b. Humidity

No positive correlation between humidity with dengue cases recorded.

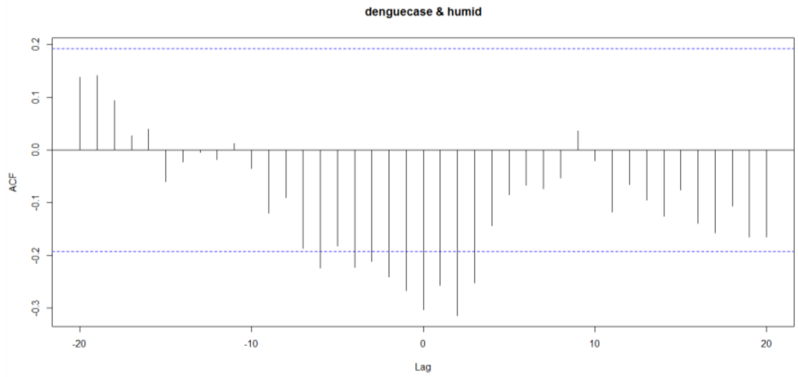


FIGURE 4.17. Cross-correlation between dengue cases recorded and humidity in Gombak

c. Rainfall

No correlation between rainfall and dengue cases recorded

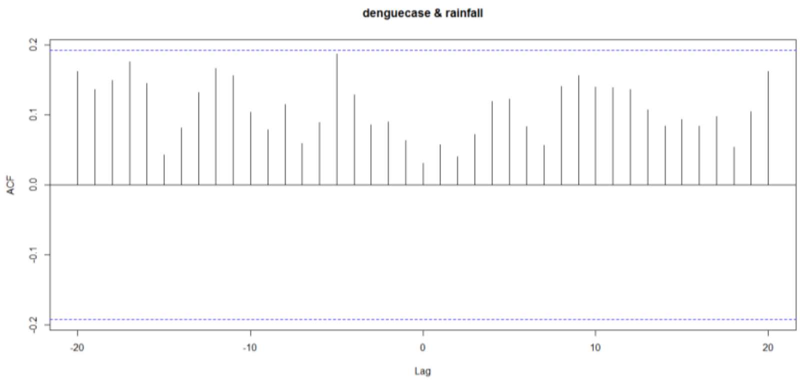


FIGURE 4.18. Cross-correlation between dengue cases recorded and rainfall in Gombak

4.2.5 Model Forecasting Step (Univariate ARIMA)

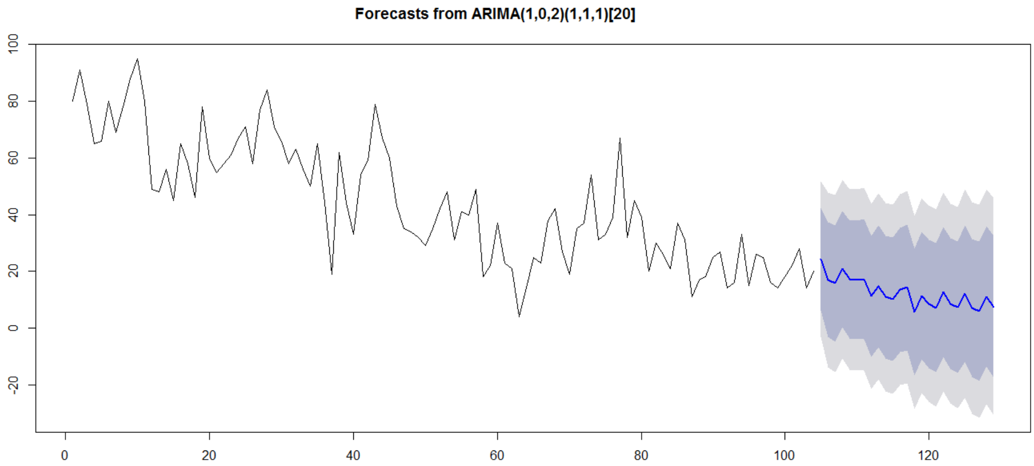


FIGURE 4.19. Forecasts from the ARIMA$(1,0,2)(1,1,1)_{20}$ model applied to the dengue recorded case data.

4.2.6 Model Forecasting Step (Multivariate ARIMA)

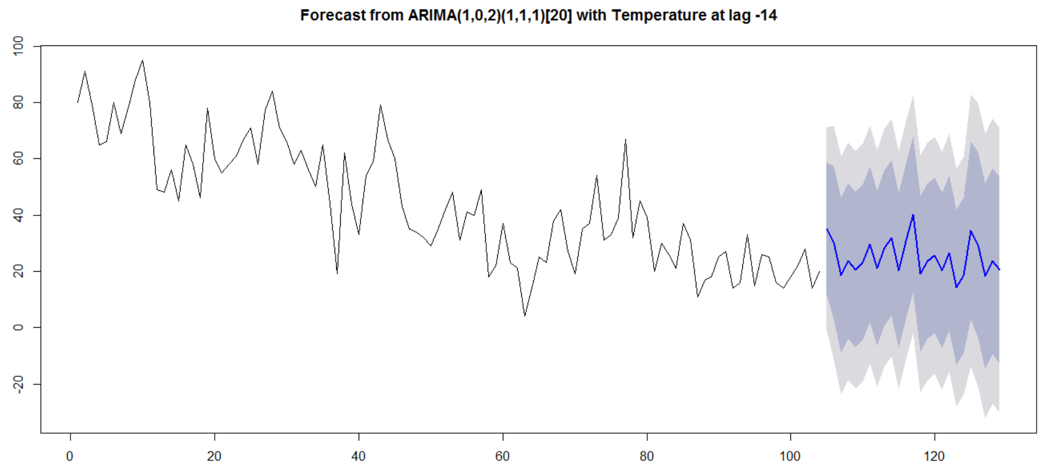


FIGURE 4.20. Forecasts from the ARIMA$(1,0,2)(1,1,1)_{20}$ model with temperature variable of lag-14 applied to the dengue recorded case data.

4.2.7 Model comparison

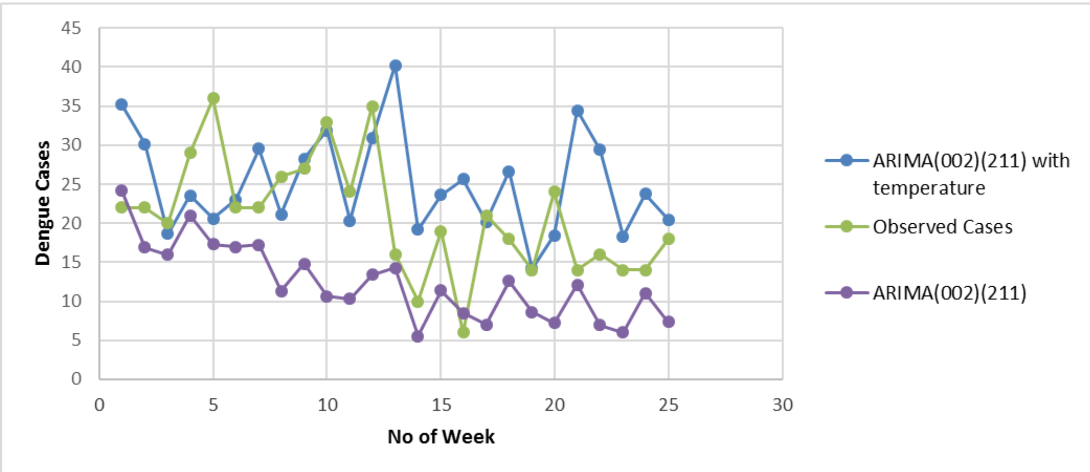


FIGURE 4.21. Model comparison between original dengue cases, forecasted cases using Univariate ARIMA and Multivariate ARIMA

4.3 Klang

4.3.1 Model Identification Step

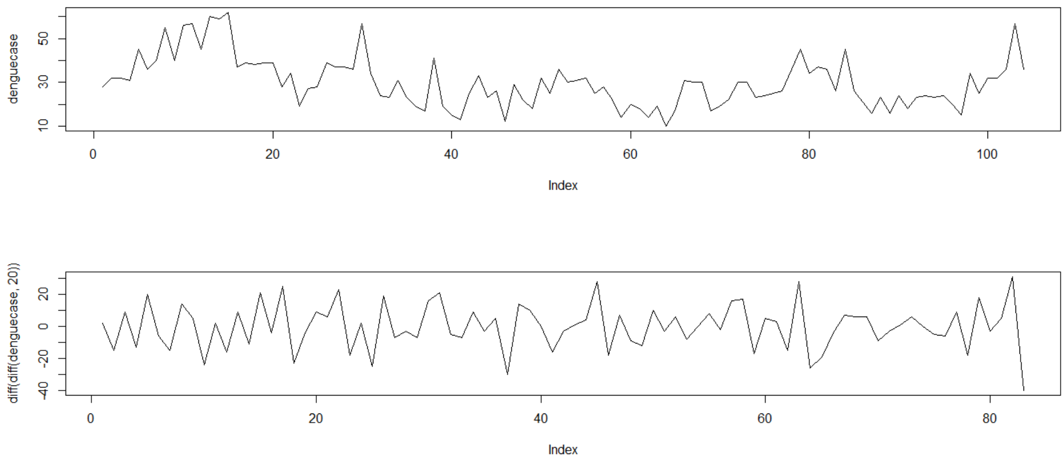


FIGURE 4.22. A. Plot of original dengue cases B. Plot of dengue cases after first seasonal difference

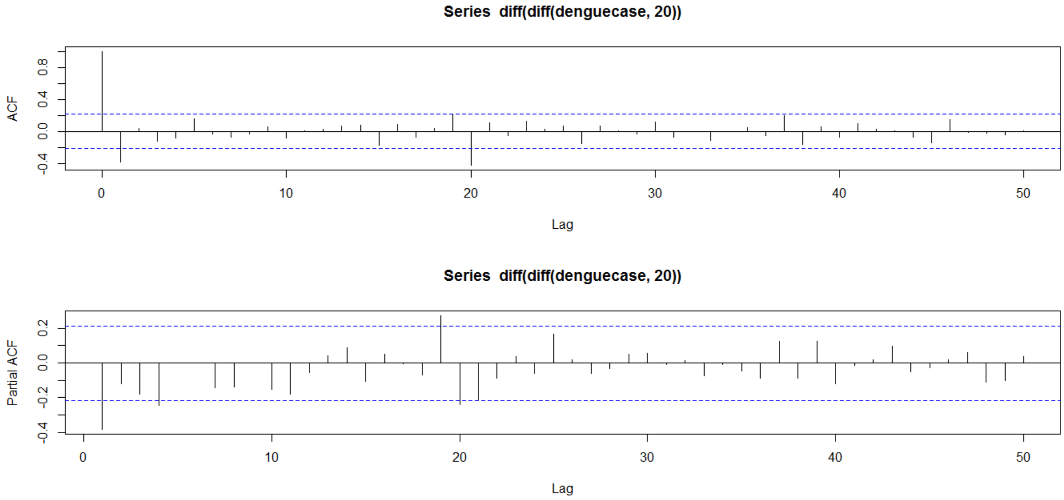


FIGURE 4.23. Plot of ACF and PACF of dengue cases after first non-seasonal and seasonal difference

Based on plot of PACF, maximum p value is 2, P parameter is 1. Meanwhile ACF plot shows that maximum q value is 2 with Q parameter of 1 maximum. Hence numbers of ARIMA is identified and assessed.

4.3.2 Model Estimation Step

TABLE 4-5. AIC values for different ARIMA model

Models	AIC	Models	AIC
$ARIMA(0,1,1)(1,1,1)_{20}$	631.53	$ARIMA(1,1,2)(0,1,1)_{20}$	639.85
$ARIMA(0,1,2)(1,1,1)_{20}$	633.53	$ARIMA(2,1,0)(0,1,1)_{20}$	634.6
$ARIMA(1,1,0)(1,1,1)_{20}$	634.2	$ARIMA(2,1,1)(0,1,1)_{20}$	634.14
$ARIMA(1,1,1)(1,1,1)_{20}$	632.87	$ARIMA(2,1,2)(0,1,1)_{20}$	636.13
$ARIMA(1,1,2)(1,1,1)_{20}$	635.52	$ARIMA(0,1,1)(1,1,0)_{20}$	635.53
$ARIMA(2,1,0)(1,1,1)_{20}$	634.86	$ARIMA(0,1,2)(1,1,0)_{20}$	637.5
$ARIMA(2,1,1)(1,1,1)_{20}$	634.68	$ARIMA(1,1,0)(1,1,0)_{20}$	639.85
$ARIMA(2,1,2)(1,1,1)_{20}$	637.33	$ARIMA(1,1,1)(1,1,0)_{20}$	716.42
$ARIMA(0,1,1)(0,1,1)_{20}$	631.55	$ARIMA(1,1,2)(1,1,0)_{20}$	638.26
$ARIMA(0,1,2)(0,1,1)_{20}$	633.55	$ARIMA(2,1,0)(1,1,0)_{20}$	640.42
$ARIMA(1,1,0)(0,1,1)_{20}$	633.62	$ARIMA(2,1,1)(1,1,0)_{20}$	638.13
$ARIMA(1,1,1)(0,1,1)_{20}$	633.55	$ARIMA(2,1,2)(1,1,0)_{20}$	640.09

*$ARIMA(0,1,1)(1,1,1)_{20}$ is selected for having lowest AIC

4.3.3 Model Validation Step

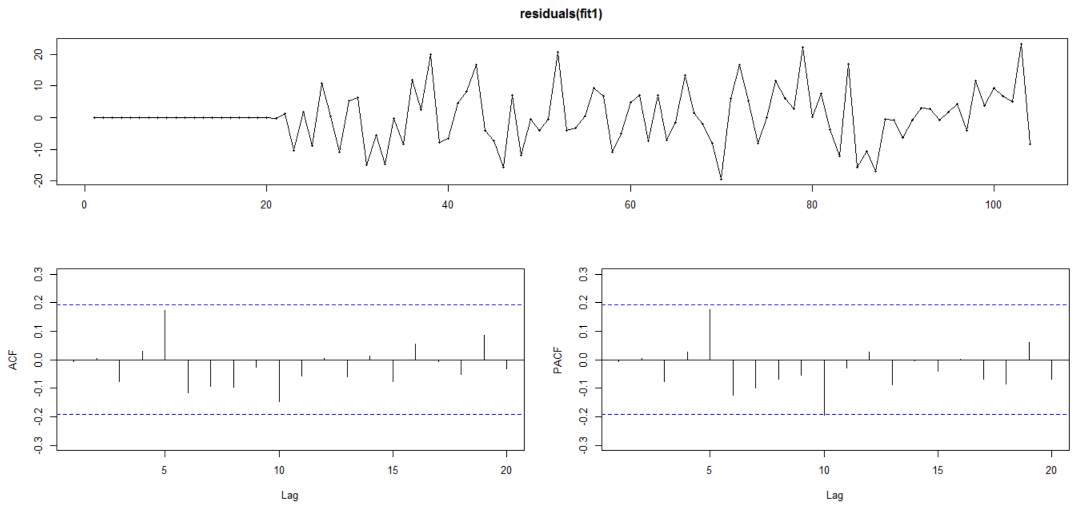


FIGURE 4.24. Residuals from the $ARIMA(0,1,1)(1,1,1)_{20}$ model

ACF of residual after applying model is within significance limits shows that no other information can be extracted and it is equal to white noise

4.3.4 Influence of climate variable

a. Temperature

Maximum lag of correlation between temperature with dengue cases recorded is at lag -5. The dengue cases recorded is higher after 5 weeks of higher temperature.

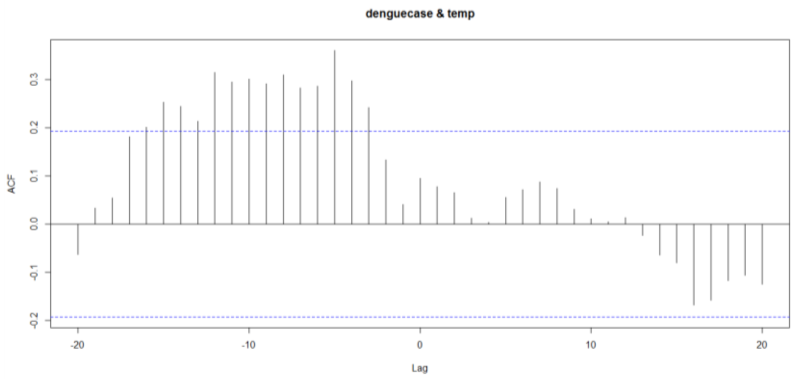


FIGURE 4.25. Cross-correlation between dengue cases recorded and mean temperature in Klang

b. Humidity

Maximum lag of correlation between humidity with dengue cases recorded is at lag -20. The dengue cases recorded is higher after 20 weeks of higher humidity.

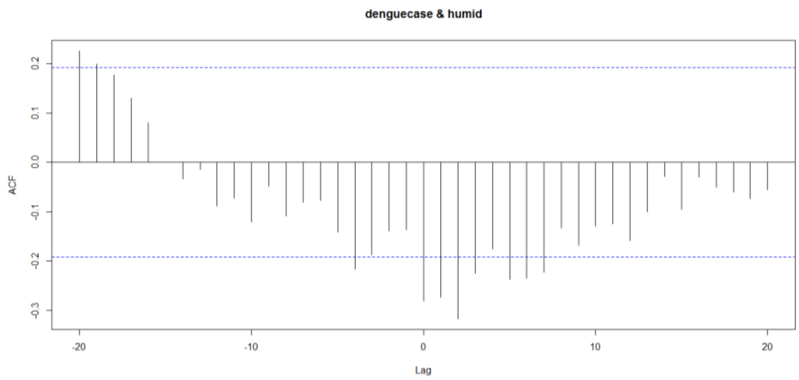


FIGURE 4.26. Cross-correlation between dengue cases recorded and humidity in Klang

c. Rainfall

Maximum lag of correlation between rainfall with dengue cases recorded is at lag +8. The dengue cases recorded is higher before 8 weeks of higher rainfall rate.

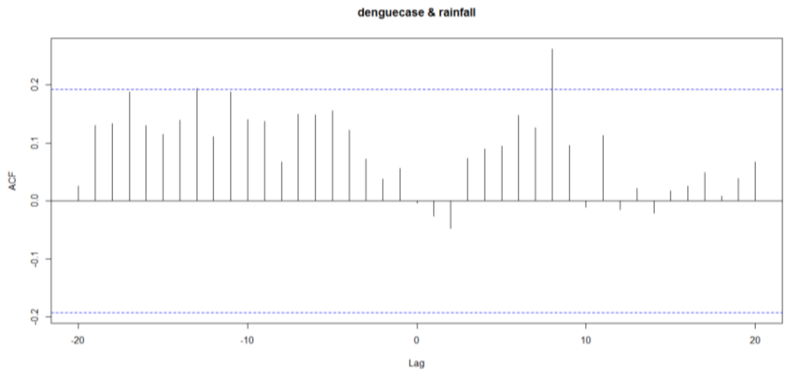


FIGURE 4.27. Cross-correlation between dengue cases recorded and rainfall in Klang

4.3.5 Model Forecasting Step (Univariate)

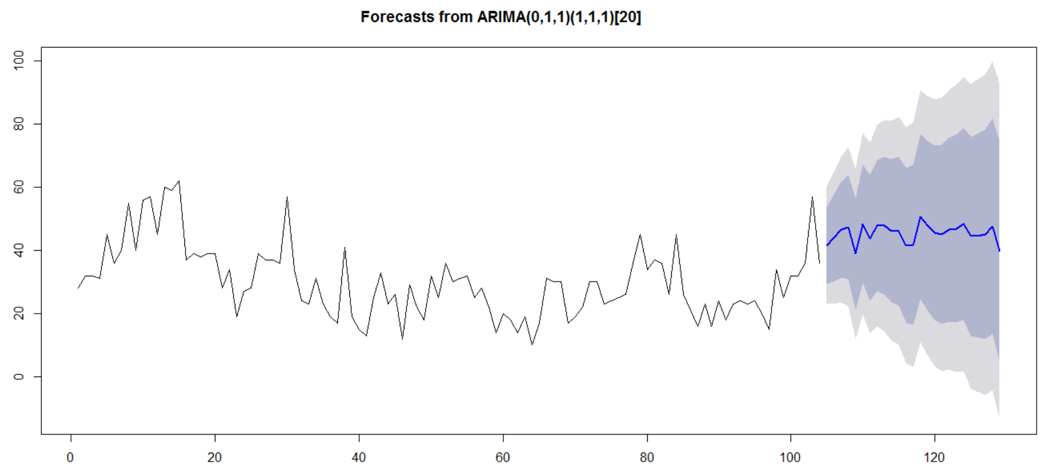


FIGURE 4.28. Forecasts from the ARIMA$(0,1,1)(1,1,1)_{20}$ model applied to the dengue recorded case data.

4.3.6 Model Forecasting Step (Multivariate with Temperature)

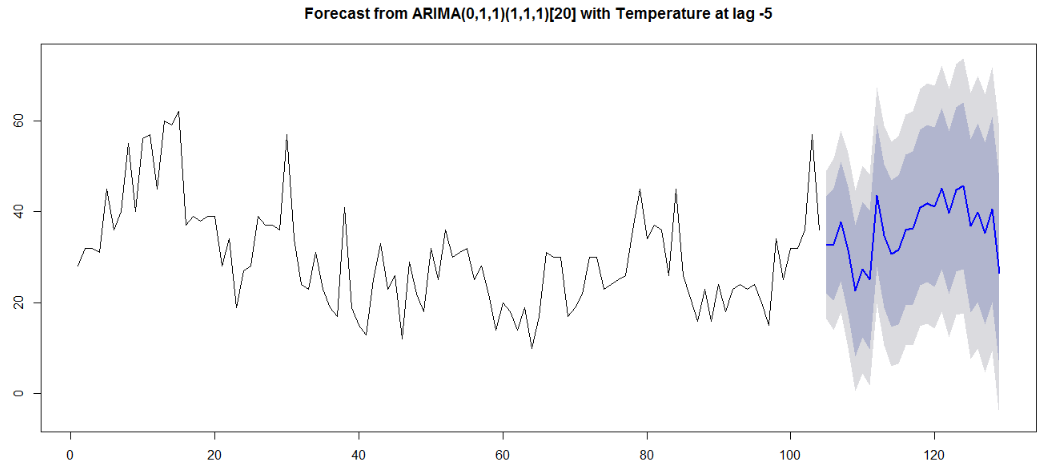


FIGURE 4.29. Forecasts from the ARIMA$(0,1,1)(1,1,1)_{20}$ model with temperature variable of lag -5 applied to the dengue recorded case data.

4.3.7 Model Forecasting Step (Multivariate with Humidity)

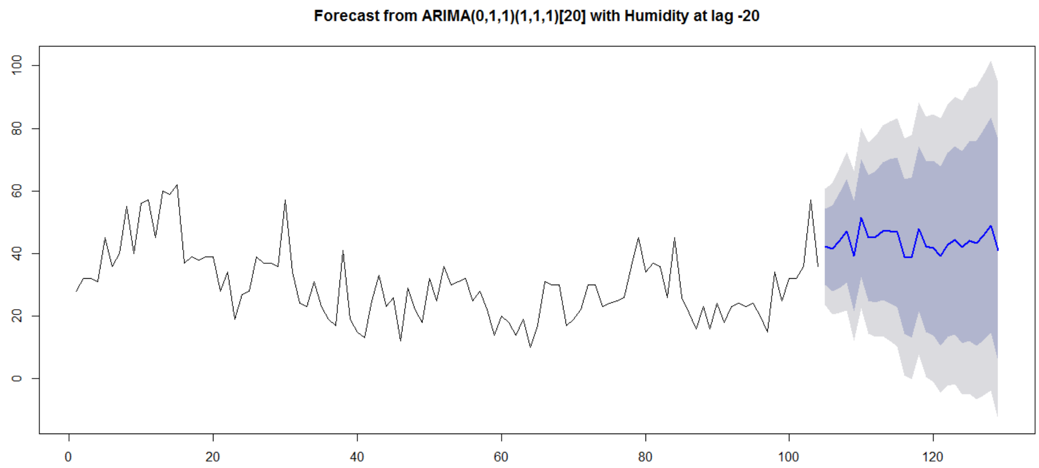


FIGURE 4.30. Forecasts from the ARIMA$(0,1,1)(1,1,1)_{20}$ model with humidity variable of lag -20 applied to the dengue recorded case data.

4.3.8 Model Forecasting Step (Multivariate with Rainfall)

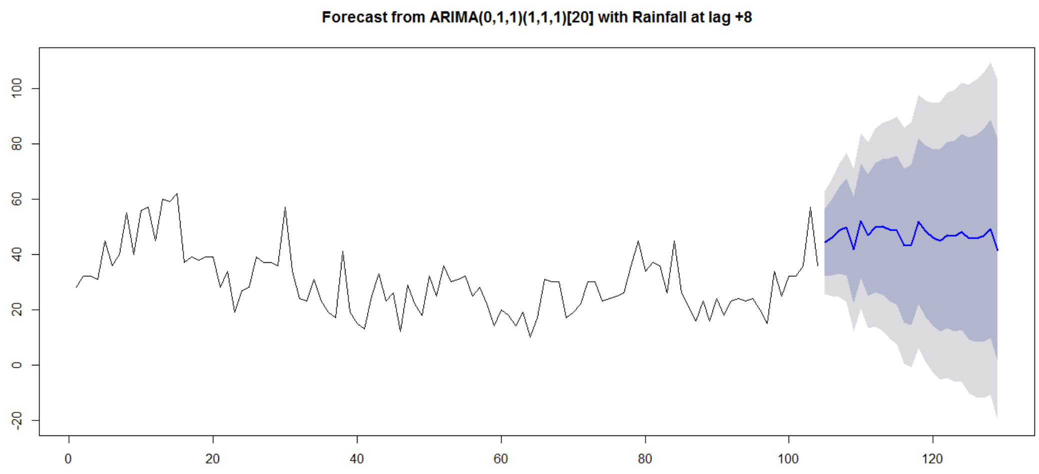


FIGURE 4.31. Forecasts from the ARIMA$(0,1,1)(1,1,1)_{20}$ model with rainfall variable of lag +8 applied to the dengue recorded case data.

4.3.9 Model comparison

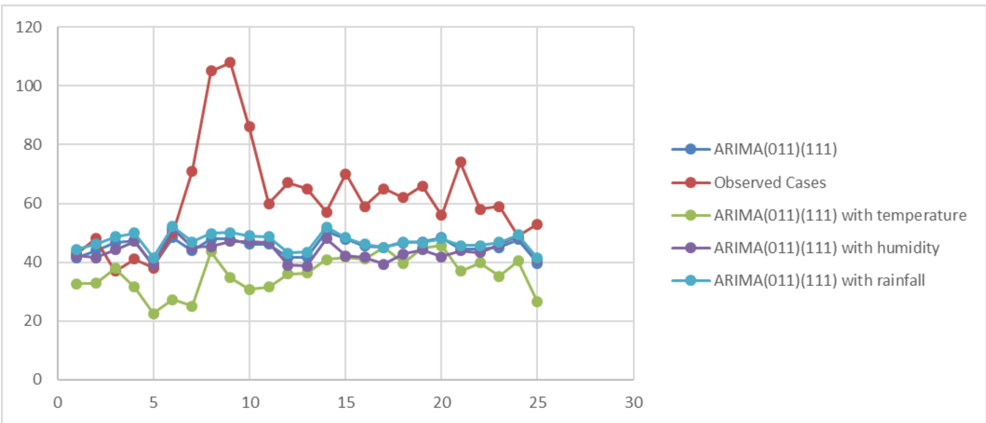


FIGURE 4.32. Model comparison between original dengue cases, forecasted cases using Univariate ARIMA and Multivariate ARIMA

The pattern of the original observed cases compared to forecasted ARIMA shows obvious different. This means that the observed cases has other external factor than climate as all climate variable cannot predict the highest number of dengue cases at week 8 and 9 except temperature variable which shows some increase of cases at that week. The result can be concluded that temperature might give proper forecasted value except the value is lower than the actual value.

4.3.10 Final Model

TABLE 4-6 Coefficient value of the final ARIMA model.

Model	MA1	SAR1	SMA1	AIC	RMSE
ARIMA$(0,1,1)(1,1,1)_{20}$	-0.4728	-0.2968	-0.5481	631.53	8.439636
ARIMA$(0,1,1)(1,1,1)_{20}$ with Temperature	0.1104	-0.5103	-0.2849	635	6.411847
ARIMA$(0,1,1)(1,1,1)_{20}$ with Humidity	-0.4611	-0.4873	-0.9995	631.55	7.677376
ARIMA$(0,1,1)(1,1,1)_{20}$ with Rainfall	-0.4387	-0.2596	-0.6358	634.86	8.376592

ARIMA $(0,1,1)(1,1,1)_{20}$ with Temperature has the lowest error while univariate ARIMA recorded the highest error. Results shows that ARIMA model with external variable has successfully improve the model shown as the lower value of RMSE for each external variable.

4.4 Kuala Selangor

4.4.1 Model identification step

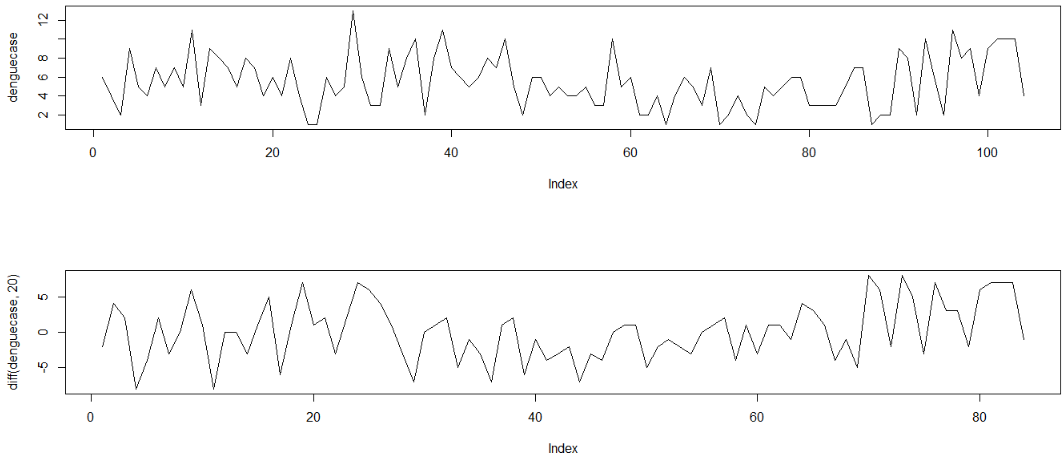


FIGURE 4.33. A. Plot of original dengue cases B. Plot of dengue cases after first seasonal difference

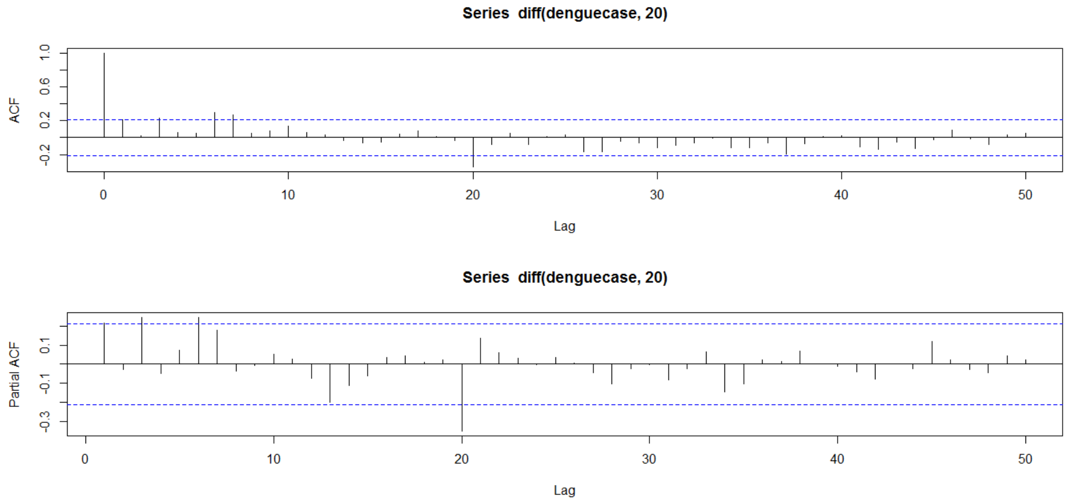


FIGURE 4.34. Plot of ACF and PACF of dengue cases after first seasonal difference

Based on plot of PACF, maximum p value is 1, P parameter is 1. Meanwhile ACF plot shows that maximum q value is 2 with Q parameter of 1 maximum. Hence numbers of ARIMA is identified and assessed.

4.4.2 Model Estimation Step

TABLE 4-7. AIC values for different ARIMA model

Models	AIC	Models	AIC
ARIMA(0,0,1)(1,1,1)$_{20}$	447.11	ARIMA(1,0,1)(0,1,1)$_{20}$	447.41
ARIMA(0,0,2)(1,1,1)$_{20}$	449	ARIMA(1,0,2)(0,1,1)$_{20}$	448.85
ARIMA(1,0,0)(1,1,1)$_{20}$	447.24	ARIMA(0,0,1)(1,1,0)$_{20}$	452.67
ARIMA(1,0,1)(1,1,1)$_{20}$	444.88	ARIMA(0,0,2)(1,1,0)$_{20}$	454.44
ARIMA(1,0,2)(1,1,1)$_{20}$	446.7	ARIMA(1,0,0)(1,1,0)$_{20}$	452.79
ARIMA(0,0,1)(0,1,1)$_{20}$	446.25	ARIMA(1,0,1)(1,1,0)$_{20}$	448.59
ARIMA(0,0,2)(0,1,1)$_{20}$	448.01	ARIMA(1,0,2)(1,1,0)$_{20}$	450.58
ARIMA(1,0,0)(0,1,1)$_{20}$	446.53		

*ARIMA(1,0,1)(1,1,1)$_{20}$ is selected for having lowest AIC

4.4.3 Model Validation Step

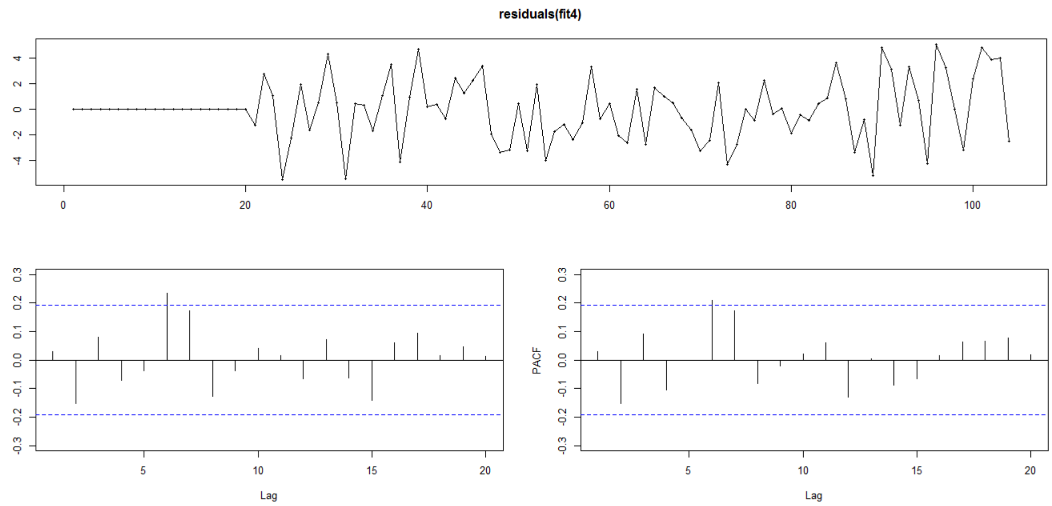


FIGURE 4.35. Residuals from the ARIMA(1,0,1)(1,1,1)$_{20}$ model.

ACF of residual after applying model is within significance limits shows that no other information can be extracted and it is equal to white noise

4.4.4 Influence of Climate Variable

a. Temperature

Maximum lag of correlation between temperature with dengue cases recorded is at lag +9. The dengue cases recorded is higher before 9 weeks of higher temperature.

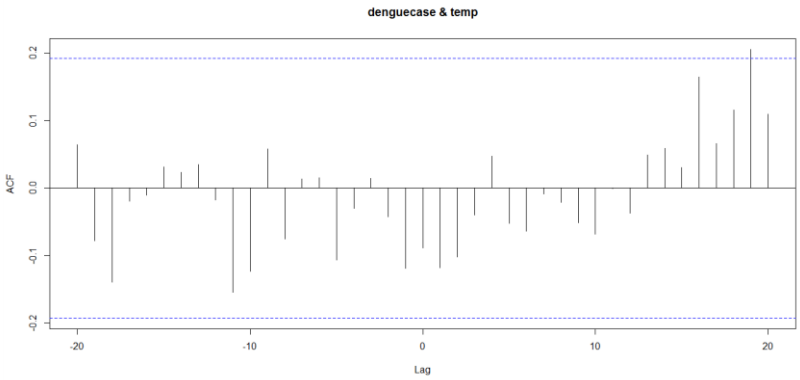


FIGURE 4.36. Cross-correlation between dengue cases recorded and mean temperature in Kuala Selangor

b. Humidity

Maximum lag of correlation between humidity with dengue cases recorded is at lag -5. The dengue cases recorded is higher after 5 weeks of higher humidity.

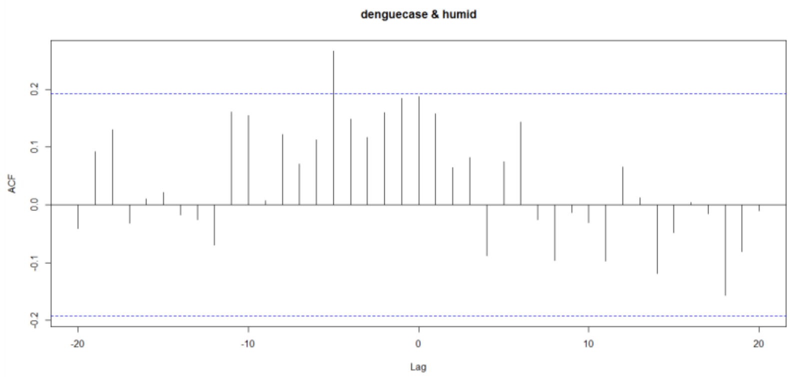


FIGURE 4.37. Cross-correlation between dengue cases recorded and humidity in Kuala Selangor

c. Rainfall

Maximum lag of correlation between rainfall with dengue cases recorded is at lag +6. The dengue cases recorded is higher before 6 weeks of higher rainfall rate. However, the correlation can be considered as no correlation at all as the correlation is not obvious.

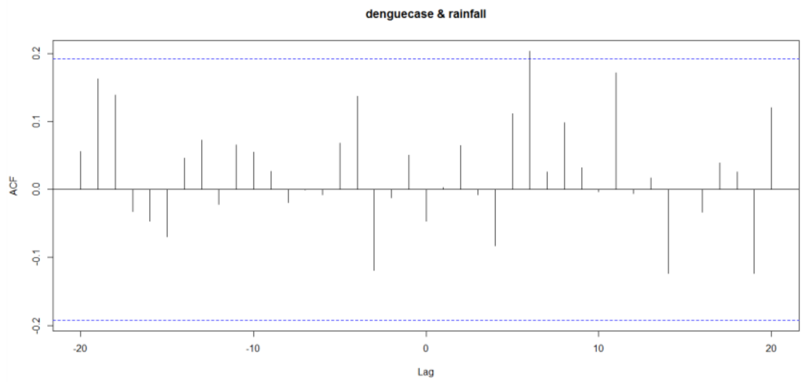


FIGURE 4.38. Cross-correlation between dengue cases recorded and rainfall in Kuala Selangor

4.4.5 Model Forecasting Step (Univariate)

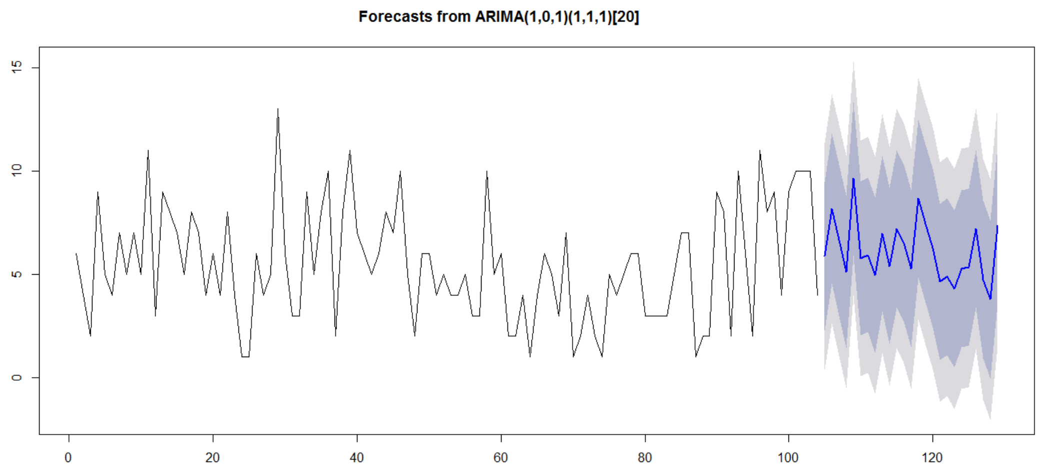


FIGURE 4.39. Forecasts from the ARIMA$(1,0,1)(1,1,1)_{20}$ model applied to the dengue recorded case data

4.4.6 Model Forecasting Step (Multivariate with Temperature)

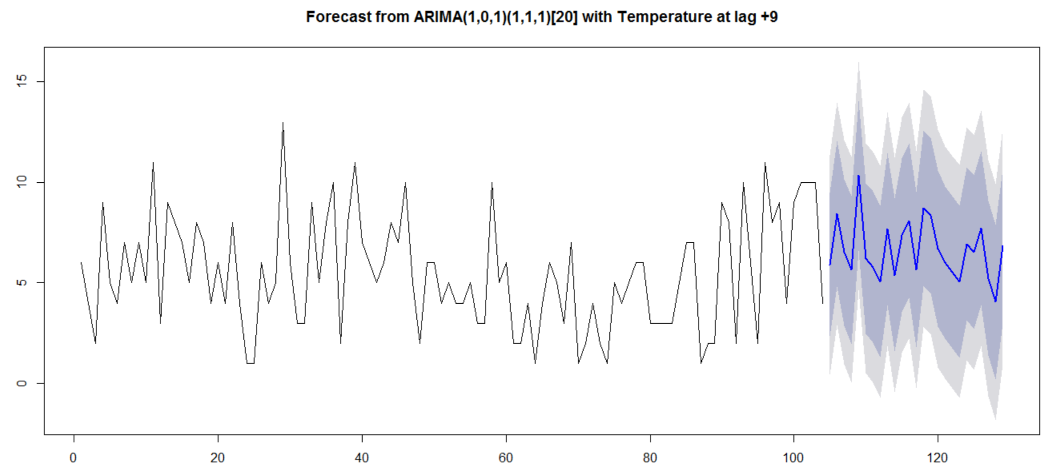


FIGURE 4.40. Forecasts from the ARIMA$(1,0,1)(1,1,1)_{20}$ model with temperature variable of lag +9 applied to the dengue recorded case data.

4.4.7 Model Forecasting Step (Multivariate with Humidity)

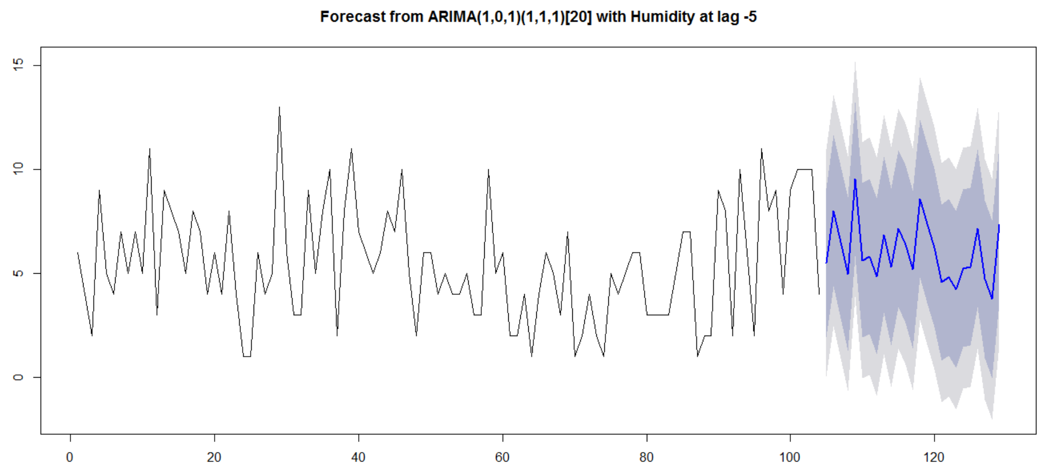


FIGURE 4.41. Forecasts from the ARIMA$(1,0,1)(1,1,1)_{20}$ model with humidity variable of lag -5 applied to the dengue recorded case data.

4.4.8 Model comparison

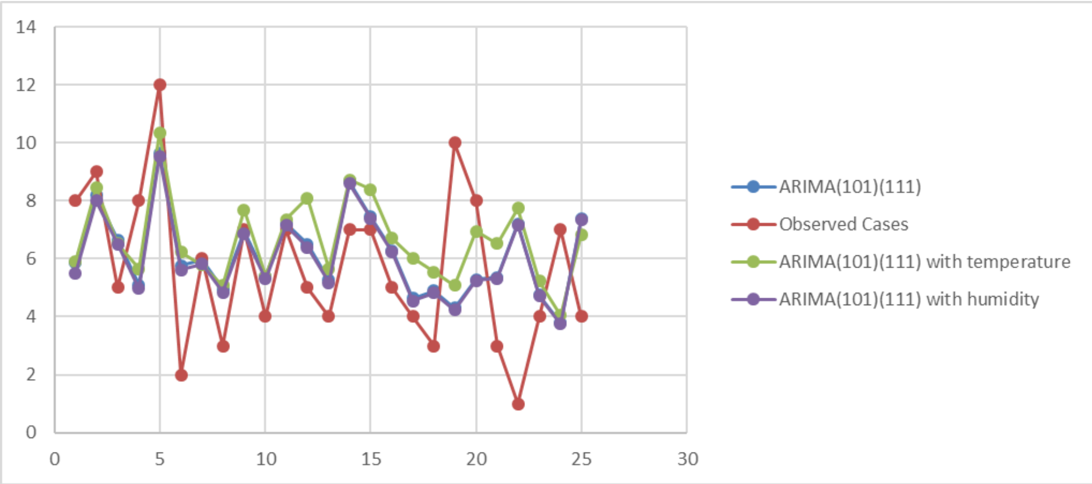


FIGURE 4.42. Model comparison between original dengue cases, forecasted cases using Univariate ARIMA and Multivariate ARIMA

Comparing the original observed cases with the predicted value, the ARIMA has closely accurate at the beginning of the weeks, from week 1 to 17, however the predicted value cannot evaluate the increasing in cases at week 19 and 24. The increase in dengue cases might be influence from other factor than climate since both temperature and humidity variable cannot evaluate the situation.

4.4.9 Final Model

TABLE 4-8 Coefficient value of the final ARIMA model.

Model	AR1	MA1	SAR1	SMA1	AIC	RMSE
ARIMA(1,0,1)(1,1,1)$_{20}$	0.9518	-0.818	-0.215	-0.947	444.88	2.342431
ARIMA(1,0,1)(1,1,1)$_{20}$ with Temperature	0.9549	-0.799	-0.991	-0.060	446.39	2.213418
ARIMA(1,0,1)(1,1,1)$_{20}$ with Humidity	0.9537	-0.768	-0.218	-0.973	446.7	2.310442

ARIMA (1,0,1)(1,1,1)$_{20}$ with external variable recoded lower error compared to univariate ARIMA model. Results shows that ARIMA model with external variable has successfully improve the model shown as the lower value of RMSE for both external variables.

4.5 Hulu Langat

4.5.1 Model identification step

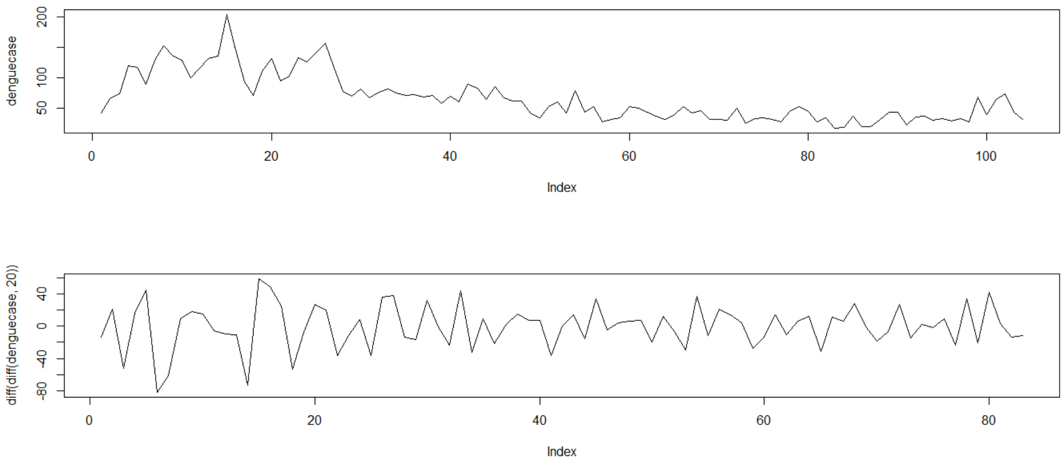


FIGURE 4.43. A. Plot of original dengue cases B. Plot of dengue cases after first non-seasonal and first seasonal difference

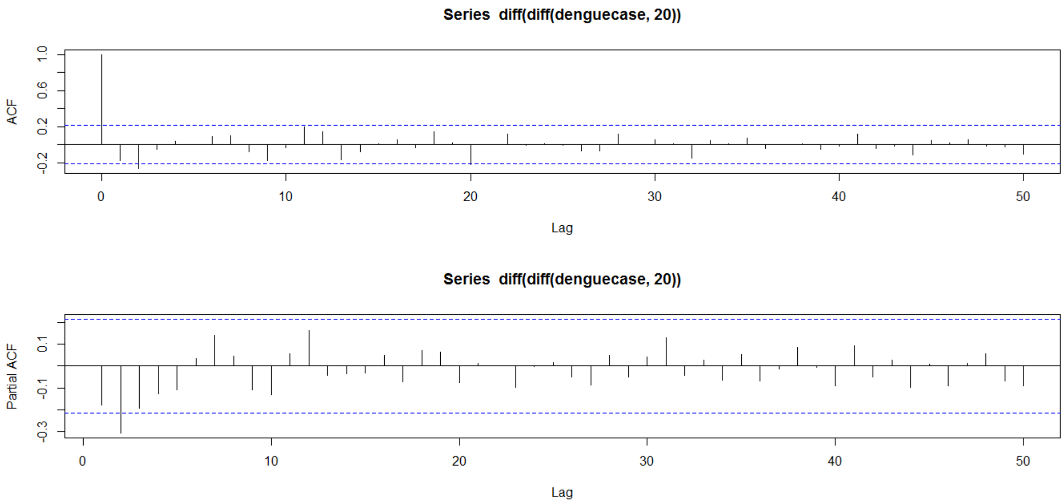


FIGURE 4.44. Plot of ACF and PACF of dengue cases after first non-seasonal and seasonal difference

Based on plot of PACF, maximum p value is 0, P parameter is 1. Meanwhile ACF plot shows that maximum q value is 2 with Q parameter of 2 maximum. Hence numbers of ARIMA is identified and assessed.

4.5.2 Model Estimation Step

TABLE 4-9. AIC values for different ARIMA model

Models	AIC	Models	AIC
$ARIMA(0,1,1)(1,1,1)_{20}$	772.68	$ARIMA(1,1,1)(0,1,1)_{20}$	768.18
$ARIMA(0,1,2)(1,1,1)_{20}$	769.14	$ARIMA(1,1,2)(0,1,1)_{20}$	769.26
$ARIMA(1,1,0)(1,1,1)_{20}$	776.15	$ARIMA(0,1,1)(1,1,0)_{20}$	772.11
$ARIMA(1,1,1)(1,1,1)_{20}$	770.14	$ARIMA(0,1,2)(1,1,0)_{20}$	769.56
$ARIMA(1,1,2)(1,1,1)_{20}$	771.29	$ARIMA(1,1,0)(1,1,0)_{20}$	776.6
$ARIMA(0,1,1)(0,1,1)_{20}$	770.71	$ARIMA(1,1,1)(1,1,0)_{20}$	770.42
$ARIMA(0,1,2)(0,1,1)_{20}$	767.37	$ARIMA(1,1,2)(1,1,0)_{20}$	778.87
$ARIMA(1,1,0)(0,1,1)_{20}$	774.15		

*$ARIMA(0,1,2)(0,1,1)_{20}$ is selected for having lowest AIC

4.5.3 Model Validation Step

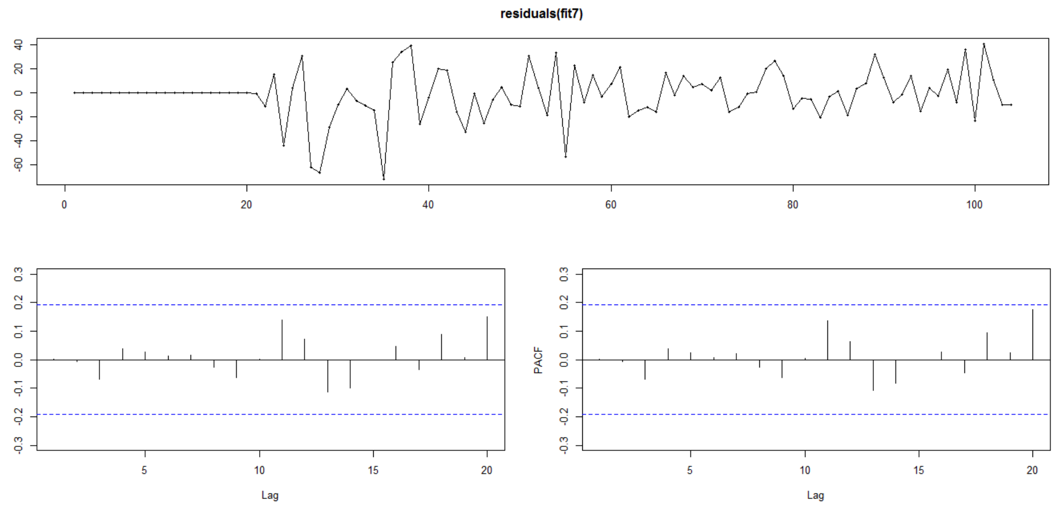


FIGURE 4.45. Residuals from the $ARIMA(0,1,2)(0,1,1)_{20}$ model.

ACF of residual after applying model is within significance limits shows that no other information can be extracted and it is equal to white noise

4.5.4 Influence of climate variable

a. Temperature

Maximum lag of correlation between temperature with dengue cases recorded is at lag -5. The dengue cases recorded is higher after 5 weeks of higher temperature.

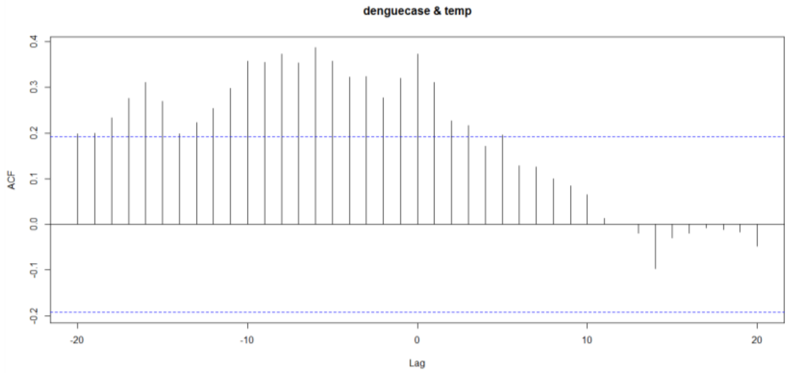


FIGURE 4.46. Cross-correlation between dengue cases recorded and mean temperature in Hulu Langat

b. Humidity

No positive correlation between humidity with dengue cases recorded.

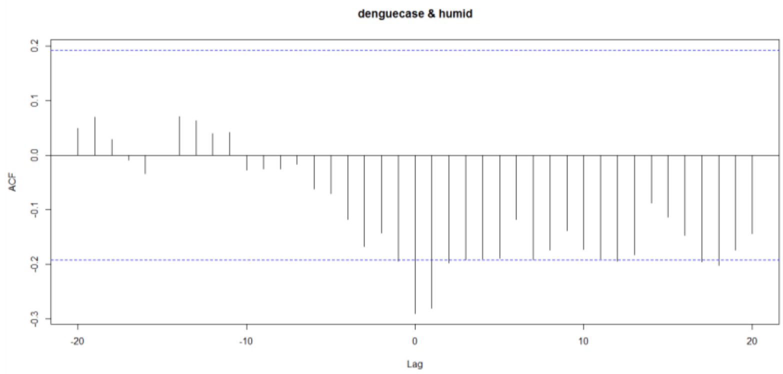


FIGURE 4.47. Cross-correlation between dengue cases recorded and humidity in Hulu Langat

c. Rainfall

Maximum lag of correlation between rainfall with dengue cases recorded is at lag +6. The dengue cases recorded is higher before 6 weeks of higher rainfall rate.

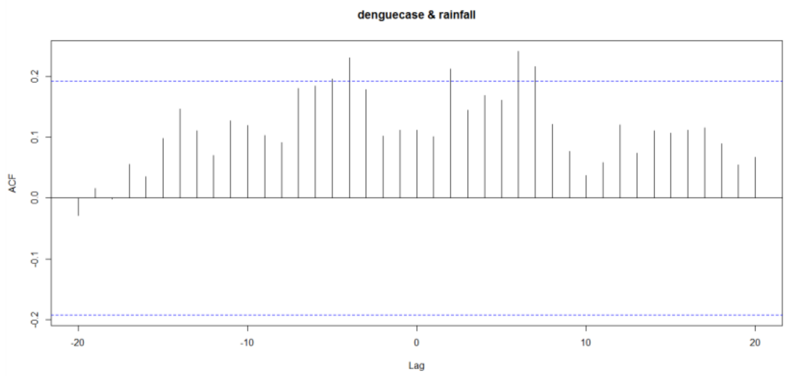


FIGURE 4.48. Cross-correlation between dengue cases recorded and rainfall in Hulu Langat

4.5.5 Model Forecasting Step (Univariate)

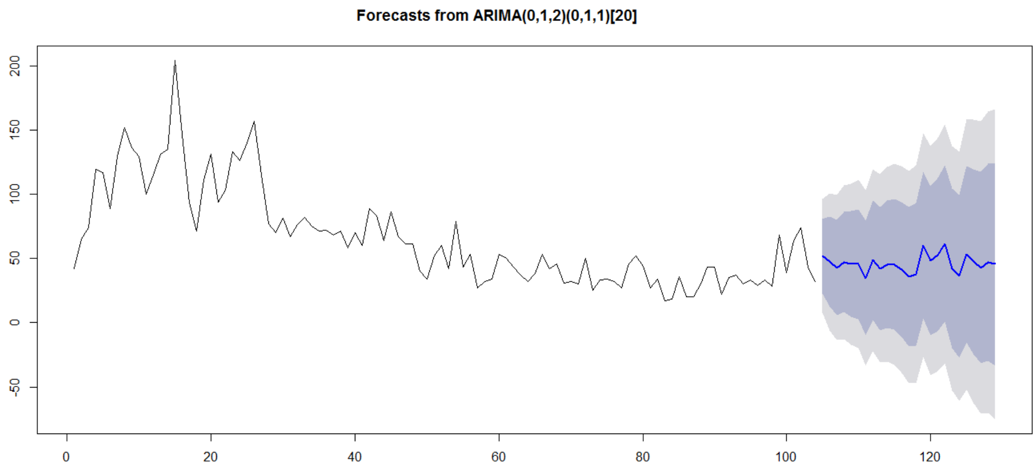


FIGURE 4.49. Forecasts from the ARIMA$(0,1,2)(0,1,1)_{20}$ model applied to the dengue recorded case data

4.5.6 Model Forecasting Step (Multivariate with Temperature)

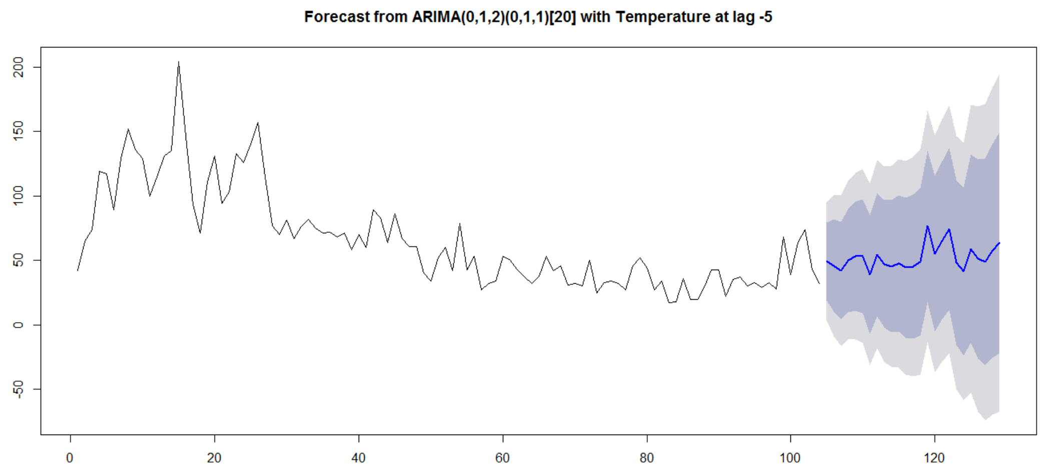


FIGURE 4.50. Forecasts from the ARIMA$(0,1,2)(0,1,1)_{20}$ model with temperature variable of lag -5 applied to the dengue recorded case data.

4.5.7 Model Forecasting Step (Multivariate with Rainfall)

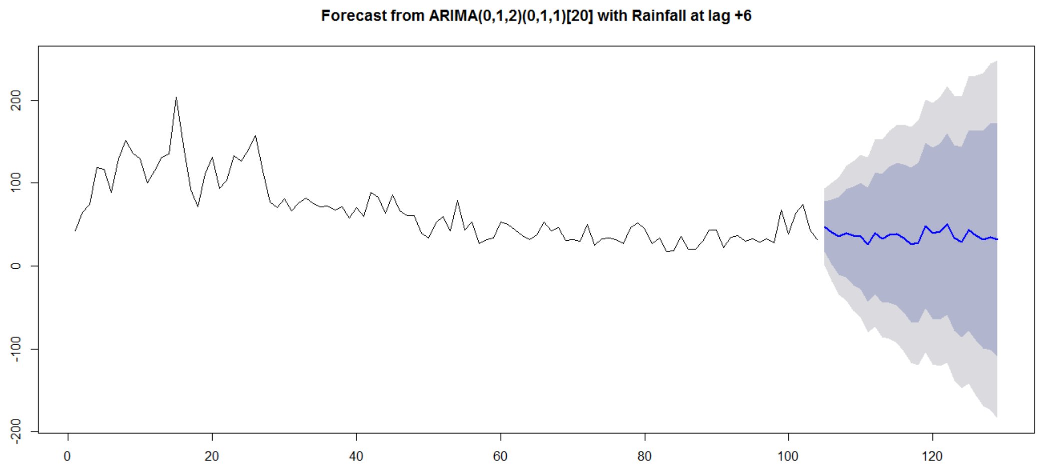


FIGURE 4.51. Forecasts from the ARIMA$(0,1,2)(0,1,1)_{20}$ model with rainfall variable of lag +6 applied to the dengue recorded case data.

4.5.8 Model comparison

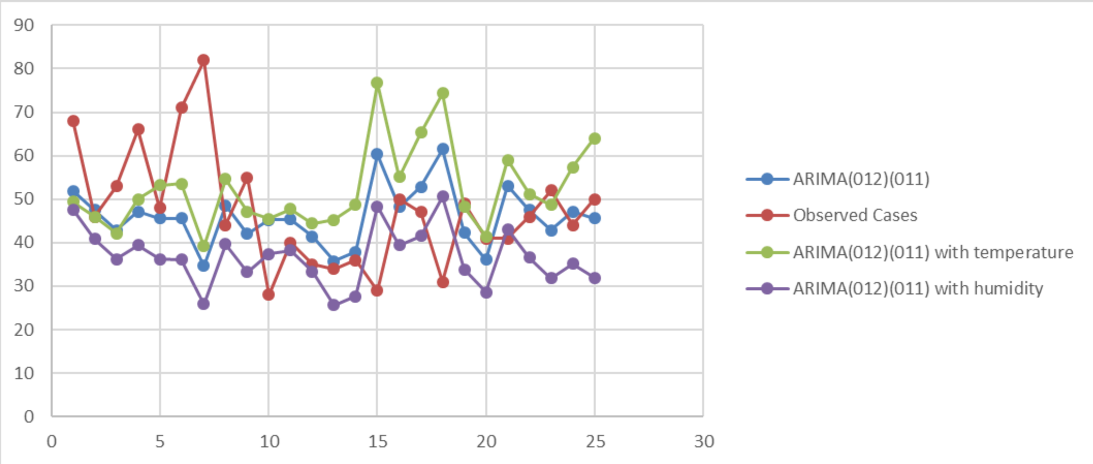


FIGURE 4.52. Model comparison between original dengue cases, forecasted cases using Univariate ARIMA and Multivariate ARIMA

Based on Figure 4.52, the model of univariate and multivariate having not much difference. Comparing forecasted value with original value, they are having huge difference value in term of outbreak. For example, at week 7, the original data shows sudden increase, however the forecasted value for both univariate and multivariate shows the opposite.

4.5.9 Final Model

TABLE 4-10 Coefficient value of the final ARIMA model.

Model	MA1	MA2	SMA1	AIC	RMSE
$ARIMA(0,1,2)(0,1,1)_{20}$	-0.328	-0.242	-0.510	767.37	20.1770
$ARIMA(0,1,2)(0,1,1)_{20}$ with Temperature	-0.344	-0.225	-0.352	769.56	20.8651
$ARIMA(0,1,2)(0,1,1)_{20}$ with Rainfall	-0.051	-0.186	-0.576	774.15	21.0581

$ARIMA(0,1,2)(0,1,1)_{20}$ with external variable recoded lowest error. Results shows that ARIMA model with external variable has failed to improve the model shown as the higher value of RMSE for both external variables. Even so, the original univariate ARIMA model is also considered fail to predict accurately.

4.6 Hulu Selangor

4.6.1 Model Identification Step

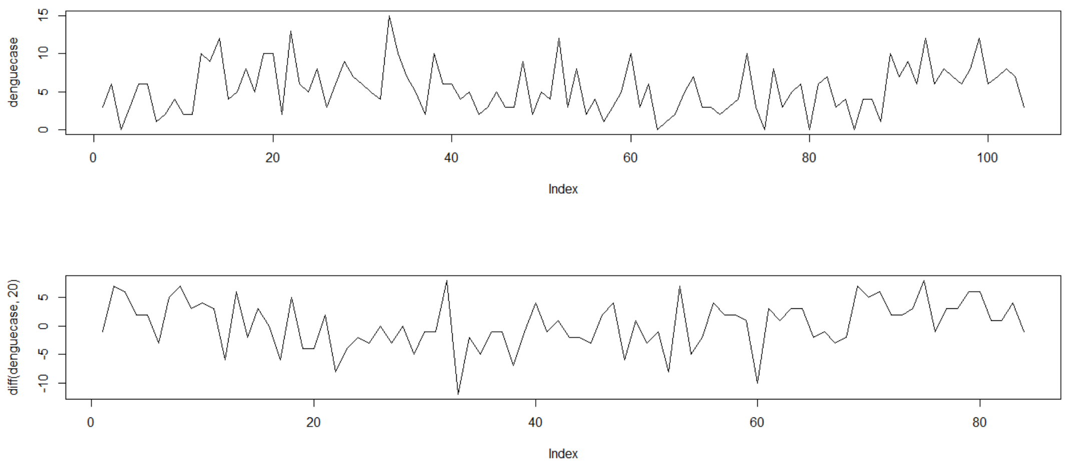


FIGURE 4.53. A. Plot of original dengue cases B. Plot of dengue cases after first seasonal difference

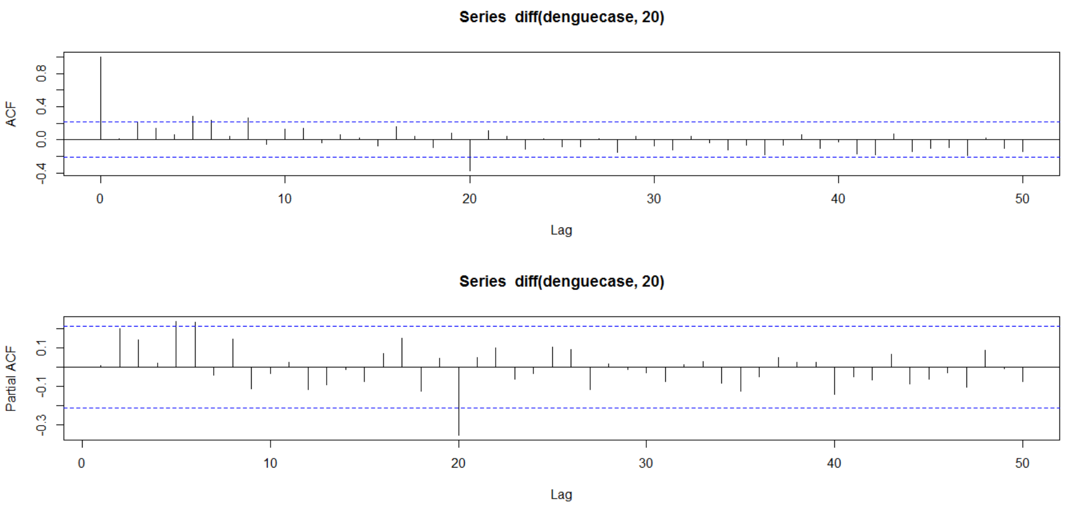


FIGURE 4.54. Plot of ACF and PACF of dengue cases after first non-seasonal and seasonal difference

Based on plot of PACF, maximum p value is 0, P parameter is 1. Meanwhile ACF plot shows that maximum q value is 1 with Q parameter of 1 maximum. Hence numbers of ARIMA is identified and assessed.

4.6.2 Model Estimation Step

TABLE 4-11. AIC values for different ARIMA model

Models	AIC	Models	AIC
ARIMA$(0,0,1)(1,1,1)_{20}$	462.26	ARIMA$(1,0,1)(0,1,1)_{20}$	455.22
ARIMA$(0,0,2)(1,1,1)_{20}$	459.74	ARIMA$(1,0,2)(0,1,1)_{20}$	453.1
ARIMA$(1,0,0)(1,1,1)_{20}$	776.15	ARIMA$(0,0,1)(1,1,0)_{20}$	469.37
ARIMA$(1,0,1)(1,1,1)_{20}$	457.02	ARIMA$(0,0,2)(1,1,0)_{20}$	466.68
ARIMA$(1,0,2)(1,1,1)_{20}$	463.77	ARIMA$(1,0,0)(1,1,0)_{20}$	469.1
ARIMA$(0,0,1)(0,1,1)_{20}$	460.3	ARIMA$(1,0,1)(1,1,0)_{20}$	461.26
ARIMA$(0,0,2)(0,1,1)_{20}$	457.79	ARIMA$(1,0,2)(1,1,0)_{20}$	470.9
ARIMA$(1,0,0)(0,1,1)_{20}$	460.29		

*ARIMA$(1,0,2)(0,1,1)_{20}$ is selected for having lowest AIC

4.6.3 Model Validation Step

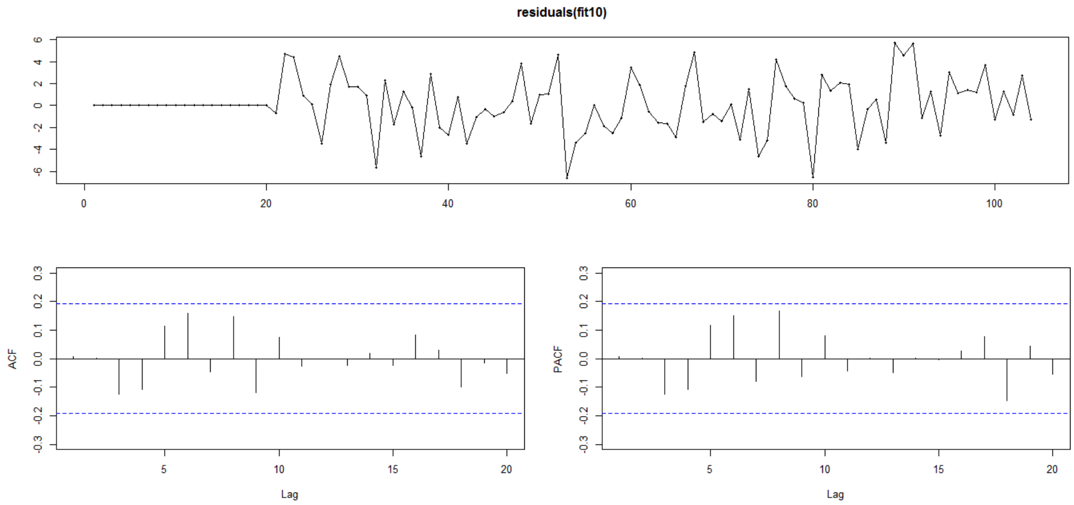


FIGURE 4.55. Residuals from the ARIMA$(1,0,2)(0,1,1)_{20}$ model.

ACF of residual after applying model is within significance limits shows that no other information can be extracted and it is equal to white noise.

4.6.4 Influence of climate variable

a. Temperature

Maximum lag of correlation between temperature with dengue cases recorded is at lag +7. The dengue cases recorded is higher before 7 weeks of higher temperature.

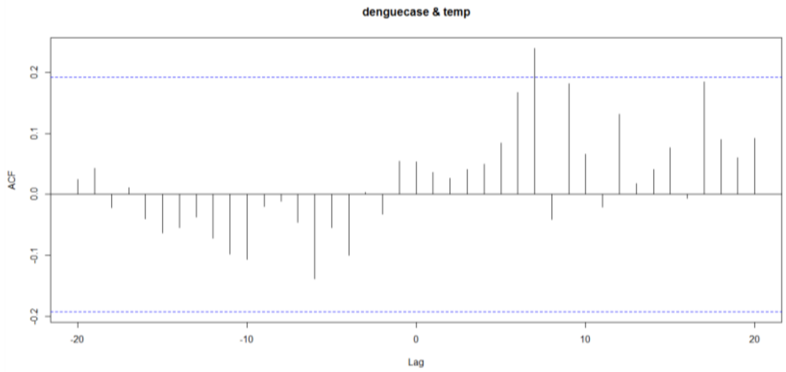


FIGURE 4.56. Cross-correlation between dengue cases recorded and mean temperature in Hulu Selangor

b. Humidity

Maximum lag of correlation between humidity with dengue cases recorded is at lag +4. The dengue cases recorded is higher before 4 weeks of higher humidity.

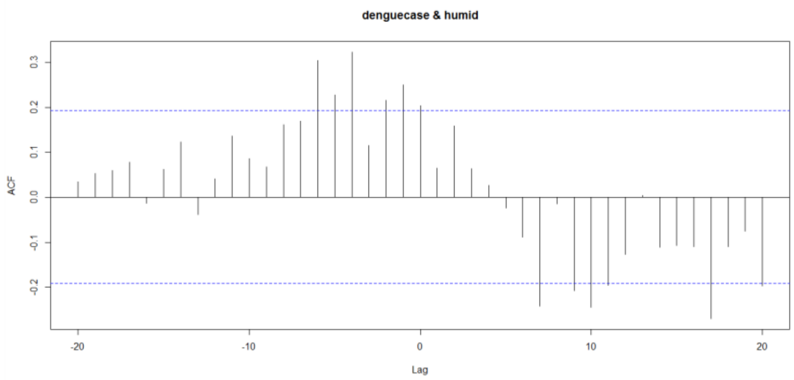


FIGURE 4.57. Cross-correlation between dengue cases recorded and humidity in Hulu Selangor

c. Rainfall

No positive correlation between humidity with dengue cases recorded.

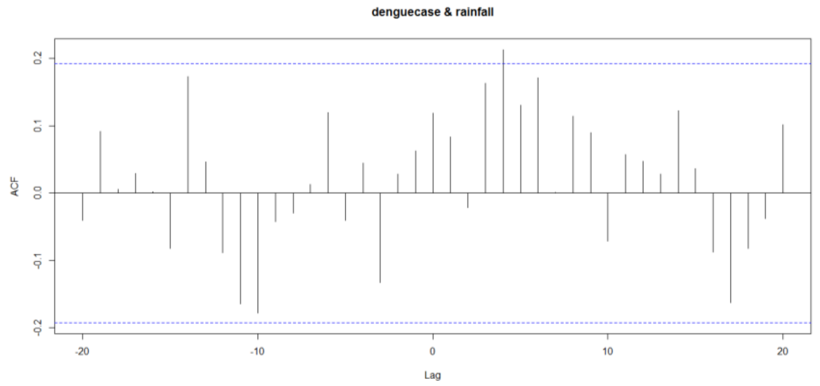


FIGURE 4.58. Cross-correlation between dengue cases recorded and rainfall in Hulu Selangor

4.6.5 Model Forecasting Step (Univariate)

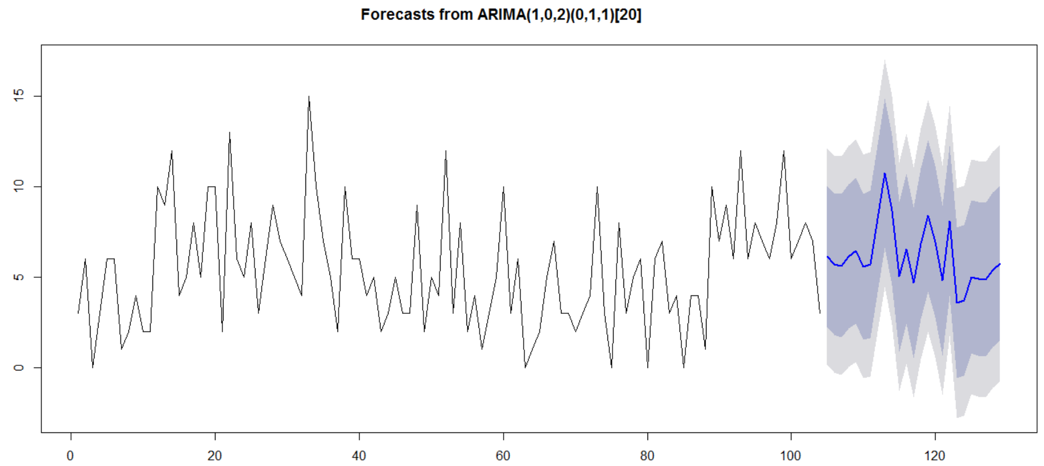


FIGURE 4.59. Forecasts from the ARIMA$(1,0,2)(0,1,1)_{20}$ model applied to the dengue recorded case data

4.6.6 Model Forecasting Step (Multivariate with Temperature)

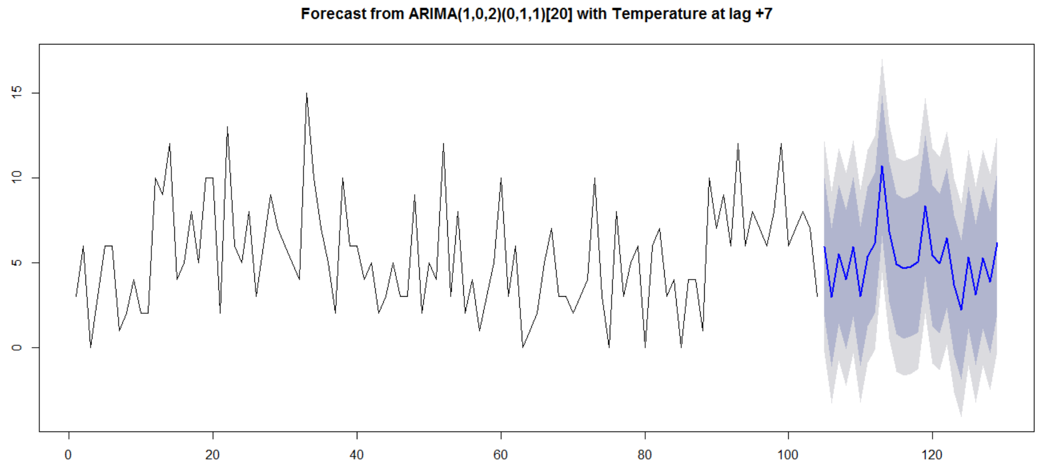


FIGURE 4.60. Forecasts from the ARIMA(1,0,2)$(0,1,1)_{20}$ model with temperature variable of lag +7 applied to the dengue recorded case data.

4.6.7 Model Forecasting Step (Multivariate with Humidity)

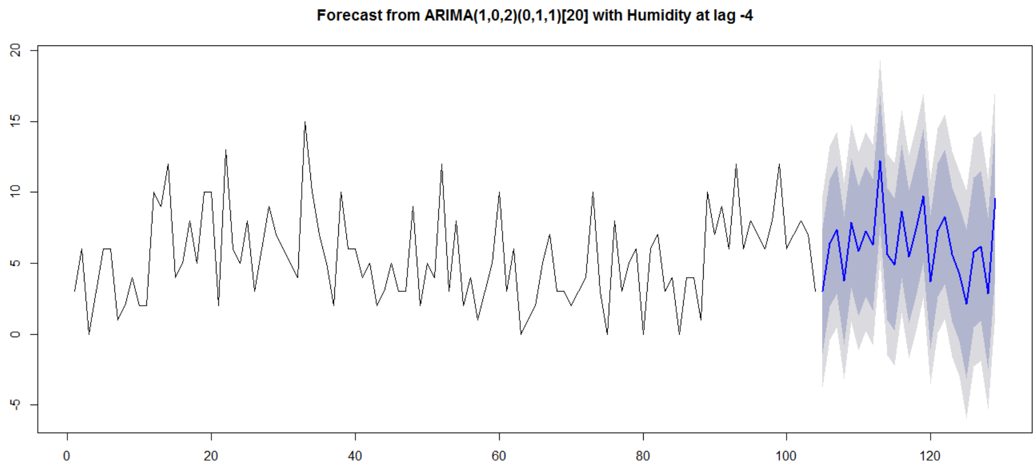


FIGURE 4.61. Forecasts from the ARIMA(1,0,2)$(0,1,1)_{20}$ model with humidity variable of lag -4 applied to the dengue recorded case data.

4.6.8 Model comparison

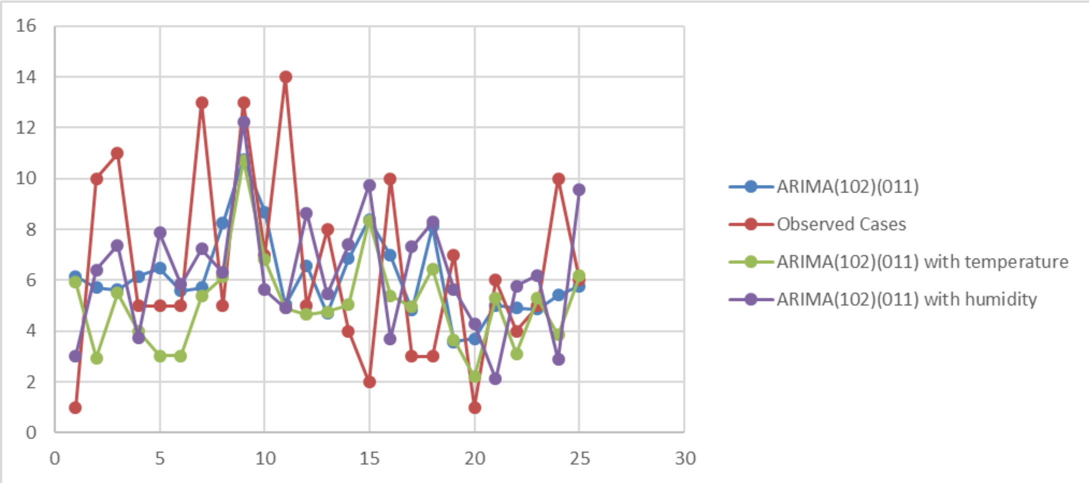


FIGURE 4.62. Model comparison between original dengue cases, forecasted cases using Univariate ARIMA and Multivariate ARIMA

4.6.9 Final Model

TABLE 4-12. Coefficient value of the final ARIMA model.

Model	AR1	MA1	MA2	SMA1	AIC	RMSE
$ARIMA(1,0,2)(0,1,1)_{20}$	0.9506	-1.040	0.2238	-0.999	453.1	2.6110 29
$ARIMA(1,0,2)(0,1,1)_{20}$ with Temperature	-0.994	1.0080	0.0769	-0.999	463.77	2.5004 65
$ARIMA(1,0,2)(0,1,1)_{20}$ with Humidity	0.9496	-0.821	-0.068	-0.534	461.26	3.1011 31

ARIMA $(1,0,2)(0,1,1)_{20}$ with temperature variable recoded lowest error. Results shows that ARIMA model with humidity variable has failed to improve the model shown as the higher value of RMSE. Multivariate ARIMA with temperature variable has been successfully improved the model even though some of the pattern is opposite from the original value.

4.7 Sepang

4.7.1 Model Identification step

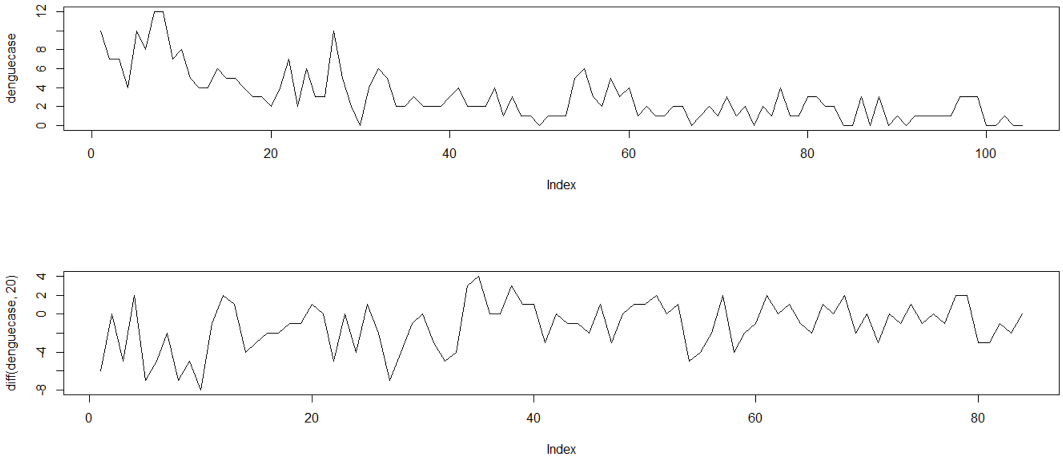


FIGURE 4.63. A. Plot of original dengue cases B. Plot of dengue cases after first seasonal difference

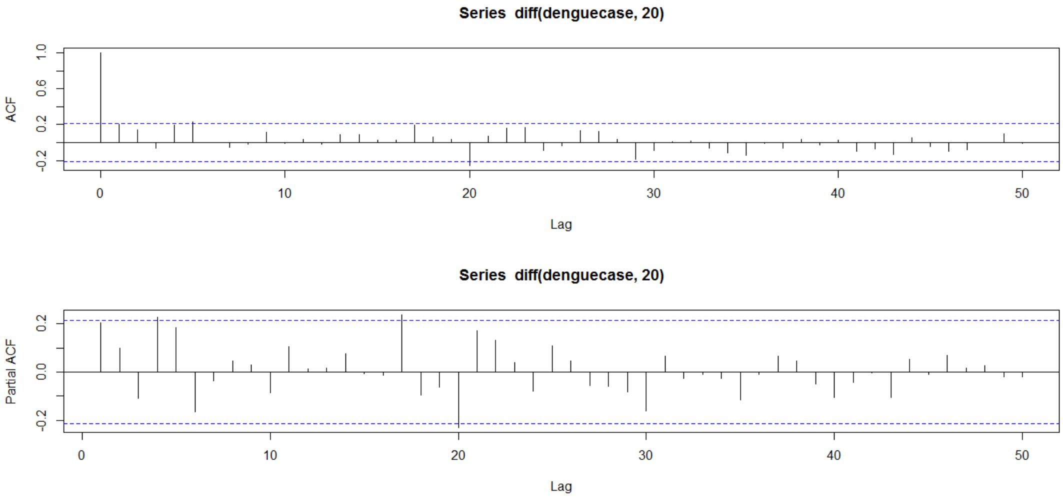


FIGURE 4.64. Plot of ACF and PACF of dengue cases after first seasonal difference

Based on plot of PACF, maximum p value is 0, P parameter is 1. Meanwhile ACF plot shows that maximum q value is 1 with Q parameter of 1 maximum. Hence numbers of ARIMA is identified and assessed.

4.7.2 Model Estimation Step

TABLE 4-13. AIC values for different ARIMA model

Models	AIC	Models	AIC
$ARIMA(0,0,1)(1,1,1)_{20}$	408.89	$ARIMA(1,0,1)(0,1,1)_{20}$	385.95
$ARIMA(0,0,2)(1,1,1)_{20}$	399.99	$ARIMA(1,0,2)(0,1,1)_{20}$	386.98
$ARIMA(1,0,0)(1,1,1)_{20}$	400.36	$ARIMA(0,0,1)(1,1,0)_{20}$	408.73
$ARIMA(1,0,1)(1,1,1)_{20}$	387.94	$ARIMA(0,0,2)(1,1,0)_{20}$	398.13
$ARIMA(1,0,2)(1,1,1)_{20}$	388.98	$ARIMA(1,0,0)(1,1,0)_{20}$	398.39
$ARIMA(0,0,1)(0,1,1)_{20}$	409.84	$ARIMA(1,0,1)(1,1,0)_{20}$	387.89
$ARIMA(0,0,2)(0,1,1)_{20}$	399.17	$ARIMA(1,0,2)(1,1,0)_{20}$	470.9
$ARIMA(1,0,0)(0,1,1)_{20}$	399.67		

*$ARIMA(1,0,1)(0,1,1)_{20}$ is selected for having lowest AIC

4.7.3 Model Validation Step

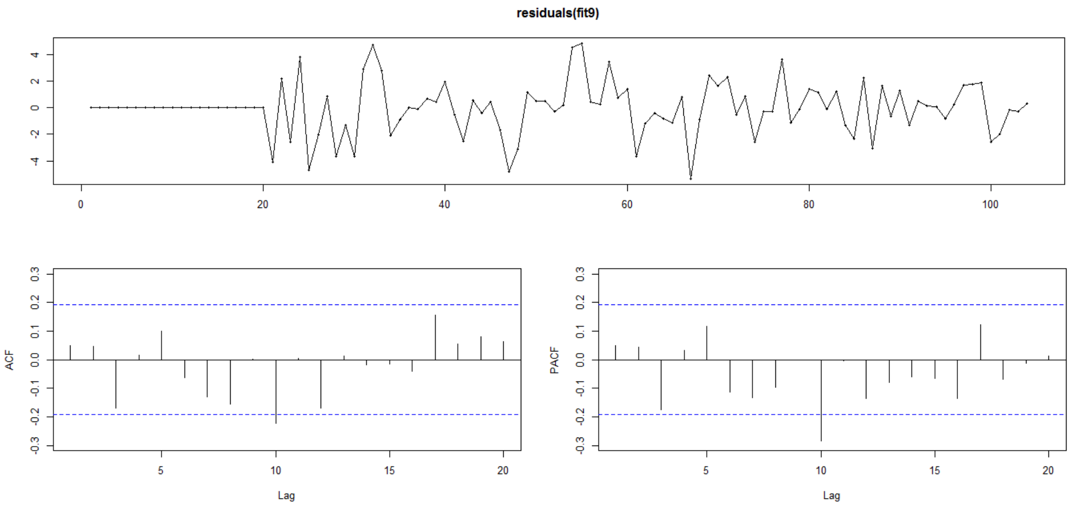


FIGURE 4.65. Residuals from the $ARIMA(1,0,1)(0,1,1)_{20}$ model.

ACF of residual after applying model is within significance limits shows that no other information can be extracted and it is equal to white noise.

4.7.4 Influence of Climate Variable

a. Temperature

Maximum lag of correlation between temperature with dengue cases recorded is at lag -14. The dengue cases recorded is higher after 14 weeks of higher temperature.

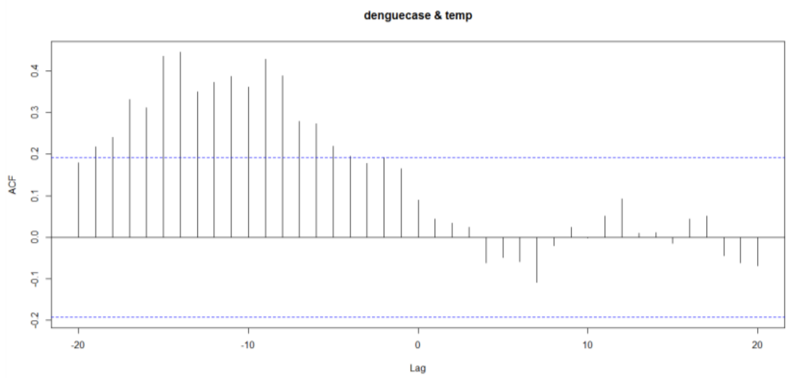


FIGURE 4.66. Cross-correlation between dengue cases recorded and mean temperature in Sepang

b. Humidity

No positive correlation between humidity with dengue cases recorded.

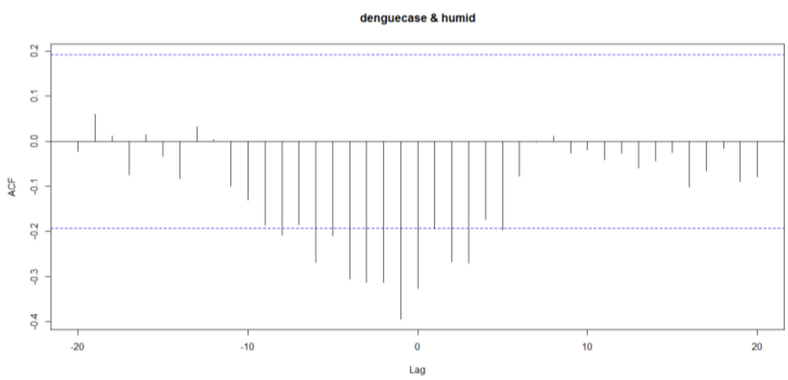


FIGURE 4.67. Cross-correlation between dengue cases recorded and humidity in Sepang

c. Rainfall

Maximum lag of correlation between rainfall with dengue cases recorded is at lag -12. The dengue cases recorded is higher after 12 weeks of higher rainfall rate.

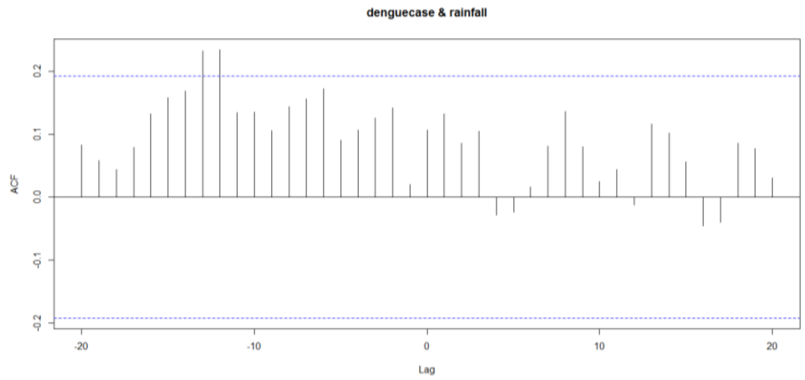


FIGURE 4.68. Cross-correlation between dengue cases recorded and rainfall in Sepang

4.7.5 Model Forecasting Step (Univariate)

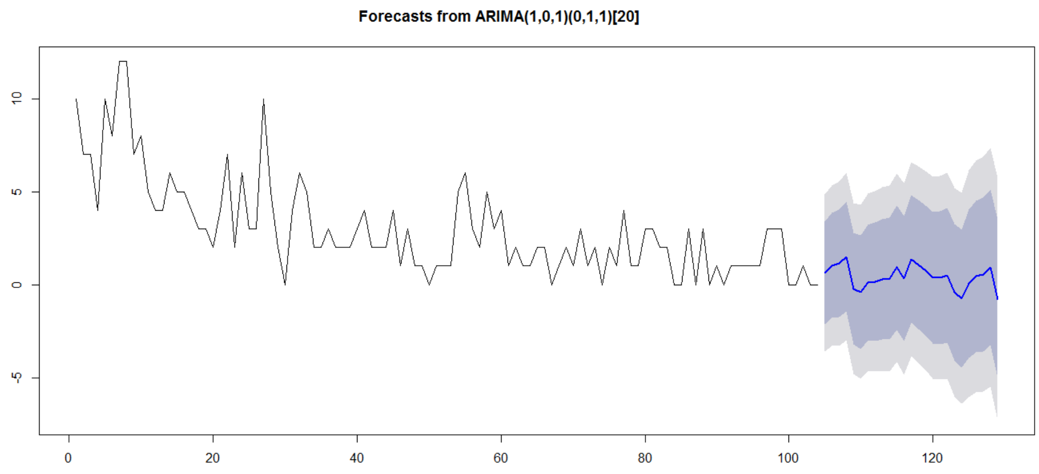


FIGURE 4.69.Forecasts from the ARIMA$(1,0,1)(0,1,1)_{20}$ model applied to the dengue recorded case data

4.7.6 Model Forecasting Step (Multivariate with Temperature)

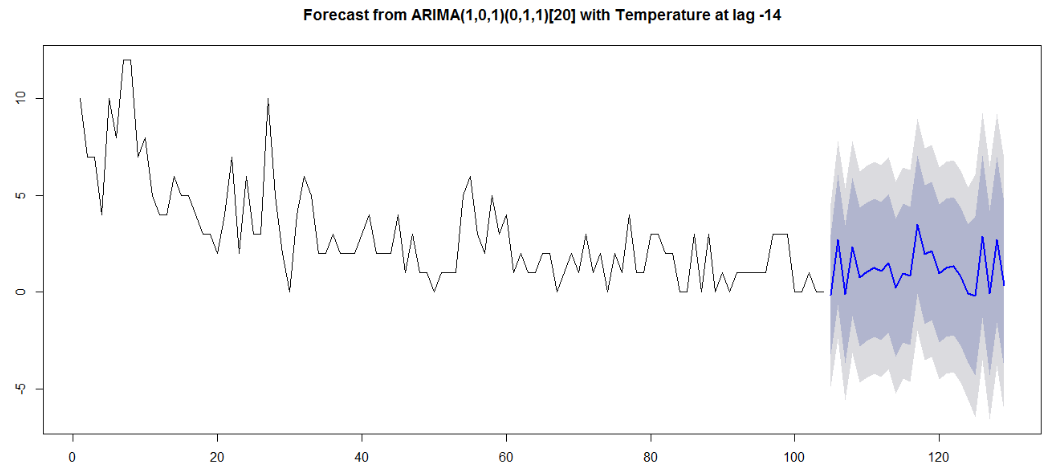


FIGURE 4.70. Forecasts from the ARIMA$(1,0,1)(0,1,1)_{20}$ model with temperature variable of lag -14 applied to the dengue recorded case data.

4.7.7 Model Forecasting Step (Multivariate with Rainfall)

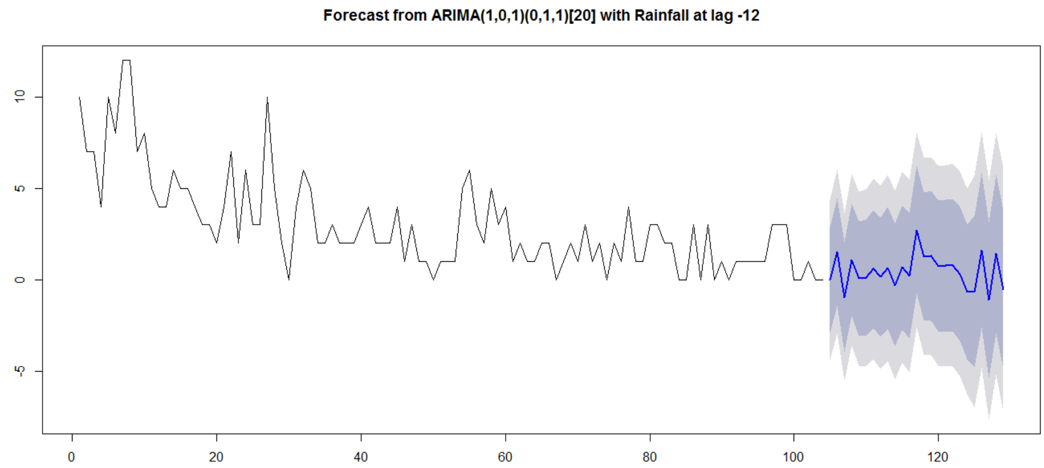


FIGURE 4.71. Forecasts from the ARIMA$(1,0,1)(0,1,1)_{20}$ model with rainfall variable of lag -12 applied to the dengue recorded case data.

4.7.8 Model comparison

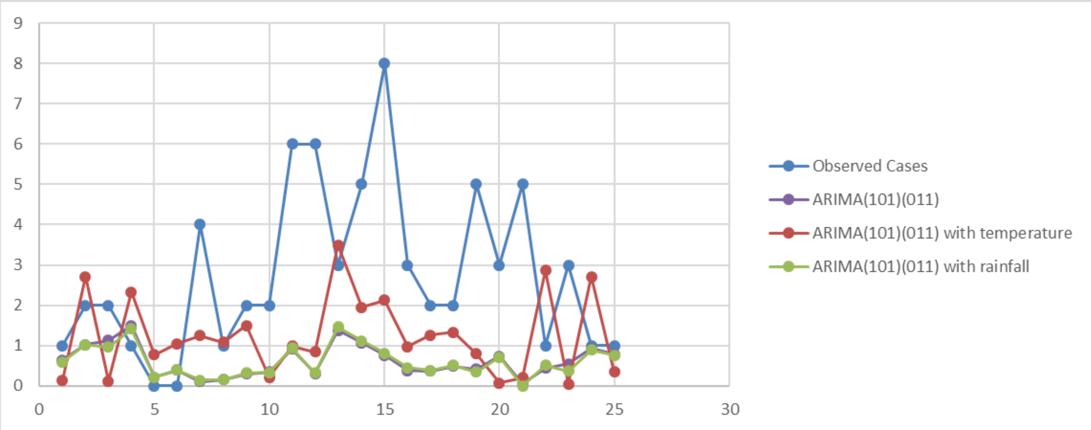


FIGURE 4.72. Model comparison between original dengue cases, forecasted cases using Univariate ARIMA and Multivariate ARIMA

4.7.9 Final Model

TABLE 4-14 Coefficient value of the final ARIMA model.

Model	AR1	MA1	SMA1	AIC	RMSE
$ARIMA(1,0,1)(0,1,1)_{20}$	0.9932	-0.7739	-0.6546	385.95	1.925673
$ARIMA(1,0,1)(0,1,1)_{20}$ with Temperature	0.4104	0.4188	0.2092	399.99	2.164527
$ARIMA(1,0,1)(0,1,1)_{20}$ with Rainfall	0.978	-0.7561	-0.4866	387.89	2.011531

$ARIMA(1,0,1)(0,1,1)_{20}$ with temperature variable recoded highest error. However, the value of the multivariate ARIMA with temperature variable shows the nearest to the observed cases compared to both univariate and multivariate with rainfall which value almost 0 for every week. Results shows that multivariate ARIMA model has failed to improve the model shown as the higher value of RMSE.

4.8 Summary

TABLE 4-15 Summary of correlation between climate variable and dengue cases recorded

District	Model	External Climate Variable		
		Temperature	Humidity	Rainfall
Gombak	$ARIMA(0,0,2)(2,1,1)_{20}$	Correlation at lag -14	No positive correlation	No correlation
Petaling	$ARIMA(1,1,2)(1,1,0)_{20}$	Correlation at lag -13	No positive correlation	Correlation at lag -5 and lag -16
Klang	$ARIMA(0,1,1)(1,1,1)_{20}$	Correlation at lag -5	Correlation at lag -20	Correlation at lag +8
Kuala Selangor	$ARIMA(1,0,1)(1,1,1)_{20}$	Correlation at lag +9	Correlation at lag -5	Correlation at lag +6
Hulu Langat	$ARIMA(0,1,2)(0,1,1)_{20}$	Correlation at lag -5	No positive correlation	Correlation at lag +6 and lag -4
Hulu Selangor	$ARIMA(1,0,2)(0,1,1)_{20}$	Correlation at lag +7	Correlation at lag -4	Correlation at lag +4
Sepang	$ARIMA(1,0,1)(0,1,1)_{20}$	Correlation at lag -14	No positive correlation	Correlation at lag -12

Based on the results, temperature has greatest influence among the three climate variable with average lag of -8. Temperature may indirectly influence the development of mosquitoes by reducing the duration of their life cycle which is 6 weeks. Averagely, mosquito has 6 weeks to spread the dengue. Rainfall is observed capable to either increase the transmission by promoting breeding places or eliminating breeding sites through heavy rainfall. It is because the correlation between rainfall and dengue cases recorded is sometimes ahead and before the dengue outbreak. Heavy rainfall usually will wash out the small water reservoir such as tires or vase, while drizzle rain might create the places for the larvae. Lastly, humidity plays indefinite role on dengue incidence as most of the ARIMA with humidity does not improve the model and sometimes no positive correlation between humidity and dengue outbreak. Hence, it is concluded that humidity does not affect much to dengue outbreak.

In short, the temperature variable is increasing the ARIMA model power in most district in Selangor, while rainfall is considered unreliable to as external regressors for ARIMA model as it is not constant either ahead or before of the dengue cases. The rainfall variable results are listed as the failed variable to be include as reliable variables. In addition, humidity variables are also cannot be considered successful variables as in some district there are no correlation at all with the dengue cases recorded.

CHAPTER 5
CONCLUSION

Forecasting an outbreak is important in the future in order to reduce the amount of victims and dengue cases and to act as secondary plan to support primary action. This project research is hopefully discovering the best and effective method to forecast dengue virus with optimum lag time. ARIMA is chosen as the basic model, where the author is specifying the variable and procedure to use the model efficiently. The climate considered as the most reliable variable that have high influence to the dengue cases. R provides a platform for easiness and accurate modelling.

This project is focusing on research and learning process. The author is having difficulties in assessing the methodology and results of the ARIMA models as limited knowledge on the statistical area. However, the process of developing an ARIMA model is actually feasible for an electrical engineer as the basics statistical tool have been learned in earlier semester. Moreover, basics formulation and technical details for each tools are not necessary to perform ARIMA model as it can always be performed using R function and packages. The cross correlation between the model and external climate variable is done to know the optimum lag of climate with dengue cases.

Learning and studying ARIMA model can be very interesting, to know the relationship between real study area with the limited data, yet the study still can be achieved successfully. Hence, application of time series model can be widely used in any outbreak to predict the future of the disease as long as the main factor is known.

REFERENCE

[1] S. Bhatt, P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, *et al.*, "The global distribution and burden of dengue," *Nature,* vol. 496, pp. 504-507, 04/25/print 2013.

[2] C. P. Simmons, J. J. Farrar, N. van Vinh Chau, and B. Wills, "Dengue," *New England Journal of Medicine,* vol. 366, pp. 1423-1432, 2012/04/12 2012.

[3] L. P. Lounibos, "INVASIONS BY INSECT VECTORS OF HUMAN DISEASE," *Annual Review of Entomology,* vol. 47, pp. 233-266, 2002/01/01 2002.

[4] W. H. Organization, S. P. f. Research, T. i. T. Diseases, W. H. O. D. o. C. o. N. T. Diseases, W. H. O. Epidemic, and P. Alert, *Dengue: guidelines for diagnosis, treatment, prevention and control*: World Health Organization, 2009.

[5] D. J. Gubler, E. E. Ooi, S. Vasudevan, and J. Farrar, *Dengue and Dengue Hemorrhagic Fever, 2nd Edition*: CABI, 2014.

[6] Bernama, "Selangor records 75% rise in dengue deaths," in *The Star*, ed, 2015.

[7] J. S. Armstrong, "Illusions in regression analysis," *International Journal of Forecasting,* vol. 28, pp. 689-694, 7// 2012.

[8] A. Earnest, S. B. Tan, and A. Wilder-Smith, "Meteorological factors and El Niño Southern Oscillation are independently associated with dengue infections," *Epidemiology & Infection,* vol. 140, pp. 1244-1251, 2012.

[9] M.-J. Chen, C.-Y. Lin, Y.-T. Wu, P.-C. Wu, S.-C. Lung, and H.-J. Su, "Effects of Extreme Precipitation to the Distribution of Infectious Diseases in Taiwan, 1994–2008," *PLoS ONE,* vol. 7, p. e34651, 2012.

[10] F. J. Colón-González, I. R. Lake, and G. Bentham, "Climate Variability and Dengue Fever in Warm and Humid Mexico," *The American Journal of Tropical Medicine and Hygiene,* vol. 84, pp. 757-763, 10/25/ 2011.

[11] Y. L. Hii, H. Zhu, N. Ng, L. C. Ng, and J. Rocklöv, "Forecast of Dengue Incidence Using Temperature and Rainfall," *PLoS Neglected Tropical Diseases,* vol. 6, p. e1908, 11/29 2012.

[12] C. D. Nazri, A. Abu Hassan, Z. Abd Latif, and R. Ismail, "Impact of climate and landuse variability based on dengue epidemic outbreak in Subang Jaya," in *Humanities, Science and Engineering (CHUSER), 2011 IEEE Colloquium on*, 2011, pp. 907-912.

[13] W. Hu, A. Clements, G. Williams, and S. Tong, "Dengue fever and El Niño/Southern Oscillation in Queensland, Australia: a time series predictive model," *Occupational and Environmental Medicine,* vol. 67, pp. 307-311, May 1, 2010 2010.

[14] M. Gharbi, P. Quenel, J. Gustave, S. Cassadou, G. Ruche, L. Girdary*, et al.*, "Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors," *BMC Infectious Diseases,* vol. 11, pp. 1-13, 2011/06/09 2011.

[15] R. D. Farmer and J. Emami, "Models for forecasting hospital bed requirements in the acute sector," *Journal of Epidemiology and Community Health*, vol. 44, pp. 307-312, 1990.

[16] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*: Wiley, 2015.

[17] K. CHOI and S. B. THACKER, "AN EVALUATION OF INFLUENZA MORTALITY SURVEILLANCE, 1962–1979: II. PERCENTAGE OF PNEUMONIA AND INFLUENZA DEATHS AS AN INDICATOR OF INFLUENZA ACTIVITY," *American Journal of Epidemiology,* vol. 113, pp. 227-235, March 1, 1981 1981.

[18] F. F. Nobre, A. B. S. Monteiro, P. R. Telles, and G. D. Williamson, "Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology," *Statistics in Medicine,* vol. 20, pp. 3051-3069, 2001.

[19] M. Hurtado-Díaz, H. Riojas-Rodríguez, S. J. Rothenberg, H. Gomez-Dantés, and E. Cifuentes, "Short communication: Impact of climate variability on the incidence of dengue in Mexico," *Tropical Medicine & International Health,* vol. 12, pp. 1327-1337, 2007.

[20] A. L. Corwin, R. P. Larasati, M. J. Bangs, S. Wuryadi, S. Arjoso, N. Sukri*, et al.*, "Epidemic dengue transmission in southern Sumatra, Indonesia," *Transactions of the Royal Society of Tropical Medicine and Hygiene,* vol. 95, pp. 257-265, 2001.

[21] K. Nakhapakorn and N. K. Tripathi, "An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence," *International Journal of Health Geographics,* vol. 4, p. 13, 2005.

[22] D. A. Focks and R. Barrera, "Dengue transmission dynamics: assessment and implications for control," in *WHO Report of the Scientific Working Group meeting on Dengue, Geneva*, 2006.

[23] M. A. Johansson, F. Dominici, and G. E. Glass, "Local and Global Effects of Climate on Dengue Transmission in Puerto Rico," *PLoS Negl Trop Dis,* vol. 3, p. e382, 2009.

[24] B. Cazelles, M. Chavez, A. J. McMichael, and S. Hales, "Nonstationary Influence of El Niño on the Synchronous Dengue Epidemics in Thailand," *PLoS Med,* vol. 2, p. e106, 2005.

[25] G. Chowell, B. Cazelles, H. Broutin, and C. V. Munayco, "The influence of geographic and climate factors on the timing of dengue epidemics in Perú, 1994-2008," *BMC Infectious Diseases,* vol. 11, pp. 164-164, 06/08 2011.

[26] K. T. D. Thai, B. Cazelles, N. V. Nguyen, L. T. Vo, M. F. Boni, J. Farrar*, et al.*, "Dengue Dynamics in Binh Thuan Province, Southern Vietnam: Periodicity, Synchronicity and Climate Variability," *PLoS Negl Trop Dis,* vol. 4, p. e747, 2010.

[27] W. Hu, A. Clements, G. Williams, S. Tong, and K. Mengersen, "Spatial Patterns and Socioecological Drivers of Dengue Fever Transmission in Queensland, Australia," *Environmental Health Perspectives,* vol. 120, pp. 260-266, 10/20 2012.

[28] E. Descloux, M. Mangeas, C. E. Menkes, M. Lengaigne, A. Leroy, T. Tehei*, et al.*, "Climate-Based Models for Understanding and Forecasting Dengue Epidemics," *PLoS Negl Trop Dis,* vol. 6, p. e1470, 2012.

[29] M. N. Karim, S. U. Munshi, N. Anwar, and M. S. Alam, "Climatic factors influencing dengue cases in Dhaka city: A model for dengue prediction," *The Indian Journal of Medical Research,* vol. 136, pp. 32-39, 11/02/received 2012.