Bulletin of Electrical Engineering and Informatics
Vol. 8, No. 1, March 2019, pp. 305~312
ISSN: 2302-9285, DOI: 10.11591/eei.v8i1.1447

Haze alarm visual map (HazeViz): an intelligent haze forecaster

Mohd Said Syukri Morsid[1], Syeril Azira Jamaluddin[2], Nur Azmina Hood[3], Norshahida Shaadan[4], Yap Bee Wah[5], Muthukkaruppan Annamalai[6]
[1,2,3,4,5]Centre for Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[4,5,6]Advanced Analytics Engineering Centre, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Article Info

Article history:

Received Oct 15, 2018
Revised Nov 4, 2018
Accepted Dec 20, 2018

Keywords:

Big data analytics
Haze
PM10
SARIMA
Visualization

ABSTRACT

The haze problem has intensified in recent years. The particulate matter of less than 10 microns in size, PM10 is the dominant air pollutant during haze. In this paper, we present the development of HazeViz, a Haze Alarm Visual Map forecaster, which is based on PM10. The intelligent web application allows users to visualize the pattern of PM10 in a region, forecasts PM10 value and alarms bad haze condition. HazeViz was developed using HTML, Java Script, PHP, MySQL, R Programming and Fusionex Giant. The SARIMA statistical forecasting models that underlie the application were developed using R. The PM10 trend analysis, and the consequential map and chart visualizations were implemented on the Fusionex GIANT Big Data Analytics platform. HazeViz was developed in the context of the Klang Valley, our case study. The dataset was obtained from Department of Environment Malaysia, which contains a total of 157,680 hourly PM10 data for six stations in Klang Valley, for the years 2013 to 2015. The SARIMA models were developed using maximum daily PM10 data for 2013 and 2014, and the 2015 data was used to validate the model. The fitting models were determined based on the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). While the selected models were implemented in HazeViz and successfully deployed on the web, the results show that the selected models have MAPE ranging between 35 percent and 45 percent, which implies that the models are still far from robust. Future work can consider augmented SARIMA models that can yield improved results.

Corresponding Author:

Yap Bee Wah,
Centre for Statistical and Decision Science Studies,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
Email: beewah@tmsk.uitm.edu.my

1. INTRODUCTION

The haze problem has intensified in recent years. For instance, Malaysia has been facing increasing bad haze problems since the 1990s, which typically occur during the southwest monsoon season from July till September. One of the reasons of the haze occurrence was due to the trans-boundary sources of the agriculture fires in Indonesia, which not only affected Malaysia but also the neighboring countries such as Singapore, Thailand and Philippines [1]. The haze episodes posed serious threats to the health of the Malaysian community [2]. Haze has been reported to cause eye and skin irritations, bronchitis, asthma, acute respiratory illness and cardiovascular disease [3].

Air quality monitoring is part of the strategy in the pollution prevention program in Malaysia. The Air Pollutant Index (API) is calculated by taking into consideration the concentration of air pollutants

namely, sulphur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO), Ozone (O3) and suspended particulate matter (PM) [4]. Of these pollutants, the suspended particulate matter of less than 10 microns in size (PM10) is the chief cause of the cardio-respiratory mortality and morbidity among children and elderly [5]. Malaysia Ambient Air Quality Guidelines (MAAQG) state that the daily and monthly averages of the PM10 concentration levels should not exceed 150μg/m^3 and 50 μg/m^3, respectively [6].

Since PM10 is the dominant air pollutant during haze episodes, the study sets to develop an intelligent, web-based Haze Alarm Visual Map application called HazeViz to forecast the PM10 value and indicate whether the haze condition is alarming or not, as well as to visualize the pattern of PM10 in a region. Since Klang Valley has been experiencing bad haze conditions for the past many years, the PM10 data for Klang Valley was used as case study.

HazeViz was implemented on Fusionex GIANT [7], a Big Data Analytics and Visualization platform. The underlying time series forecast models were developed using R, an open source software environment for statistical computing. The web-based visualization can provide easy, fast and direct information about the haze condition to the public and relevant authorities, which can assist the user community to take precautionary measures during bad haze conditions. This paper is organized as follows. The review of related studies is covered in Section 2, the research method is described in Section 3, the result are discussed in Section 4 and the paper is concluded in Section 5.

2. RELATED STUDIES

The suspended particulate matter PM10 has been used as a proxy measure of haze. Wu et al. studied the haze situation in China and identified the determinants of PM2.5 (even smaller than PM10) using a random-effects model and a set of OLS regressions [8]. They reported that PM2.5 is significantly correlated with the industrial proportion, the number of motor vehicles, and household gas consumption.

Oanh et al. investigated the main causes of haze episodes in the northwestern Thailand to provide early warning and prediction [9]. A stepwise regression model was developed to predict hourly PM10 for days of meteorology pattern using the February–April data of years 2007–2009. The model performed satisfactorily for dataset (R^2=81%) with the input variables PM10 averaged over two stations in Chiangmai on the previous day.

There are number of related studies in Malaysia. Juneng et al. studied spatio-temporal characteristics of PM10 concentration across Malaysia [10]. They found that the PM10 concentration fluctuates markedly in two timescale bands i.e., 10–20 days quasi-biweekly (QBW) and 30–60 days lower frequency (LF) band of the intra-seasonal timescales. Shaadan et al. used robust projection pursuit and robust Mahalanobis distance methods to detect anomalies in PM10 functional data obtained from three air-quality monitoring stations in Klang Valley [11]. Hamid et al. considered two seasons, i.e., wet season (northeastern monsoon) and dry season (southwestern monsoon) and developed seasonal autoregressive integrated moving average model to predict the PM10 concentration in Negeri Sembilan [12]. They reported that Seasonal ARIMA (SARIMA) was a suitable model in predicting the PM10 concentration levels. Lee et al. also used SARIMA for forecasting the API value in Johor [13], while Siew et al. developed ARIMA and Integrated ARFIMA models for forecasting the API reading in Shah Alam [14].

3. RESEARCH METHOD

This section covers the design and development of HazeViz that adopts the approach of machine learning [15]. The description is divided into three subsections: Data preparation, Model development and Application design and development.

3.1. Data preparation

The scope of the study covers six air quality monitoring stations located in Klang Valley, namely Klang, Petaling Jaya, Shah Alam, Kuala Selangor, Batu Muda and Banting. The PM10 dataset was obtained from the Department of Environment (DOE), Malaysia. The dataset contains a total of 157,680 hourly PM10 data for each of the six air quality monitoring stations for the years 2013 to 2015. For ease of management, the data were stored in six different CSV files according to the station. The data was checked for missing values and outliers. The missing values for each station was recorded by month and year, and were imputed by using Mean Top Bottom (MTB) method [11]. MTB averages the observation on the top and at the bottom of the missing value. The data was summarized into daily PM10 data by selecting the maximum concentration level of PM10 of day. Subsequently, the daily PM10 data was used to develop the time series forecasting models. The computations of the alarming index is based on hourly and daily basis assessment described as follows [16]:

Let y be the PM10 concentration level. Let A=1 indicates alarming condition and A=0 indicates a non- alarming condition.

a. The i^{th} hour is alarming (A=1) if y(i) > 200 ug/m^3, else A=0.
b. The j^{th} day is alarming (A=1) if there is at least one of the hours within the day is alarming, else A=0.
c. The k^{th} week is alarming (A=1) if there is at least one of the days in the week is alarming, else A=0.

3.2. Forecasting models

The forecasting models for each station were developed using time-series SARIMA to forecast maximum daily PM10. There are three stages in developing a SARIMA model [17]. At the first stage, a simple data investigation using line charts was conducted to understand the basic pattern of series to identify if any unusual observation or characteristics exists and to check if the data is stationary. Note that SARIMA requires stationary data. The ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) were plotted to get more conclusive evidence of the stationary condition.

At the second stage, the first differencing is performed if the data series appears non-stationary. If seasonality exists, then seasonal differencing is also performed. The ACF and PACF plots of the final series were used to confirm the stationary condition.

At the third state, the model is identified. Previous studies developed ARIMA and Integrated ARFIMA for forecasting API values [14]. Juneng et al. reported that SARIMA was suitable for predicting PM10 value [10]. In this study, due to the seasonal pattern of haze, we developed SARIMA to forecast maximum daily PM10 value. SARIMA $(p,d,q)\times(P,D,Q)s$ model are defined by six parameters, namely autoregressive (AR) part of order p, moving average (MA) part of order q, differencing (I) of order d, seasonal autoregressive (SAR) part of order P, seasonal moving average (SMA) part of order Q, the period of the seasonal period pattern appearing as D (I) and the period of seasonal pattern appearing (s). SARIMA can be expressed as shown equation 1 [12].

$$(1-\emptyset_1 B-\emptyset_2 B^2-\ldots-\emptyset_p B^p)(1-\beta_1 B^s-\beta_2 B^{2s}-\cdots-\beta_P B^{Ps})(1-B)^d(1-B^s)^D y_t$$
$$\text{AR}(p) \qquad \text{SAR}(P) \qquad \text{I}(d) \quad \text{I}_s(\text{D}) \qquad (1)$$
$$= C+\underbrace{(1-\psi_1 B-\psi_2 B^2-\cdots-\psi_q B^q)}_{\text{MA}(q)}\underbrace{(1-\theta_1 B^s-\theta_2 B^{2s}-\ldots-\theta_Q B^{Qs})}_{\text{SMA}(Q)}\varepsilon_t$$

The Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are error measures commonly used to determine the accuracy of the forecasting models [14]. The MAE, RMSE and MAPE measures are expressed by equations 2, 3 and 4, respectively. The variables X_i and $\hat{X}$ in the equations are the actual and the predicted values, respectively, while n is the number of observations:

$$MAE=\frac{\Sigma_{i=1}^{n}|X_i-\hat{X}|}{n} \qquad (2)$$

$$RMSE=\sqrt{\frac{\Sigma(X_i-\hat{X})^2}{n}} \qquad (3)$$

$$MAPE=\frac{\left|\frac{X_i-\hat{X}}{X_i}\right|}{n}\times 100\% \qquad (4)$$

In developing the SARIMA model, the maximum daily PM10 data for the years 2013 and 2014 were used to train the models, while the data for 2015 was used to validate the models. There is a seasonal pattern of haze every three months or so, and the presence of the seasonal component in the data series is also revealed by the ACF plots. In order to eliminate the seasonal component, seasonal differencing was performed. Through trial and error upto 105 days, we found the seasonal differencing of 60 days producing the best results. Then, first differencing was carried out to achieve stationary. Augmented Dickey-Fuller (ADF) test was conducted on the series to check if the data is stationary afterwards.

For SARIMA (p,d,q)(P,D,Q)s, the number of significant lags in the PACF plot was used to obtain the p value for AR (Auto-Regressive), while the number of significant lag in ACF was used to determine the q value for moving average (MA). We used 0 and 1 for P and Q in the seasonal part of the model. The SARIMA models and forecast error measures for each station are shown in Table 1.

Table 1. SARIMA models and error measures

Station	Model	RMSE	MAE	MAPE
Klang	$SARIMA(1,1,1)(1,1,1)_{60}$	91.0021	55.2708	36.8221
	$SARIMA(3,1,4)(1,1,1)_{60}$	90.7678	55.1412	36.8629
	$SARIMA(3,1,5)(1,1,1)_{60}$*	89.8525	54.3888	35.2812
Petaling Jaya	$SARIMA(1,1,1)(1,1,0)_{60}$	75.5410	45.1373	42.3819
	$SARIMA(3,1,5)(1,1,0)_{60}$	75.5444	45.1489	42.4421
	$SARIMA(2,1,2)(1,1,0)_{60}$*	75.5379	45.0009	42.3760
Shah Alam	$SARIMA(5,1,7)(0,1,1)_{60}$	50.1323	33.7377	50.4625
	$SARIMA(1,1,1)(0,1,1)_{60}$*	46.9684	26.7995	35.8477
	$SARIMA(5,1,1)(0,1,1)_{60}$	50.2247	33.6193	50.7124
Kuala Selangor	$SARIMA(7,1,5)(0,1,1)_{60}$*	77.4491	43.5666	36.4137
	$SARIMA(2,1,1)(0,1,1)_{60}$	77.3003	43.6652	36.4137
	$SARIMA(5,1,1)(0,1,1)_{60}$	77.2874	43.6365	37.0009
Batu Muda	$SARIMA(4,1,3)(1,1,0)_{60}$	77.1685	47.1332	45.2537
	$SARIMA(4,1,1)(1,1,0)_{60}$	77.0888	47.0397	44.9748
	$SARIMA(5,1,1)(1,1,0)_{60}$*	77.1458	46.8300	44.7359
Banting	$SARIMA(3,1,5)(1,1,0)_{60}$	85.1213	51.6883	36.2121
	$SARIMA(1,1,3)(1,1,0)_{60}$	85.1231	51.6921	36.2182
	$SARIMA(4,1,3)(1,1,0)_{60}$*	85.1199	51.5071	36.2082

* Selected model

The best forecast model for each station is selected based on the lowest error measures. In determining the fitting models, we relied on MAPE as the primary measure to determine the accuracy of a model, and supported by RMSE and MAE measures. MAPE calculates the percentage difference between the actual and forecasted values. The selected forecast model for each station is asterixed in Table 1. The mathematical expressions of the models and their graphical illustrations are shown in Figure 1. The X- and Y-axes represent the days in year and the PM10 concentration levels, respectively. The actual and foreasted PM10 lines for each station delineates the performance of selected model in forecasting the PM10 values. The model is said to perform well when its forecasted value is close to the actual value.

KLANG–SARIMA (3,1,5)(1,1,0)

$$y_t = \frac{(1-0.626\mathrm{B})(1-0.5408B^2)(1+0.0019B^3)}{(1-0.1255B^4)(1+0.29606B^5)\varepsilon_t}{(1+0.2746B)(1+0.5962B^2)(1+0.0069B^3)}{(1-0.4594B^{60})(1-B)(1-B^{60})}$$

PETALING JAYA–SARIMA (2,1,2)(1,1,0)

$$y_t = \frac{(1-0.3895B)(1-0.6105B^2)\varepsilon_t}{(1+0.0713B)(1+0.5025B^2)(1-0.4667B^{60})}{(1-B)(1-B^{60})}$$

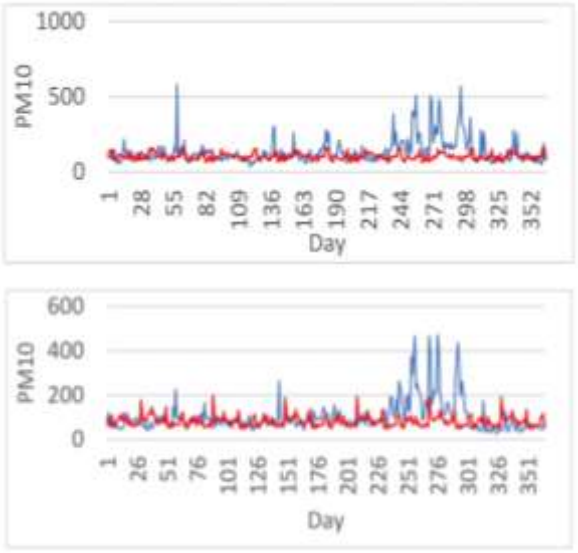

Figure 1. Selected SARIMA model of stations, and their actual and forecasted PM10 line charts

SHAH ALAM–SARIMA (1,1,1)(1,1,1)

$$y_t = \frac{(1-0.9224B)(1-0.6105B^{60})(1-0.6105B^{120})\varepsilon_t}{(1+0.5821B)\,(1-B)(1-\,B^{60})}$$

KUALA SELANGOR–SARIMA (7,1,5)(0,1,1)

$$y_t = \frac{\begin{matrix}(1-0.9806B)(1-01349B^2)(1+0.001938B^3)\\(1-1.1272B^4)(1+0.7798B^5)(1-0.9974B^{60})\varepsilon_t\end{matrix}}{\begin{matrix}(1-0.6229B)(1-0.2121B^2)(1-0.0671B^3)\\(1+0.9844B^4)(1-0.5160B^5)\\(1-0.1631B^6)(1-0.0994B^7)\\(1-B)(1-B^{60})\end{matrix}}$$

BATU MUDA–SARIMA (5,1,1)(1,1,0)

$$y_t = \frac{(1-1.0000B)\varepsilon_t}{\begin{matrix}(1-0.5208B)(1-0.1151B^2)(1+0.0005B^3)\\(1-0.0142B^4)(1-\,0.0682B^5)\\(1+0.4602B^{60})(1-B)(1-B^{60})\end{matrix}}$$

BANTING–SARIMA (4,1,3)(1,1,0)

$$y_t = \frac{(1-0.6475B)(1-0.7447B^2)(1+0.3922B^3)\varepsilon_t}{\begin{matrix}(1+0.2130B)(1+0.8250B^2)(1-0.0538B^3)\\(1-0.1055B^4)(1-0.4607B^{60})\\(1-B)(1-B^{60})\end{matrix}}$$

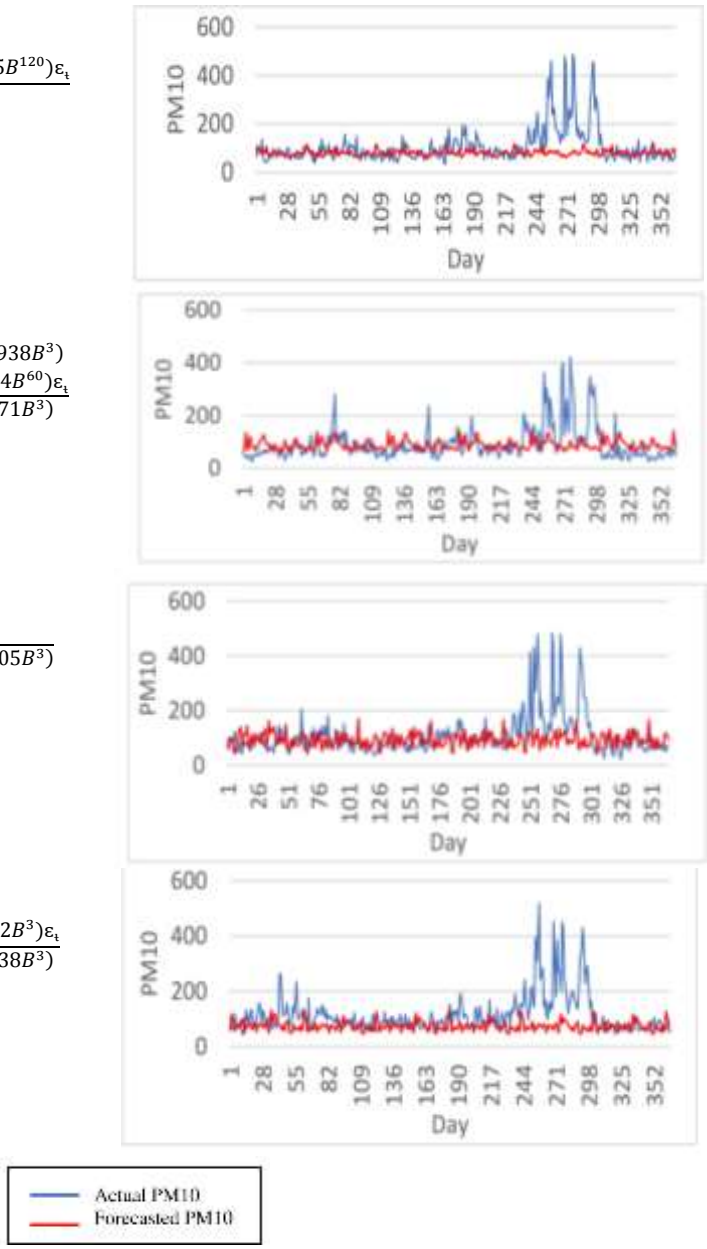

Figure 1. Selected SARIMA model of stations, and their actual and forecasted PM10 line charts

3.3. Application design and development

The design and development of HazeViz is described in terms of the program and the user-interface.

3.3.1. Program

HazeViz was designed to forecast future maximum PM10 value, as well as to retrieve past maximum PM10 value for each station for a specified date. The HazeViz main program consists of two procedures: *GetHistory* and *Forecast*, which are linked to Fusionex GIANT to allow the visualization of the map and charts as outlined in the HazeViz processing steps as shown in Figure 2. The *GetHistory* and *Forecast* procedures are described below.

Procedure *GetHistory*

a. *GetHistory* will execute *AnalyseHistory* to retrieve the PM10 Station Data (PMSD) stored in CSV format.
b. *AnalyseHistory* will extract the data from PMSD for the specified date for each station.
c. The extracted data will be saved in an output file and sent to the public web server.
d. Fusionex Giant will update the Map Chart based on the data in the output file.

e. *GetHistory* will open the webpage to display the Map Chart result.

Procedure *Forecast*

a. *Forecast* will execute *AnalyseHistory* to retrieve the PMSD stored in CSV format (c.f. item 1 above).
b. Next, *Forecast* will execute *AnalyseForecast* that runs the SARIMA model on the PMSD for each station to forecast the PM10 value for the specified date.
c. The forecasted data will be saved in two output files; one for Fusionex Giant to update the Map Chart, and another to update the Bar/ Line Chart.
d. The output files will be sent to the public web server.
e. *Forecast* will open the webpage to display the Map Chart and Bar/ Line Chart results.

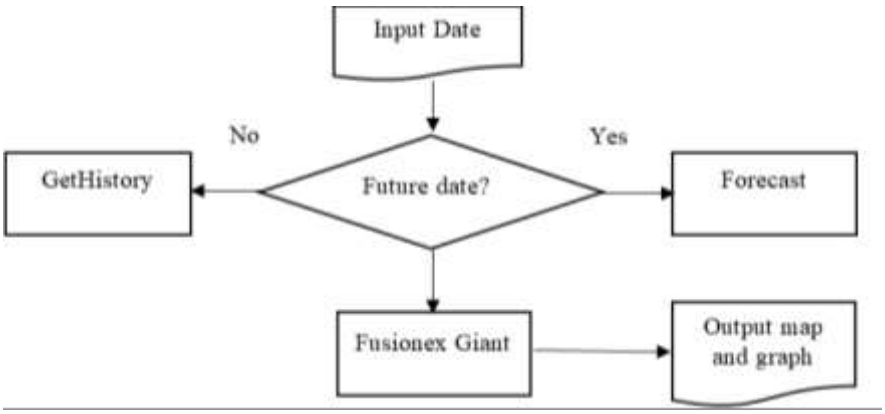

Figure 2. HazeViz processing steps

The HazeViz implementation framework is shown in Figure 3. The browser web page was constructed using HTML and JavaScript was used to extract data from the computer such as the current date, time and year. The HazeViz main program and its procedures are encoded in PHP.

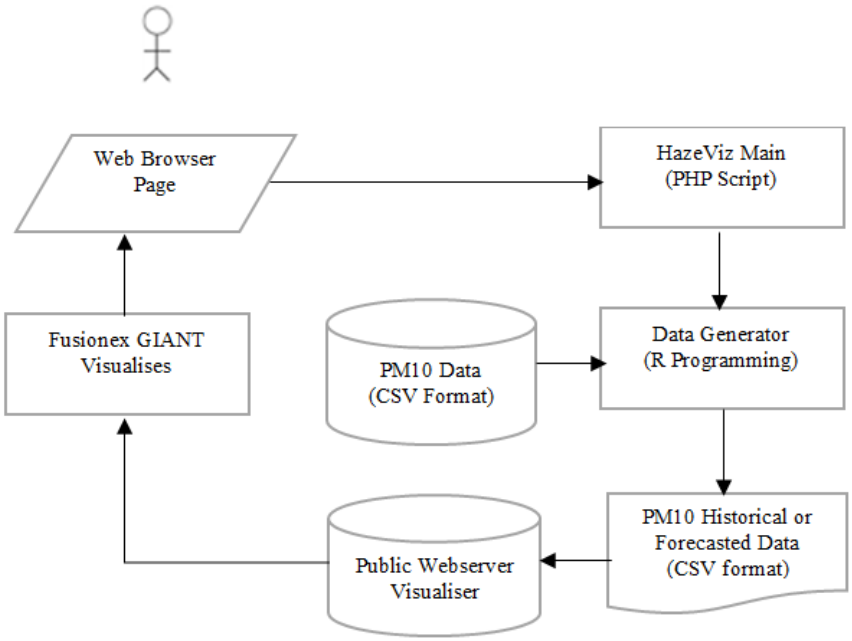

Figure 3. HazeViz implementation framework

AnalyseHistory that retrieves the PM10 Station Data (PMSD) and *AnalyseForecast* that implements the SARIMA model, are encoded in R. The two sub-procedures that produce the PM10 data that is sent to the

public web server are together referred to as *Data Generator*. The data used to visualise the map and bar/ line charts is stored in a MySQL database. The source PMSD and the intermediate data produced by Data Generator are stored in data files in CSV format. Fusionex GIANT is given access to these data files, which it uses to display the map and/ or bar/ line charts. The links to the map and/or charts produced by Fusionex GIANT are embedded in the web browser page.

3.3.2. User interface

The HazeViz web browser page has input text box for user to enter the date and request to view past haze conditions or to forecast future haze condition by clicking the *Find* or *Forecast* button, respectively. In response, the haze condition will be presented on a geographical map and the daily or weakly, historical or forecasted data is charted in a graph based on the user request. The map provides information on the severity of the haze at the location with a red or a green dot indicating alarming or non-alarming condition, respectively. The HazeViz input and output screen shot is shown in Figure 4.

Figure 4. HazeViz input and output screenshots

4. RESULT AND DISCUSSION

The HazeViz application incorporates the SARIMA forecasting models. Unit and integrated testing of the procedures were carried out to verify the SARIMA models have been properly implemented. The test results show that HazeViz is functional since it correctly extracted the required PMSD and correctly functioned to indicate the severity of haze on the map. Besides visualizing the historical data, HazeViz also reasonably forecasted the alarming PM10 concentration levels during the haze period.

While HazeViz has been successfully deployed on the web, the results show that its underlying SARIMA models have mean absolute percentage error (MAPE) ranging between 35 percent and 45 percent, implying that the selected models are still far from robust.

We decided to use SARIMA in this study because it is commonly used when data has seasonal patterns. However, our models did not perform as anticipated due to what we think is idiosyncrasies in haze data. Past observations indicate that occassional PM10 peaks occurred during different times of year. The models failed to capture these occasional spikes. It appears that SARIMA has problem forecasting irregular spikes.

Moreover, the data quality issue is a common problem for environmental cases. For example, in our study, some station data series (e.g. Batu Muda and Petaling Jaya) have many missing values. We attempted to replace the missing values using the MTB method, which may not seem adequate. Consequently, the MAPE forecasting errors for these stations are higher than the rest.

Another shortcoming in the model is the seasonal differencing value used. The observed seasonal pattern of the haze series is around 90 days. However, we encountered problems creating models with such high seasonal indices. Through trial and error testing, we settled for a shorter seasonal differencing of 60 days. Even though it turned out the best during the trials, such a characterisation of the setting does not reflect the reality of the haze. Future work can look into these deficiencies and also consider augmenting the SARIMA models to reduce the forecast errors. SARIMA-ANN appears to be a promising model to explore.

5. CONCLUSION

An intelligent, web-based application called HazeViz has been successfully developed to forecast the PM10 value and visualize the haze condition on a map. The application can also visualize the historical PM10 data using graph and chart.

The SARIMA models underlying HazeViz needs to be improvised to better forecast the haze condition. Further research to improve the models are currently being conducted. At the moment, HazeViz covers six air quality monitoring stations in Klang Valley. The application can be extended to include other stations in Malaysia so that more people can benefit from it. Further, integrating the HazeViz functionality with Weather and Calendar apps, and with mobile navigation systems such as Waze, are seen as practical ways to deploy the application on a larger scale.

ACKNOWLEDGEMENT

The authors thankfully acknowledge Universiti Teknologi MARA for support of this work, which was funded under the Ministry of Education's Fundamental Research Grant Scheme (ref. no. FRGS/1/2016/ICT01/UITM/02/3). The authors thank Fusionex for giving access to the Fusionex GIANT platform, and the Department of Environment, Malaysia for the PM10 data used in this study.

REFERENCES

[1] What causes South East Asia's haze?, BBC News, 2017. http://www.bbc.com/news/world-asia-34265922.
[2] Forest Fires in Indonesia Choke Much of South-East Asia, The Guardian, 2015. http://www.theguardian.com/environment/2015/oct/05/forest-fires-in-indonesia-choke-much-of-south-east-asia
[3] Air Quality and Haze Episodes in Malaysia, Stakeholder Consultation Workshop, 2016. http://haze.akademisains.gov.my/download/WG1.120516%20(8am)%20WG1%20Air%20Quality%20&%20Haze%20Episode%20in%20Malaysia.pdf
[4] Air Pollution Index, Wikipedia, 2017. https://en.wikipedia.org/wiki/Air_Pollution_Index
[5] N. Manan, R. Hod, M. Sahani, H. Yusoff, R. Ismail and W. Mahiyuddin, "The Impact of Air Pollution and Haze on Hospital Admissions for Cardiovascular and Respiratory Diseases," *International Journal of Public Health Research*, vol. 6, no. 1, pp. 707-712, 2016.
[6] New Malaysia Ambient Air Quality Standard, DOE, 2013. https://www.doe.gov.my/portalv1/wp-content/uploads/2013/01/Air-Quality-Standard-BI.pdf
[7] Fusionex, 2017. http://fusionexgiant.com/page/library
[8] J. Wu, P. Zhang, H. Yi and Z. Qin, "What Causes Haze Pollution? An Empirical Study of PM2.5 Concentrations in Chinese Cities," *Sustainability*, vol. 8, pp. 1-14, 2016.
[9] N. T. K. Oanh and K. Leelasakultum, "Analysis of Meteorology and Emission in Haze Episode Prevalence over Mountain-Bounded Region for Early Warning," *Science of the Total Environment*, vol. 409, pp. 2261-2271, 2011.
[10] L. Juneng, M. T. Latif, F. T. Tangang and H. Mansor, "Spatio-Temporal Characteristics of PM10 Concentration Across Malaysia," *Atmospheric Environment*, vol. 43, no. 30, pp. 4584-4594, 2009.
[11] N. Shaadan, A. A. Jemain, M. T. Latif and S. M. Deni, "Anomaly detection and assessment of PM10 functional data at several locations in the Klang Valley," *Atmospheric Pollution Research*, vol. 6, no. 2, pp. 365-375, 2015.
[12] H. A. Hamid, A. S. Yahaya, N. A Ramli, A. Z. Ul-Saufie and M. N. Yasin, "Short Term Prediction of PM10 Concentrations Using Seasonal Time Series Analysis," *MATEC Web Conferences*, vol. 47, 2016.
[13] M. H. Lee, N. H. A.Rahman, L. Suhartono, M. T. Latif, M. E. Nor and N. A. B. Kamisan, "Seasonal ARIMA for Forecasting Air Pollution Index: A Case Study," *American Journal of Applied Sciences*, vol. 9, no. 4, pp. 570-578, 2012.
[14] L. Siew, L. Chin and M. Wee, "ARIMA and Integrated ARFIMA Models for Forecasting Air Pollution Index in Shah Alam, Selangor," *The Malaysian Journal of Analytical Sciences*, vol. 12, no. 1, pp. 257-263, 2008.
[15] C. E. Sapp, "Preparing and architecting for machine learning," Gartner Technical Professional Advice, 2017.
[16] N. Shaadan, S. M. Deni and A. A. Jemain, "Assessing and Comparing PM10 Pollutant Behaviour using Functional Data Approach," *Sains Malaysian*, vol. 41, no. 11, pp. 1335-1344, 2012.
[17] M. Lazim, "Introductory Business Forecasting a practical approach," Universiti Teknologi MARA, 2011.