A combination index measurement in forecasting daily air pollutant index

Conference Paper *in* AIP Conference Proceedings · August 2019

DOI: 10.1063/1.5121130

2 authors:

Nur Haizum Abd Rahman
Universiti Putra Malaysia

Muhammad Hisyam Lee
Universiti Teknologi Malaysia

A combination index measurement in forecasting daily air pollutant index

Cite as: AIP Conference Proceedings **2138**, 050025 (2019); https://doi.org/10.1063/1.5121130
Published Online: 21 August 2019

Nur Haizum Abd Rahman, and Muhammad Hisyam Lee

A Combination Index Measurement in Forecasting Daily Air Pollutant Index

Nur Haizum Abd Rahman[1, a)] and Muhammad Hisyam Lee[2, b)]

[1]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*
[2]*Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia.*

[a)]Corresponding author: nurhaizum_ar@upm.edu.my
[b)]mhl@utm.my

Abstract. Error magnitude is a measurement commonly used in forecast evaluation. However, the purpose of forecasting air quality is to maintain the air quality within assigned guidelines. Thus, the index measurement is important to be considered. But, the problem arises when the index is used to gauge the values of different offices and these measurements are found to be degenerate in commonly occurring situations. Therefore, this study aims to overcome both of the limitations. The daily air pollutant index (API) data from year 2005 to 2011 was used to compare the forecast performance between Box-Jenkins methods, artificial neural networks (ANN) and hybrid method. The forecast accuracy measurements used include mean absolute percentage error (MAPE), root mean squared error (RMSE), mean absolute deviation (MAD), true predicted rate (TPR), false positive rate (FPR), false alarm rate (FAR) and successful index (SI) including the proposed index measurement, combination index (CI). It is found that the index measurement enhance the ability to measure the air quality forecast performance in choosing the best forecast method with CI significantly overcome the limitation of existing index measurement. Thus, this study suggests to use the appropriate measurement in accordance to the purpose of forecasting.

INTRODUCTION

One important aspect in forecasting is the evaluation of model performance. At the end of the forecast, the appropriate model's validation is required in accordance to the purpose of forecasting. Hence, the validation of forecast message is useful to the public, government, and authorities. For instance, many measures of forecasting performance have been proposed in the past with recommendation to suit the model's comparison based on time series data. This is because different conclusion can be drawn from the results. Two types of measurements will be used in this study, known as error magnitude measurements and benchmark quality index measurements.

The error magnitude measurements such as mean absolute deviation (MAD), mean absolute percentage error (MAPE) and root mean square error (RMSE) have been broadly used to evaluate forecast precision [1, 2]. However, the measurements, have owned disadvantage as describe by Hyndman and Koehler [1]. MAD and RMSE are not appropriate to evaluate forecast accuracy across different time series since both are scale dependent. Meanwhile, MAPE gives a larger penalty to over forecast compared with under forecast [3, 1].Therefore, to overcome the limitation, they proposed mean absolute scaled error (MASE).

Besides error magnitude measurements, several studies have used the benchmark quality in the model's validation [3, 4, 5]. Hence, forecast accuracy based on a threshold value, known as truely predicted rate (TPR), false positive rate (FPR), false alarm rate (FAR) and successful index (SI) imperatively considered in forecast validation. In the field of environment, the benchmark quality is important to ensure the environment remain satisfactory especially to the public according to the assigned guidelines. Thus, benchmark quality will be used as the model's validation in determining the best forecast model alongside with the common forecast model's validation, error magnitude measurements. Both group measurements will be used in real air quality data.

The key necessity for human health and well-being is clean air. However, air pollution continues to pose a major

The 4th Innovation and Analytics Conference & Exhibition (IACE 2019)
AIP Conf. Proc. 2138, 050025-1–050025-8; https://doi.org/10.1063/1.5121130
Published by AIP Publishing. 978-0-7354-1881-3/$30.00

threat to health globally. The presence of the globalized development for both developed and developing countries has increase air pollution problems [6]. Thus, the air pollution has been affecting human health, the environment and in the long run, it can exacerbate risks to the earth due to increasing global warming and the greenhouse effect [7, 8]. To reduce the worst impacts of air pollution, air quality guidelines are designed. One of the simplest developed air quality guidelines is measured based on an index. Thus, the ambient air quality measurement in Malaysia was described in terms of Air Pollutant Index (API).

Over decades, the development of numerical models that can simulate the air quality has increasingly become a major focus of the researchers. These involved many types of forecasting methods including classical, modern, advanced or combination procedures. As an example, the linear method introduced by Box-Jenkins in 1976 is among the widely important classical time series models [9, 10]. In previous research done by Wang and Lu [11], Ibrahim et al. [12] and Kurt and Oktay [13], the performance of Box-Jenkins, or also referred to as autoregressive integrated moving average (ARIMA) model, gives satisfactory results in forecasting the daily pollutants such as carbon dioxide (CO_2), ozone (O_3) and nitrogen dioxide (NO_2). Besides that, modern methods such as artificial neural networks (ANNs) have been used for forecasting in many years with a wide range of pollutants with impressive results. These methods have proven to be the better approaches compared to the linear methods [14, 15, 16].

In addition, past research in an advanced method, hybrid often leads to the improvement in the forecasting accuracy [17, 18]. Chelani and Devotta [19] and Wang and Lu [11] are among the earliest studies that look into hybrid methodology in air quality applications. Chelani and Devotta [19] used a hybrid between ARIMA and nonlinear dynamical modelling in forecasting the NO_2 concentrations in Delhi. On the other hand, Wang and Lu [11] used multi-layer perceptron (MLP) trained with a particle swarm optimization algorithm (MLP-PSO) and a hybrid Monte Carlo (HMC) method. This was applied for ground level O_3 forecasting in Hong Kong. The paper is worded as follows. Section 2, materials and methods gives a simple introduction to the study city and data set. The forecasting techniques, error measurements including the proposed method also include in this section.

MATERIALS AND METHODS

THE DATA SET

The API scale and terms based on human health implications have been used for several years [20]. The API in Malaysia was developed based on the API introduced by the United State Environmental Protection Agency (USEPA). It is determined by the calculation of sub-indexes from five main pollutants, particulate matter (PM_{10}), ozone (O_3), carbon dioxide (CO_2), sulphur dioxide (SO_2) and nitrogen dioxide (NO_2) [21]. This method can be used by government agencies to characterize the status of air quality since API providing an easy evaluation and management at the given location throughout the countries [22]. Hence, it helps general public to easily understand the air quality status for their own health precaution. The API scales and terms are used in assessing and describing the air quality status on human health. This information with different ranges reflected as "Good (0-50), Moderate (51-100), Unhealthy (101-200), Very Unhealthy (201-300) and Hazardous (301 and above)" can be the benchmark of air quality management or data interpretation in decision-making processes [23].

The chosen station located in Johor Bahru city (N01°29.815, E103°43.617). Johor Bahru is the capital of Johor state, was selected for this study because this city is the second largest metropolitan area in Malaysia after the capital city, Kuala Lumpur [24]. Apart from that, it is home to a large number of the region's industries, residential, and commercial hotspots, and therefore, has congested roads. The study used daily data set for seven years, which covered the period from year 2005 until 2011. The data were divided into two data sets: (1) a training data set of year 2005 until 2010 (2191 observations) to identify the API model, and (2) a test data set for year 2011 with a total of 365 observations to check the performance of the model. The data plot is shown in Fig 1.

HYBRID SEASONAL ARIMA WITH ARTIFICIAL NEURAL NETWORK

The objective of hybrid method in time series forecasting is to make use the unique feature of the single methods. As such, to capture different patterns in the data. The basic idea is to hybrid between the linear pattern and nonlinear pattern data. This is because the method used for each pattern, has limited capability depending on the characteristics of the data [9, 25]. It is known that the linear methods are only capable of detecting and modeling linear data pattern, meanwhile nonlinear methods are for nonlinear data pattern. Therefore, with this capability, both linear and nonlinear time series models will be combined to achieve better forecasting performance [9, 26].

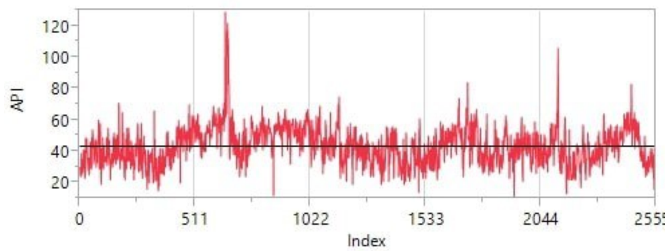

FIGURE 1. Air pollutant index (API) recorded at the sampling stations from year 2005 until 2011

A hybrid time series model which comprising a linear and a non-linear component has been employed in the experiments and can be written as [9]:

$$Y_t = L_t + N_t \tag{1}$$

where the components are, L_t is linear and N_t is nonlinear.

The hybrid model consists of two primary steps. Initially, the data will be analyzed by using the linear model. For this study, the linear model is estimated using seasonal autoregressive integrated moving average (SARIMA) method. From the SARIMA model, the linear part of hybridization is obtained. Then, by using the residual of a linear model, the nonlinear model will be discovered [9]. For non-linear component, the artificial neural networks are used. From this approach, the hybrid model has advantage because it exploiting the strength of both components.

Linear model, the Box-Jenkins or also known as ARIMA methods is capable in presenting both stationary and non-stationary time series. Most researchers have used this model to forecast univariate time series data. This method involve three main steps; (a) tentative identification, (b) parameter estimation, and (c) diagnostic checking. This method is flexible in representing different types of time series; autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and ARIMA. In the case where seasonal components are included in the model, the ARIMA model is called as SARIMA model and abbreviated as SARIMA . Generally, the SARIMA model can be written as:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B)\varepsilon_t \tag{2}$$

where $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$, $\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \ldots - \phi_p B^P$, $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^q$ and $\Theta_q(B) = 1 - \Theta_1 B - \theta_2 B^2 - \ldots - \Theta_p B^Q$. B is denoted as the backward shift operator, d and D are denoted as the non-seasonal and seasonal orders of difference respectively.

For non-linear patterns, the ANNs are used. ANNs are a branch of artificial intelligence that were firstly developed in the 1950s [27]. This method aim to imitate the biological brain architecture. Thus, progressive growth of ANN has been reported in the literature, which is mainly interested in developing convenient hardware for fast data analysis and information processing [28]. The significant advantage of the ANNs model is that no prior assumption of the model form is required in the model building process because the network model is highly determined by the characteristics of the data. Many different ANN models have been proposed however, the most influential ANNs used is MLP [29].

The ANNs are characterized by three simple layers of processing units which are the input layer, the hidden layer, and the output layer. Each layer are connected to each other and consists of processing elements (PEs), called neurons. The processes start when the input layer receives the external information. This is known as the input nodes correspond to the number of variables or the number of lagged observations that are used to discover the underlying pattern in the time series Zhang et al. [30]. The last or the highest layer is the output layer, where the problem solution is obtained. The input and output layers are separated by one or more intermediate layers called as the hidden layer. Most authors used only one or two hidden layers for forecasting purposes. Moreover, hidden nodes are used to process the information received from the input nodes to perform a nonlinear mapping between input and output. Without hidden nodes in the hidden layer, the ANN is equivalent to a linear statistical time series forecasting model. In addition, the hidden layer is included in the neural network system to increase the flexibility of the model. Thus, arrows from the input to the hidden layer and hidden layer to the output layer indicate the strength of each connection and can be measured by a quantity called weight.

The MLP relationship between the output, y_t and the inputs, $y_{t-1}, y_{t-2}, \ldots, y_{t-n}$ has the following mathematical representation:

$$y_t = w_0 + \sum_{i=1}^{n} w_i \, y_i \tag{3}$$

$$y_t = w_0 + \sum_{j=1}^{q} w_j \, g\left(w_{0.j} + \sum_{i=1}^{p} w_{ij} \, y_{t-i} + \varepsilon_t\right) \tag{4}$$

where w_j (j = 1, 2, .., q) and $w_{i\,j}$ (i = 1, 2, ..., p; j = 1, 2, ..., q) are the model parameters that are often called as the connection weights; p is the number of input nodes and q is the number of hidden nodes. Commonly, in the hidden layer, the activation function used is the logistic function, which is $f(x) = 1/(1 - exp^{-x})$,, meanwhile the linear function, $f(x) = x$, is used at the output stage.

FORECAST ACCURACY MEASUREMENT

The forecast data that have been produced by forecasting models need to be evaluated by comparing with the corresponding out-sample data. The out-sample data is denoted as $Y_t = y_{t+1}, y_{t+2}, \dots, y_{t+n}$ where is the number of forecast data.

Error magnitude measurements

This measurement measures the difference between forecast values with actual values. The measurements of forecasting accuracy are normally based on the mean absolute percentage error (MAPE), the root mean squared error (RMSE), and mean absolute deviation (MAD). The MAPE, RMSE and MAD can be written as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{(y_t - \hat{y}_t)}{y_t}\right|(100); y_t \neq 0 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \tag{6}$$

$$MAD = \frac{1}{n}\sum_{t=1}^{n} |y_t - \hat{y}_t| \tag{7}$$

where y_t is an actual value, $\hat{y}_t$ is a predicted value, and n is the number of predicted value. MAPE, RMSE, and MAD can be computed directly from the actual and forecast values and are not involved with unknown parameters that need to be estimated in which these statistical computations are widely applied in the literature [31]. Thus, all the statistical computations measure the residual errors with the smallest value and are chosen as the best model to be used in forecasting.

Index measurements

Index measurements used in this study are the incorrect forecast error based on the benchmark quality, which are also known as the exceedance measures. The binary system is used to determine and compare the occurrence of forecast and observed data, whether the occurrence exceeds the threshold value assigned [32]. Furthermore, definitions and values of statistical indexes related to the model's ability to forecast reliably the exceedances of the threshold value can be defined as A, the number of occurrences in the forecast and observed, B, the number of occurrences only in observed, C, the number of occurrences only in the forecast and D, the number of non-occurrence in the forecast and observed. From the count number of A, B, C and D, the statistical indexes can be defined as a truly predicted rate (TPR), falsely predicted rate (FPR), false alarm rate (FAR) and success index (SI). The TPR, FPR, FAR and SI can be computed as follows:

$$TPR = \frac{A}{A + B} \tag{8}$$

$$FPR = \frac{C}{C + D} \tag{9}$$

$$FAR = \frac{C}{C + A} \tag{10}$$

$$SI = \frac{A + D}{A + B + C + D} \tag{11}$$

All these indices have values ranging from 0 to 1. For instance, TPR represents the fraction between correct occurrences over a total of occurrences in observed, where the perfect score will be equal to 1 with the fraction of unexpected occurrences is given by ($1 - TPR$) or $B/A + B$. FPR represents the occurrences only in the forecast over total non-occurrences in observed while FAR represents the occurrences only in the forecast over total occurrences in the forecast; thus, the perfect score will be equal to 0. Finally, SI represents the correct forecast of occurrences and non-occurrences over the total number of forecasts, with 1 as the perfect score.

In some situation, these measures have the disadvantage of being infinite or undefined if the denominator is equal to 0. This infinite number reflects that no boundaries or limits are impossible to measure or calculate. In past research, the measurements remain infinite [5, 27]. However, based on Eq (8) to Eq (11) presented previously, the values are based on the binary system. This is due to the number of data presenting A, B, C and D in the equation based on the ratio between numerator and denominator. Thus, the infinite has a limit value, which is bounded between 0 and 1.

According to Schaefer (1990), problems arise in indices measurements when finding the optimal result where indices are used to measure the values from different groups. The groups are:

Group 1: False group with the perfect score of zero.

Group 2: True group with the perfect score of one.

By combining different index measurements, different aspect of the underlying conclusions can be captured. Thus, from hybrid methodology presented previously, the combination index measurement between group 1 and group 2 can be proposed. Let the first group, which is Group 1 to represent the false performance which are FPR and FAR be the negative value and Group 2 to represent the true performance which are TPR and SI as the positive value.

Thus, as a solution new combination index (CI) based on the accuracy obtained from Eq (8) to Eq (11) was proposed and can be presented as follows:

$$CI = TPR - FPR - FAR + SI \tag{12}$$

From the Eq (12), the interval of CI is given by the following conditions.

Condition 1: Set the worst combination from both groups. Group 1 is set as one and Group 2 is set as zero.

Condition 2: Set the best combination from both groups. Group 1 is set as zero while Group 2 is set as one.

From condition one, the FPR and FAR, the worst performance is one, while the TPR and SI, the worst performance is zero. Thus, from Eq (12) the CI becomes negative two. From condition two, the FPR and FAR, the best performance is zero, while the TPR and SI, the best performance is one. Thus, the CI becomes the positive of two. Hence, the range of CI is [-2,2]. Therefore, the perfect score for combination index measurement is positive two while the worst combination index measurement is negative two. Negative index indicates false negatives where the number of exceedances occur in observed but not in predicted while positive index indicates false positive as the observed not exceed the threshold value. In addition, the index has the value zero whenever a test gives the same proportion of positive and negative groups.

RESULTS AND DISCUSSION

Classical method, SARIMA, the modern method, ANNs and the advance method, hybrid method have been used in this study to forecast daily air pollutant index (API). For classical method, the SARIMA model was implemented via MINITAB version 16 and SAS software. Before ascertaining the tentative model, identification of the differencing process is necessary for both non-seasonal and seasonal parts to obtain the stationary data set before model development can be undertaken. This is because, based on Fig 1, the data is not stationary because of the existent of seasonality. As the data contain a seasonal pattern, the SARIMA was used. By taking the difference d = 1 for non-seasonal, D = 1 and S = 365 with for seasonal part, the data had become stationary series. Then, the tentative model based on the autocorrelation function (ACF) and partial autocorrelation function (PACF) from stationary series was determined to find the best combination order of SARIMA. By using the Ljung-Box statistics, the adequacy of the tentatively identified model was identified. The result is shown in Table 1 where the possible model were SARIMA(3,1,3)(0,1,1)^{365} and SARIMA(3,1,3)(2,1,0)^{365} with the best model, SARIMA(3,1,3)(0,1,1)^{365} and written as:

$$(1-\phi_1 B^1-\phi_2 B^2-\phi_3 B^3)(1-B)(1-B)^{365}Y_t=(1-\theta_1 B^1-\theta_2 B^2-\theta_3 B^3)(1-\Theta_1 B^{365})\varepsilon_t \tag{13}$$

TABLE 1. Error magnitude statistics for the validation of SARIMA, ANN, and hybrid models.

Models	MAPE	MAE	RMSE
SARIMA			
(3,1,3)(0,1,1)^{365}	**19.39**	**8.36**	**10.52**
(3,1,3)(2,1,0)^{365}	23.22	10.47	13.26
ANN			
All lags	**22.17**	**8.70**	**10.57**
Seasonal Lags 365,730	23.87	9.02	10.90
Lag 1 with Seasonal Lags ± 1	22.46	9.03	11.06
Hybrid			
All lags	**13.91**	**5.41**	**6.81**
Seasonal Lags 365,730	19.41	8.34	10.48
Lag 1 with Seasonal Lags ± 1	14.65	6.03	7.46

For ANN, the computations were performed using Matlab software. The data pre-processing was carried out first by normalizing the original data into the standard normal distribution with mean zero and variance equal to 1, making the range of the original data into [-1,1] and [0,1]. Next, the selected input, p was found through the ARIMA model [33, 34]. The possible lags are 1, 2, 3, 4, 365, 366, 367, 368 and 369. The single hidden layer was used as it is sufficient for any complex nonlinear function suggested by Zhang et al. [30]. The forecast performance result also given in Table 1.

Based on Table 1, SARIMA obtained the best performance in forecasting API in the selected stations compared to ANNs. As mentioned in the introduction, many results in the air quality literature find ANNs to have the best performance. However, our results show contrary findings to previous studies. With this mixed conclusion, the analysis is extended to forecast the API using hybrid methodology. The basic principles and modeling process of the ARIMA and ANN models were used. By using the residual from the best model SARIMA model, the ANN model was built. Based on Eq (1), the hybrid forecast was obtained and the results are presented in Table 1. The hybrid model gave the most accurate forecast and had been consistent in all four magnitude error measurements as shown in Table 1. The results are 13.91, 5.41 and 6.81 for MAPE, MAE and RMSE measurements respectively.

The purpose of forecasting the air quality aims to act as an early warning system for air quality control and management which to maintain the air quality within assigned guidelines. Thus, index measurement was carried out. The results using the index measurements were presented in Table 2 where it based on the best forecast performance for each model, SARIMA, ANN and hyrbid. Normally, the concentration of fine dust, PM_{10} is the highest compared to other pollutants and, therefore, PM_{10} value will determine the API levels. According to WHO and MAAQS, the limit value assigned is 50 μg/m^3 [35]. Thus, the API limit value of 50 was selected in order to verify the forecast quality of developed models.

The out sample data contain 365 data with 92 data exceeded and the remaining 273 did not exceed the assigned limit value. In Table 2, the ANNs are able to capture all the data that do not exceed the limit value. However, ANNs perform worst in capturing the exceedances in the observed data as shown in A. For SARIMA, the frequency in determining non-exceedances is higher than for hybrid model, but in determining the exceedance observed value, the hybrid model performs better. More importantly, the forecast obtained is used to build a warning system that is able to give prior notice to the public especially among the sensitive groups. Thus, the frequency of B, the occurrence in observed, but not predicted is needed to be aware as SARIMA and ANN recorded higher number compared to a hybrid. In other situation, the forecast value predicts the threshold value, although the observed does not. This can be called as a false alarm. Denoted by C, the hybrid shows that 19 days of false alarm occurred.

TABLE 2. Index measurements for the validation of SARIMA, ANN, and hybrid models.

Model	Frequency				Measurement				
	A	B	C	D	TPR	FPR	FAR	SI	CI
SARIMA	1	91	7	266	0.01	0.03	0.86	0.73	-0.16
ANN	0	92	0	273	0	0	0	0.75	0.75
Hybrid	46	46	19	254	0.50	0.07	0.29	0.82	0.96

For validation of the model, the performance statistics using the frequency information were defined. The hybrid model with the TPR values of 0.5 indicated that the model predicted exceeded successfully compared to SARIMA and ANN. In particular, the FPR with the best model equal to 0 is shown by the ANN following SARIMA and hybrid model. As mentioned earlier, this measurement could not be defined. This situation occurs when finding FAR for ANN where the values of the denominator are equal to zero. However, the value can be defined because it based on binary system. Thus, in Table 2, the FAR infinite result of ANN model is defined by 0 since the frequency of nominator is equal to 0. Hence, ANN is the best model followed by hybrid and SARIMA. In addition, Table 2 shows the SI measurement where the hybrid model gives the best result in identifying the capability of the prediction to capture the observed either exceeding or not exceeding the agreed threshold value. In order to support that the models can predict accurately the exceeds based on the assigned limit, the values of TPR and SI must be high while FPR and FAR must be low. Thus, the hybrid model is better in TPR and SI measurements, while ANN is better in FPR and FAR measurements. Besides, the problem arises when the index is used to gauge the values of different offices, true and false. Therefore, the combination of the index measurements presented here with hybrid gives the best forecast. Concerning both groups of forecast accuracy measurements, the hybrid model becomes the best-developed model to forecast the API values.

CONCLUSION

A detailed study of daily air pollutant index (API) in urban areas, Johor Bahru has been reported in this study. Based on the data recorded, the selected station shows seasonal data pattern with high API in the southwest monsoon (May to September) period due to biomass burning. Concerning the air pollution affecting the human health, forecasting using time series reported here is hoped to give an advanced notice to the public and the authorities. Therefore, the SARIMA, ANNs and hybrid method of SARIMA and ANNs were used in this study. To forecast the API with the highest accuracy, various accuracy measurements have been tested. This includes the measurements based on the magnitude of forecasted and observed and the model's validation based on the ability to forecast the exceedances of the assigned limit value. Comparing the best three models using magnitude error measurements (MAPE, MAE, RMSE), the hybrid model showed better skills in forecasting API compared to SARIMA model and ANNs model. Besides, comparing the three models based on index measurement, the hybrid model gives the best result in forecasting the capability to forecast these exceedances with a limit value of 50. As a conclusion, research on air quality concerning the assigned guidelines. Therefore, the chosen best forecast must consider the index measurement together with the statistical performance using the magnitude error.

ACKNOWLEDGMENTS

This study was supported by Universiti Putra Malaysia, Malaysia under Putra-IPM grant, 9587700. We would like to thanks the Department of Environment (DOE), Malaysia for providing air pollutants data.

REFERENCES

1. R. J. Hyndman and A. B. Koehler, International Journal of Forecasting **22**, 679 – 688 (2006).
2. J. McKenzie, Economics Letters **113**, 259 – 262 (2011).
3. J. Armstrong and F. Collopy, International Journal of Forecasting **8**, 69 – 80 (1992).

4. U. Schlink, S. Dorling, E. Pelikan, G. Nunnari, G. Cawley, H. Junninen, A. Greig, R. Foxall, K. Eben, T. Chatterton, J. Vondracek, M. Richter, M. Dostal, L. Bertucco, M. Kolehmainen, and M. Doyle, Atmospheric Environment **37**, 3237 – 3253 (2003).
5. A.-L. Dutot, J. Rynkiewicz, F. E. Steiner, and J. Rude, Environmental Modelling & Software **22**, 1261 – 1269 (2007).
6. H. do Nascimento Camelo, P. S. Lucio, J. B. V. L. Junior, P. C. M. de Carvalho, and D. von Glehn dos Santos, Energy **151**, 347 – 357 (2018).
7. J.-S. Heo and D.-S. Kim, Science of The Total Environment **325**, 221 – 237 (2004).
8. U. Kumar and V. K. Jain, Stochastic Environmental Research and Risk Assessment **24**, 751–760 (2010).
9. G. Zhang, Neurocomputing **50**, 159 – 175 (2003).
10. S. Suhartono, *Journal of Mathematics and Statistics*, **7**, 20–2701 (2011).
11. X.-K. Wang and L. Wei-Zhen, *Chemosphere*, **63**, 1261–7206 (2006).
12. M. Z. Ibrahim, R. Zailan, M. Ismail, and M. S. Lola, American Journal of Environmental Sciences **5**, 625–632 (2009).
13. A. Kurt and A. B. Oktay, Expert Systems with Applications **37**, 7986 – 7992 (2010).
14. M. Gardner and S. Dorling, Atmospheric Environment **32**, 2627 – 2636 (1998).
15. U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, and S. Vitabile, Atmospheric Environment **41**, 2967 – 2995 (2007).
16. J. Wang, H. Li, and H. Lu, Applied Soft Computing **71**, 783 – 799 (2018).
17. J. M. Bates and C. W. J. Granger (Cambridge University Press, New York, NY, USA, 2001) Chap. The Combination of Forecasts,, pp. 391–410.
18. S. Shen, G. Li, and H. Song, Annals of Tourism Research **38**, 72 – 89 (2011).
19. A. B. Chelani and S. Devotta, Atmospheric Environment **40**, 1774 – 1780 (2006).
20. A. Kumar and P. Goyal, Science of The Total Environment **409**, 5517 – 5523 (2011).
21. D. of Environment, "Malaysia environmental quality report," (2011).
22. N. Sansuddin, N. A. Ramli, A. S. Yahaya, N. F. F. M. Yusof, N. A. Ghazali, and W. A. A. Madhoun, Environmental Monitoring and Assessment **180**, 573–588 (2011).
23. R. Afroz, M. N. Hassan, and N. A. Ibrahim, Environmental Research **92**, 71 – 77 (2003).
24. A. Rizzo and J. Glasson, Cities **29**, 417 – 427 (2012).
25. F.-M. Tseng, H.-C. Yu, and G.-H. Tzeng, Technological Forecasting and Social Change **69**, 71 – 87 (2002).
26. T. Taskaya-Temizel and M. C. Casey, Neural Networks **18**, 781 – 789 (2005).
27. K. P. Moustris, I. C. Ziomas, and A. G. Paliatsos, Water, Air, & Soil Pollution **209**, 29–43 (2010).
28. M. Khashei and M. Bijari, Expert Systems with Applications **37**, 479 – 489 (2010).
29. W. S. Sarle, "Neural networks and statistical models," in *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, Vol. 1994 (SAS Institute Inc., 1994), pp. 1–13.
30. G. Zhang, B. E. Patuwo, and M. Y. Hu, International Journal of Forecasting **14**, 35 – 62 (1998).
31. G. Elliott, A. Timmermann, and I. Komunjer, Review of Economic Studies **72**, 1107–1125 (2005).
32. R. van Aalst and F. de Leeuw, "National ozone forecasting systems and international data exchange in northwest europe," Tech. Rep. (European Environment Agency, 1997).
33. F. Julian and C. Chris, Journal of the Royal Statistical Society: Series C (Applied Statistics) **47**, 231–250, https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9876.00109 .
34. G. Zhang and M. Qi, European Journal of Operational Research **160**, 501 – 514 (2005), decision Support Systems in the Internet Age.
35. W. H. Organization, "Who air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment," Tech. Rep. (World Health Organization, 2006).