PROFILING AND FORECASTING AIR POLLUTANT INDEX FOR MALAYSIA

NUR HAIZUM BINTI ABD. RAHMAN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

FEBRUARY 2017

To my beloved husband and daughter

♣

Azle Bin Abd Ghalim
Eryna Nur Batrisya binti Azle

And my family

♣

Abd Rahman bin Ahamad
Nora binti Alias
Nur Hidayah binti Abd Rahman

ACKNOWLEDGEMENT

First and foremost, all praise belongs to Allah, the Almighty and the Benevolent for His blessing and guidance for giving me strength and inspiration to complete this project.

I would like to express my appreciation and gratitude to my supervisor, Prof. Dr. Muhammad Hisyam Lee for his guide and support throughout the accomplishment of this research. This research was unable to be completed as expected without him, since he helped in many forms throughout the process of completing this research. His formal and informal supervision, criticism, advice and guidance enabled me to produce the best result of this study.

Finally, special thank goes to my family members for their unconditional love and support and also to my lecturers and friends for the encouraging and boasting supports while I was in the process of completing my degree Doctor of Philosophy.

ABSTRACT

Detection of poor air quality is important to provide an early warning system for air quality control and management. Thus, air pollutant index (API) is designed as a referential parameter in describing air pollution levels to provide information to enhance public awareness. This study aims to study API trend, time series forecasting methods, their performance evaluations and missing values effect for accurate early warning system using several approaches. First, a calendar grid visualization is introduced to effectively display API daily profiling for the whole of Malaysia in identifying the exact point of poor air quality. Second, comparisons between classical and modern forecasting methods, artificial neural network (ANN), fuzzy time series (FTS) and hybrid are carried out to identify the best model in Johor sampling stations; industrial, urban and suburban. Third, due to the issue of different perfect score in existing index measurement to evaluate forecast performance, a combination index measures is proposed alongside error magnitude measurement. Fourth, decomposition and spatial techniques are compared to find the effect of high accuracy imputations in API missing values. The finding presented that the air quality trend across the day, week, month and year are more significant due to the daily arrangement in the calendar grid visualization. The ANN model gives the best forecasting model of API for industrial and urban area while the hybrid model provide the best forecasting for suburban area. The forecasting performance for industrial and urban areas improve between 14% to 20% and 20% to 55% in error magnitude and index measurements, respectively when high accuracy missing values imputation is conducted. In conclusion, the profiling using calendar grid visualization is useful to guide the control actions of early warning system. Forecasting using modern methods give promising result in API and the improvements in measurements will assist in choosing the best forecasting method. Missing values imputation in data series can enhance the forecasting performance.

ABSTRAK

Pengesanan kualiti udara tidak bermutu penting bagi menyediakan sistem amaran awal untuk kawalan dan pengurusan kualiti udara. Maka, indeks pencemaran udara (IPU) direka sebagai parameter rujukan dalam menggambarkan tahap pencemaran udara bagi memberikan maklumat untuk meningkatkan kesedaran umum. Kajian ini bertujuan untuk mengkaji trend IPU, kaedah siri masa ramalan, penilaian prestasi dan kesan data hilang bagi sistem amaran awal yang tepat dengan menggunakan beberapa pendekatan. Pertama, pengvisualan grid kalendar diperkenalkan untuk memaparkan secara efektif profil IPU harian di seluruh Malaysia dalam mengenalpasti titik tepat kualiti udara tidak bermutu. Kedua, perbandingan antara kaedah ramalan klasik dan moden, rangkaian neural buatan (ANN), siri masa kabur (FTS) dan hibrid dijalankan untuk mengenalpasti model yang terbaik di stesen pensampelan Johor; industri, bandar dan pinggir bandar. Ketiga, disebabkan isu perbezaan skor sempurna bagi ukuran indeks sedia ada untuk menilai prestasi ramalan, gabungan ukuran indeks dicadangkan bersama dengan ukuran ralat magnitud. Keempat, teknik penguraian dan ruang dibandingkan untuk mencari kesan ketepatan imputasi yang tinggi dalam data hilang IPU. Dapatan menunjukkan trend kualiti udara bagi harian, mingguan, bulanan dan tahunan lebih signifikan disebabkan aturan harian dalam pengvisualan grid kalendar. Model ANN memberikan model ramalan IPU yang terbaik di kawasan industri dan bandar manakala model hibrid menyediakan ramalan terbaik di kawasan pinggir bandar. Prestasi ramalan di kawasan industri dan bandar bertambah baik antara 14% hingga 20%, dan 20% hingga 55% bagi ralat magnitud dan ukuran indeks apabila ketepatan imputasi data hilang yang tinggi dijalankan. Kesimpulannya, pemprofilan dengan menggunakan pengvisualan grid kalendar adalah berguna sebagai panduan untuk tindakan kawalan bagi sistem amaran awal. Ramalan menggunakan kaedah moden memberikan hasil yang menggalakkan bagi IPU dan penambahbaikan dalam pengukuran akan membantu untuk memilih kaedah ramalan yang terbaik. Imputasi data hilang bagi siri data boleh meningkatkan prestasi ramalan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DOE	-	Department of Environment
API	-	Air Pollutant Index
USEPA	-	United State Environmental Protection Agency
PM_{10}	-	Particulate matter 10 microns diameter
O_3	-	Ozone
CO	-	Carbon monoxide
SO_2	-	Sulphur dioxide
NO_2	-	Nitrogen dioxide
MAPE	-	Mean absolute percentage error
MAD	-	Mean absolute deviation
RMSE	-	Root mean square error
TPR	-	Truely predicted rate
FPR	-	False positive rate
FAR	-	False alarm rate
SI	-	Successful index
FTS	-	Fuzzy time series
ANN	-	Artificial neural network
MSE	-	Mean square error
CIM	-	Combination index measurement
AQI	-	Air Quality Index
US	-	United State
AQHI	-	Air quality health index
CAI	-	Comprehensive Air Quality Index
PSI	-	Pollutant Standard Index
EHRs	-	Electronic health records
ARIMA	-	Autoregressive integrated moving average model
AR	-	Autoregressive

MA	-	Moving average
SARIMA	-	Seasonal autoregressive integrated moving average model
FLR	-	Fuzzy logical relationship
I	-	Integrated
MLP	-	Multilayer perceptron
ARFIMA	-	Autoregressive fraction integrated moving average
ACF	-	Autocorrelation function
PACF	-	Partial autocorrelation function
FFNN	-	Feed-forward neural network
FLRG	-	Fuzzy logical relation group
AA	-	Arithmetic average
ID	-	Inverse distance
CC	-	Correlation coefficients
NR	-	Normal ratio
NRM	-	Modified normal ratio based on correlation
MNR-T	-	Modified normal ratio with inverse distance

LIST OF SYMBOLS

$\hat{Y}_{t+1}$	-	Forecast Value at time $t+1$
t	-	Time
L_t	-	Estimate for the level factor of the time series at time t
T_t	-	Estimate for the trend factor of the time series at time t
S_t	-	Estimate for the seasonal factor of the time series at time t
α	-	Smoothing constant for the level
β	-	Smoothing constant for the trend
γ	-	Smoothing constant for the seasonal
ϕ_p	-	Non-seasonal autoregressive of order p
θ_q	-	Non-seasonal moving average of order q
Φ_P	-	Seasonal autoregressive of order P
Θ_Q	-	Seasonal moving average of order Q
B	-	Backshift operator
a_t	-	White noise
d	-	Non-seasonal differencing
D	-	Seasonal differencing
λ	-	Parameter in Box-Cox transformation
b_i	-	Bias
x_j	-	Inputs variable
n_i	-	i th neuron at hidden layer
$w_{i,j}$	-	Weight from inputs and i th neuron at hidden layer
γ_0	-	Output bias
γ_j	-	Weights from n_i to output

U	-	Universe of discourse
A_i	-	Fuzzy set of U
f_{A_i}	-	Membership function of A_i
$F(t)$	-	Fuzzy set defined on $Y(t)$
m	-	Seasonal period in fuzzy
m_{jk}	-	Midpoint
L_t	-	Linear component at time t
N_t	-	Non-linear component at time t
y_t	-	Sample observed value at time t
$\hat{y}_t$	-	Sample forecast value at time t
n	-	Number of data
A	-	The number of exceedances in observed and forecasted
B	-	The number of exceedances only in observed
C	-	The number of exceedances only in observed
D	-	The number of non-exceedances in observed and forecasted
C_t	-	Cyclic at time t
I_t	-	Irregular at time t
Y_t	-	Estimate value at target station
N	-	Number of neighboring station
W	-	Weight for ith neighboring station
d_{it}	-	Distance between the target station and neighboring stations
r_{it}	-	Correlation between the target station and neighboring stations
μ_t	-	Sample mean at target station
μ_i	-	Sample mean at neighboring station

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Pollution can take many forms, such as air pollution, water pollution, ground pollution and noise pollution. But, the fundamental pollution problem in many parts of the world is air pollution (Kurt and Oktay, 2010). Air pollution is a problem that is designated to have multiple spatial and temporal scales, which includes complex chemical and physical mechanism. It escalates as a consequence of human activity, and it is highly nonlinear as a problem (Karatzas et al., 2008). Frequently recurring situation of air pollution have substantial effect towards both social and economic managements (Caselli et al., 2009). Energy production from power plants, industrial processes, residential heating, fuel burning vehicles, and natural disasters are some factors that contribute to air pollution (Kurt and Oktay, 2010).

In Europe, reducing the exposure of air pollution still remains an important issue (World Resources Institute, 2002). Air conditions in some European countries have worsen substantially since the 1970s, which call for the improvement of air quality all over the region (Marco and Bo, 2013). However, since 1997, the measured concentrations of particulate matter and ozone in the air have not shown any significant improvements despite the decrease in emissions. The issue of air quality is now a major concern of most European citizens. Many European countries face this problem such as the United Kingdom, Greece and Italy with London being the most polluted city in Europe (Vidal, 2010).

Urban cities that are overwhelmed with industrial activities are mostly located in Asia, and particularly in China and India (CAI-Asia, 2010). China is well known as the world's fastest growing economy, and as a consequence of this economic growth, the quality of its air has deteriorated. The main factor that contributes to China's increasing air pollution is its extraordinary daily traffic. On the other hand, India suffers from an appalling air pollution because of its varying industrial wastes. Due to these activities, both China and India were recorded to have the worst conditions of air pollution in the world (Alles, 2009). In addition, four out of 10 cities with the worst air pollution in the world are located in India which are Gwalior, Allahabad, Patna and Raipur (Bhattacharya, 2016).

There are worldwide concerns over the consequences of air pollutant towards environment as its effects are diverse and numerous. The negative effects of air pollution are not only directed to human health, but also towards the forest, waters, and the ecosystem as a whole (Cisneros et al., 2010). The air we breathe everyday could be contaminated by polluting substances. For instance, PM_{10} a particulate matter with an aerodynamic diameter smaller than 10 µm, could cause nose and throat irritations that could lead to death (Caselli et al., 2009, Pope et al., 2002). Moreover, a study done in Italy, shows that ozone concentrations at ground levels modulate oxidative DNA damage in the circulating lymphocytes of residents in polluted areas (Palli et al., 2009). In addition, pollution caused by ozone, O_3 can decrease the lung function, and it has been reported to increase cardiopulmonary mortality and the risk of lung cancer (Ghazali et al., 2010, Pope et al., 2002). Furthermore, the negative effects of air pollution have increased the numbers of premature deaths, with the highest annual incidence to be noted in China (Platt, 2007).

The manifestation of haze in the atmosphere indicates the poor condition of the air. In Malaysia, a series of haze episodes were reported since the 1980s, beginning in the year 1983 and then in 1990, 1991, 1994 and in 1997 (Awang et al., 2000). The worst haze episode ever reported in Malaysia was in 1997 (Afroz et al., 2003, Lim et al., 2008) which was due to a large scale burning of forests in parts of Kalimantan and Sumatra as was apparent from satellite image (Awang et al., 2000). The winds has made it easier for the heavy haze to be transported, and as the result it reaches all over Southeast Asia namely Indonesia, Singapore, Brunei, and Malaysia.

Malaysia is divided into two main regions, namely Peninsular Malaysia and Malaysian Borneo. Peninsular Malaysia was divided into two areas, west coast states and east coast states. The west coast states are the most developed, and as a result they are the most polluted area in Malaysia. The Malaysian government has approved the building of industrial zones, particularly in forestland and uninhabited areas. This was due to the changes made by the government to shift Malaysia's industrial activities from agricultural to manufacturing and heavy industries (Afroz et al., 2003, Awang et al., 2000). The major development for manufacturing and heavy industries are mostly located in the industrial zone of Shah Alam, Selangor. As a consequence, it is now a heavily populated area and is considered as one of the most polluted areas in Malaysia.

The major contributor to Malaysia's worsened air quality are not only in heavy industries, but also vehicle emissions, as well as illegal open burning (Afroz et al., 2003). In the west coast, besides Selangor, cities with an unhealthy air quality were found in Kuala Lumpur, Penang, Perak, Negeri Sembilan, Johor and Melaka. The main cause of the unhealthy air quality in these states was due to the ground level ozone and PM_{10} as stated by Department of Environment in Malaysia Environment Quality Report (2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012). Between the Northern and the Southern regions, the Southern region which includes Melaka, Negeri Sembilan and Johor, was recorded as the most polluted due to the frequent episodes of unhealthy air quality. However, among these three states, Johor is the most severe as the monitoring stations have recorded more poor air quality days than good air quality days (Department of Environment, 2005).

The Department of Environment (DOE) is responsible for monitoring and managing Malaysia's air quality. Stations were built near industrial and residential areas to detect the significant changes in air quality that may harm human health and the environment. Since 2004 and until now, the DOE reports the environmental conditions in Malaysia in their report named Malaysia Environmental Quality Report, which covers all aspects of environmental quality in Malaysia and air quality is put in the primary pages of that report.

1.2 Background of Study

Clean air is considered a crucial necessity for human health and well-being. Thus, continuous air pollution pose a major threat to human health globally. The presence of globalized growth in developed and developing countries has contributed to the escalation of air pollution problems (Hassanzadeh et al., 2009). Besides being harmful to human health and the environment, in the long-term, these air pollution problems tend to damage the earth by contributing to the global warming and the greenhouse effect (Heo and Kim, 2004, Kumar and Jain, 2010, Kurt and Oktay, 2010). Therefore, it is important to monitor the air pollution in the atmosphere by providing guidance on effective control actions, especially in severe air quality conditions where greater forces are needed.

Southeast Asia, a sub-region of Asia, faces frequent air pollution problems. Human-based activities are the main contributor to air pollution, activities such as open burning activities, industrial processes and vehicle emission (Afroz et al., 2003, Kurt and Oktay, 2010, Wang and Lu, 2006b). Typically, air pollution in Southeast Asia becomes worsens in the dry season due to the heavy smokes of peatlands fires in Sumatra and the Kalimantan region of Borneo Islands (Heil and Goldammer, 2001). Thus, several countries in Southeast Asia, such as Brunei, Indonesia, Malaysia, Singapore and Southern Thailand, are still affected by the continuous haze crisis for several decades.

As mention earlier, among the earliest worst haze phenomenon in Southeast Asia was the one reported in 1997. Yet, that did not stop it from recurring continuously until today. The widespread haze causes a limited atmospheric visibility, and it inflicts serious health problems. In addition, high levels of air pollution will affect the economy by disrupting air travel, interrupting business activities, and increasing the expenditure on health care. As mention earlier, Malaysia is one of the most affected countries, and that is due to strong winds and dry weather that would carry the smog from Sumatra and affects the Peninsular Malaysia, while the smog from Kalimantan affects East Malaysia (Sastry, 2002). Thus, to identify the severity of air pollution, the ambient air quality measurement in Malaysia is described in terms of Air Pollutant Index (API).

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 the sub-indices of five main pollutants, namely particulate matter (PM_{10}), ozone (O_3), carbon monoxide (CO), sulphur dioxide (SO_2) and nitrogen dioxide (NO_2). Hence, the highest value among these sub-indexes is chosen as the API for the time in question. According to Malaysia's Department of the Environment (2004), different categories of sub-indices represent different effects on human health. These information, with different ranges, are reflected as "Good (0-50), Moderate (51-100), Unhealthy (101-200), Very Unhealthy (201-300) and Hazardous (301 and above)". These categories can be a benchmark for air quality management or data interpretation for decision making processes (Afroz et al., 2003).

The API scales and its terms are used in this study in order to measure the air quality, since the detection of poor air quality is important as an early warning system for air quality control and management. From the recorded API data, this study aims to build an API data profiling throughout Malaysia. The profiling will provide timely information of air quality conditions to the public, government officials, and administrative users. The profiling is developed in terms of graphics presentations. These graphics presentations described the data within the range, indicating different health status used to give visual information. Moreover, it is also a great instrument to highlight polluted areas and the time information period to improve the actions that should be taken.

Air quality forecasting is also important for the air pollution assessment and management (Lim et al., 2008). It can provide an early notice and a warning to individuals and communities, in order to help them in limiting the exposure, reduce asthma attacks, prevent the irritation of the eye, nose, and throat, avoid respiratory and cardiovascular problems, and save lives (Kampa and Castanas, 2008, Kumar and Goyal, 2011, Kurt and Oktay, 2010). Research in air quality forecasting has increased and has become an area of interest. However, dealing with air quality is not as easy since the recorded air quality are not physically produced nor manufactured. For this reason, forecasting accuracy should be periodically maintained by using statistical and mathematical tools to obtain the best forecast.

In order to find the best forecasting methods, the accuracy measurements play an important role in reaching the conclusion of any data analysis (Hyndman and Koehler, 2006, Willmott et al., 1985). In air quality, the measurements of error magnitude which analyses the difference between the observed and the predicted are usually used in forecast evaluations. Mean absolute percentage error (MAPE), mean absolute deviation (MAD) and root mean squared error (RMSE) are among the measurements that are commonly used to assess forecast accuracy (Armstrong and Collopy, 1992). However, accuracy in terms of error magnitude alone is not enough, especially in the field of air quality as it needs to relate with decision making. Thus, index measurement is also used, which aims to maintain the air quality within assigned guidelines (Moustris et al., 2010, Dutot et al., 2007, Schaefer, 1990).

Index measurement uses the benchmark quality in the model’s validation to ensure that the environment remain acceptable to the public (Armstrong and Collopy, 1992, Dutot et al., 2007, Schlink et al., 2003, Vautard et al., 2001). Thus, forecast accuracy based on threshold values, namely as truely predicted rate (TPR), false positive rate (FPR), false alarm rate (FAR) and successful index (SI) are taken into consideration for forecast validation. However, these measurements have some disadvantage where the obtained results are possible of getting infinite values. Besides, different perfect score could lead to different conclusion of the best model. The effect of missing values in forecasting are also determined to find the optimal forecast model for API data sets as the problem of missing values is common and unavoidable.

1.3 Problem Statement

Air quality data has been recorded in Malaysia since 1996 and the huge amount of data usually presented in the form of text information. Thus, air quality information are difficult to be reviewed, especially for the public understanding. Moreover, the public, especially those in high risk groups such as asthmatic individuals, children, and elderly, need to be alerted beforehand about the cases of poor air quality. Therefore, to implement air quality management and public warning strategies for pollution levels, a reasonably accurate forecasts of air quality is necessary. This can be achieved by using forecasting. Evaluation of performances are also important to find the best forecast performance. Thus, using the common error magnitude measurements is not enough to assess air quality. Index measurements are also important to evaluate the performance of air quality forecasting, because if the forecast fails to effectively predict poor air quality, it could cause a huge negative impact not only to the public health but also to the economy. Missing data is another problem that occurs when recording data due to many reasons such as instrument malfunction for a period of time. The results of air quality models and forecast could be influenced by considering the incomplete series of recorded data as an input in the analysis. Therefore, the estimations to replace missing values are always important in air quality studies.

The study will be focused on API data set with the following problems:

a. What is the profiling of Malaysia's API as a whole?

b. What is the most appropriate method to model and forecast the API data, the classical methods or modern methods?

c. What is the suitable criterion for evaluating and selecting the best model for air pollution data and subsequently improve the forecast evaluation?

d. Is there any effect of missing values toward forecasting performance?

1.4 Objectives of the Study

This study embarks on the following objective:

a. To develop the profiling for API for the whole of Malaysia by using visualization approach.

b. To improve the API forecasting by using time series method; either the classical methods or modern methods.

c. To enhance the API forecast performance evaluation by using index measurement together with magnitude measurement to find the best model.

d. To introduce the combination approach for index measurement to improve the forecast evaluations.

e. To apply decomposition and spatial method in missing values imputations to achieve high accuracy in API forecasting.

1.5 Significance of the Study

Air contamination remains as continuing area of interest, which concern the effects of poor air quality on the human health and the natural environments around the world. Therefore, poor air quality and anticipation approaches are important areas of the study, especially in developing countries like Malaysia. Thus, an early warning system is essential in order to take the control action.

Firstly, alerts could be made from the visualization of the results to give an overview of air quality in Malaysia. The trend of air quality can be easily identified especially by detecting the seasonality of the data set. Secondly, the air quality level could be monitored by forecasting. There is no clear proof to conclude which model performs best in all situations. Therefore, it is appropriate to apply forecasting competitions in order to find the best forecasting model (Athanasopoulos et al., 2011, Makridakis et al., 1993). The outcomes of the forecasting models in this study will cover the industrial, urban and suburban areas in order to provide information to predict the quality of the contaminated air. Hence, the harm to the public health and environment could be minimized.

The forecasting accuracy that has been discussed in this study could provide some basic guidelines. Therefore, the identification of the best model would be more sensible, particularly in the field of air quality. In addition, the changes in forecast performance with the presence of missing data is examined through a comparison study. This will provide a beneficial information guideline in deciding the appropriate model in forecasting whenever the historical data have missing data.

The profiling, the different forecasting methods, performance evaluations and imputation of missing data that are considered in this study will provide the information to analyse the air quality. Therefore, the practitioners will be able to compare between the methods discussed and choose the appropriate approach in relation to their context of the study. Finally, the study is crucial to assist the Department of Environment or any related agencies to take a quick action in preventing environmental deterioration. Consequently, the public will benefit from the study as the accuracy and the up to date information on air quality will provide prompt warnings for their daily activities.

1.6 Scope of the Study

This study used univariate data where the API data are obtained from the Malaysian Department of Environment. The data are from 52 sampling stations that located in Malaysia and they are accessible starting from the year 1996. For API profiling, all sampling stations are taken where the daily data are used from January 2005 to December 2011. Meanwhile, the data that are involved in the comparison of forecasting models performance include monthly API data and daily API data that are located in Johor. The Johor sampling stations consist of three different background, namely industrial, urban and suburban areas. The monthly and daily data used are from January 2000 to December 2009 and January 2005 to December 2011 respectively.

For forecasting comparison, the classical time series methods that were applied in the monthly data were Box-Jenkins method, time series regression method and winter's exponential smoothing method. For the daily data, only Box-Jenkins was used as the classical approach. The Box-Jenkins used in this study is based on seasonal autoregressive integrated moving average (SARIMA) method. The modern methods were also implemented as a comparison to the classical methods. Fuzzy time series (FTS) based on Chen's, Yu's and Cheng's methods, artificial neural network (ANN) and a hybrid method between Box-Jenkins and ANN were all used in both monthly and daily data.

The forecast accuracy of all these methods will be evaluated and compared by using the error magnitude measurements, namely mean absolute error (MAE), mean absolute percentage error (MAPE), mean square error (MSE) and root mean square error (RMSE). In addition, the index measurements were used, namely true predicted rate (TPR), false predicted rate (FPR), false alarm rate (FAR) and successful index (SI) includes the proposed combination index measurement (CIM). For the missing values imputation, the decomposition method and spatial interpolation weighting methods were used.

1.7 Thesis Structure

This thesis consists of five chapters. The first chapter, gave general information and a background of the study. The second chapter presents the literature review which encompasses Malaysia's air quality forecasting and modelling, forecasting methods that includes the classical and modern methods, forecast accuracy evaluations and imputation methods for missing values. Then, chapter three explains the methodology in detail, including the procedure implemented in the study. Next, the results of the study will be explained and discussed in chapter four. Finally, chapter five presents the conclusion, summary and recommendation for future studies.

REFERENCES

Afroz, R., Hassan, M. N. & Ibrahim, N. A. 2003. Review of air pollution and health impacts in Malaysia. *Environmental Research,* 92**,** 71-77.

Alles, D. L. 2009. Asian Air Pollution. Washington: Western Washington University

Armstrong, J. S. & Collopy, F. 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting,* 8**,** 69-80.

Athanasopoulos, G., Hyndman, R. J., Song, H. & Wu, D. C. 2011. The tourism forecasting competition. *International Journal of Forecasting,* 27**,** 822-844.

Awang, M. B., Jaafar, A. B., Abdullah, A. M., Ismail, M. B., Hassan, M. N., Abdullah, R., Johan, S. & Noor, H. 2000. Air quality in malaysia: Impacts, management issues and future challenges. *Respirology,* 5**,** 183-196.

Azmi, S. Z., Latif, M. T., Ismail, A. S., Juneng, L. & Jemain, A. A. 2010. Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. *Air Quality, Atmosphere, & Health,* 3**,** 53-64.

Bates, J. M. & Granger, C. W. J. 1969. The combination of forecasts. *OR,* 20**,** 451-468.

Bernard, F. 2003. Fuzzy environmental decision-making: Applications to air pollution. *Atmospheric Environment,* 37**,** 1865-1877.

Bernard, F. 2006. Fuzzy approaches to environmental decisions: Application to air quality. *Environmental Science & Policy,* 9**,** 22-31.

Bhattacharya, S. 2016. Which cities in the world have the worst air pollution? Available: http://www.wsj.com/news/us/ [Accessed June 17, 2016].

Bowden, G. J., Dandy, G. C. & Maier, H. R. 2003. Data transformation for neural network models in water resources applications. *Journal of Hydroinformatics,* 5**,** 245-258.

Bowerman, B. L., O'connell, R. & Koehler, A. B. 2005. *Forecasting, Time Series and Regression: An Applied Approach*, South-Western College Pub.

Box, G. E. P. & Jenkins, G. M. 1976. *Time series analysis: Forecasting and control,* San Fransisco, Holden-Day.

Brunelli, U., Piazza, V., Pignato, L., Sorbello, F. & Vitabile, S. 2007. Two-days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy. *Atmospheric Environment,* 41**,** 2967-2995.

Cai-Asia 2010. Air quality in asia: Status and trends. *In:* Center, C. a. I. F. a. C. C.-A. (ed.). Pasig City, Philippines: Clean Air Initiative for Asian Cities (CAI-Asia) Center.

Canada, E. a. C. C. 2015. *Air quality health index* [Online]. Available: http://www.ec.gc.ca/cas-aqhi/ [Accessed April 12 2016].

Cao, K., Zhu, Q., Iqbal, J. & Chan, J. W. C. 2007. A trend pattern assessment approach to microarray gene expression profiling data analysis. *Pattern Recognition Letters,* 28**,** 1472-1482.

Caselli, M., Trizio, L., De Gennaro, G. & Ielpo, P. 2009. A simple feedforward neural network for the PM10 forecasting: Comparison with a radial basis function network and a multivariate linear regression model. *Water, Air, and Soil Pollution,* 201**,** 365-377.

Chatfield, C. 1992. A commentary on error measures. *International Journal of Forecasting,* 8**,** 100-102.

Chelani, A. B. & Devotta, S. 2006. Air quality forecasting using a hybrid autoregressive and nonlinear model. *Atmospheric Environment,* 40**,** 1774-1780.

Chen, S.-M. 1996. Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems,* 81**,** 311-319.

Chen, S.-M. 2002. Forecasting enrollments based on high-order fuzzy time series. *Cybernetics and Systems,* 33**,** 1-16.

Cheng, C.-H., Chen, T.-L., Teoh, H. J. & Chiang, C.-H. 2008. Fuzzy time-series based on adaptive expectation model for TAIEX forecasting. *Expert Systems with Applications,* 34**,** 1126-1132.

Cisneros, R., Bytnerowicz, A., Schweizer, D., Zhong, S., Traina, S. & Bennett, D. H. 2010. Ozone, nitric acid, and ammonia air pollution is unhealthy for people and ecosystems in southern Sierra Nevada, California. *Environmental Pollution,* 158**,** 3261-3271.

Cryer, J. D. 1986. *Time series analysis* United States of America, Duxbury Press.

Dam, J.-W. V. & Velden, M. V. D. 2015. Online profiling and clustering of Facebook users. *Decision Support Systems,* 70**,** 60-72.

Department of Environment 2005. Malaysia Environment Quality Report 2004. Putrajaya: Department of Environment.

Department of Environment 2006. Malaysia Environment Quality Report 2005. Putrajaya: Department of Environment.

Department of Environment 2007. Malaysia Environment Quality Report 2006. Putrajaya: Department of Environment.

Department of Environment 2008. Malaysia Environment Quality Report 2007. Putrajaya: Department of Environment.

Department of Environment 2009. Malaysia Environment Quality Report 2008. Putrajaya: Department of Environment.

Department of Environment 2010. Malaysia Environment Quality Report 2009. Putrajaya: Department of Environment.

Department of Environment 2011. Malaysia Environment Quality Report 2010. Putrajaya: Department of Environment.

Department of Environment 2012. Malaysia Environment Quality Report 2011. Putrajaya: Department of Environment.

Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G. & Moncada-Herrera, J. A. 2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment,* 42**,** 8331-8340.

Dickey, D. A. Stationarity issues in time series models. SAS Conference Proceedings: SAS Users Group International 30, 10-13 April 2005 Philadelphia, Pennsylvania. SAS Institute Inc., Cary, NC, 1-17.

Dilla, W. N. & Raschke, R. L. 2015. Data visualization for fraud detection: Practice implications and a call for future research. *International Journal of Accounting Information Systems,* 16**,** 1-22.

Dorr, B. & Herbert, P. Data profiling: Designing the blueprint for improved data quality SAS Conference Proceedings: SAS Users Group International 30, 10-13 April 2005 Philadelphia, Pennsylvania. SAS Institute Inc., Cary, NC, 1-10.

Dutot, A.-L., Rynkiewicz, J., Steiner, F. E. & Rude, J. 2007. A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environmental Modelling & Software,* 22**,** 1261-1269.

Eischeid, J. K., Bruce Baker, C., Karl, T. R. & Diaz, H. F. 1995. The quality control of long-term climatological data using objective data analysis. *Journal of Applied Meteorology,* 34**,** 2787-2795.

Elliott, G., Komunjer, I. & Timmerman, A. 2005. Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies,* 72**,** 1107-1125.

Faraway, J. & Chatfield, C. 1998. Time series forecasting with neural networks: A comparative study using the air line data. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 47**,** 231-250.

Finley, J. P. 1884. Tornado predictions. *American Meteorological Journal,* 1**,** 85-88.

Gao, Y., Tian, G. Y., Li, K., Ji, J., Wang, P. & Wang, H. 2015. Multiple cracks detection and visualization using magnetic flux leakage and eddy current pulsed thermography. *Sensors and Actuators A: Physical,* 234**,** 269-281.

Gardner, M. W. & Dorling, S. R. 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment,* 32**,** 2627-2636.

Ghazali, N. A., Ramli, N. A., Yahaya, A. S., Yusof, N. F. F. M., Sansuddin, N. & Al Madhoun, W. A. 2010. Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques. *Environmental Monitoring and Assessment,* 165**,** 475-489.

Gheyas, I. A. & Smith, L. S. A Neural Network Approach to Time Series Forecasting Proceedings of the World Congress on Engineering, July 1 - 3 2009 Imperial College London, London. International Association of Engineers (IAENG), 1292-1296

Gualtieri, G., Crisci, A., Tartaglia, M., Toscano, P., Vagnoli, C., Andreini, B. P. & Gioli, B. 2014. Analysis of 20-year air quality trends and relationship with emission data: The case of Florence (Italy). *Urban Climate,* 10, Part 3**,** 530-549.

Hanke, J. E. & Wichern, D. W. 2005. *Business Forecasting,* Upper Saddle River, N.J, Pearson/Prentice Hall,.

Hassanzadeh, S., Hosseinibalam, F. & Alizadeh, R. 2009. Statistical models and time series forecasting of sulfur dioxide: A case study Tehran. *Environmental Monitoring and Assessment,* 155**,** 149-155.

Heil, A. & Goldammer, J. 2001. Smoke-haze pollution: A review of the 1997 episode in Southeast Asia. *Regional Environmental Change,* 2**,** 24-37.

Heo, J.-S. & Kim, D.-S. 2004. A new method of ozone forecasting using fuzzy expert and neural network systems. *Science of The Total Environment,* 325**,** 221-237.

Huarng, K. 2001. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems,* 123**,** 387-394.

Hugine, A. L., Guerlain, S. A. & Turrentine, F. E. 2014. Visualizing surgical quality data with treemaps. *Journal of Surgical Research,* 191**,** 74-83.

Hyndman, R. J. & Athanasopoulos, G. 2013. Forecasting: Principles and practice. Otexts.

Hyndman, R. J. & Koehler, A. B. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting,* 22**,** 679-688.

Ibrahim, M. Z., Zailan, R., Ismail, M. & Lola, M. S. 2009. Forecasting and time series analysis of air pollutants in several area of Malaysia. *Journal American Journal of Environmental Sciences,* 5**,** 625-632.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. & Kolehmainen, M. 2004. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment,* 38**,** 2895-2907.

Kampa, M. & Castanas, E. 2008. Human health effects of air pollution. *Environmental Pollution,* 151**,** 362-367.

Kandel, S., Parikh, R., Paepcke, A., Hellerstein, J. M. & Heer, J. Profiler: Integrated statistical analysis and visualization for data quality assessment. Proceedings of the Workshop on Advanced Visual Interfaces AVI, 2012. 547-554.

Karatzas, K. D., Papadourakis, G. & Kyriakidis, I. Understanding and forecasting atmospheric quality parameters with the aid of ANNs. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 1-8 June 2008 2008. 2580-2587.

Khashei, M. & Bijari, M. 2010. An artificial neural network (p, d, q) model for time series forecasting. *Expert Systems with Applications,* 37**,** 479-489.

Khashei, M. & Bijari, M. 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing,* 11**,** 2664-2675.

Konovalov, I. B., Beekmann, M., Meleux, F., Dutot, A. & Foret, G. 2009. Combining deterministic and statistical approaches for PM10 forecasting in Europe. *Atmospheric Environment,* 43**,** 6425-6434.

Kumar, A. & Goyal, P. 2011. Forecasting of daily air quality index in Delhi. *Science of The Total Environment,* 409**,** 5517-5523.

Kumar, U. & Jain, V. 2010. ARIMA forecasting of ambient air pollutants (O3, NO, NO2 and CO). *Stochastic Environmental Research and Risk Assessment,* 24**,** 751-760.

Kurt, A. & Oktay, A. B. 2010. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Systems with Applications,* 37**,** 7986-7992.

Lanzafame, R., Scandura, P. F., Famoso, F., Monforte, P. & Oliveri, C. 2014. Air quality data for Catania: Analysis and investigation case study 2010–2011. *Energy Procedia,* 45**,** 681-690.

Law, R. 2000. Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management,* 21**,** 331-340.

Lee, S., Kim, E. & Monsen, K. A. 2015. Public health nurse perceptions of Omaha System data visualization. *International Journal of Medical Informatics,* 84**,** 826-834.

Lemon, L. R. 1977. New severe thunderstorm radar identification techniques and warning criteria: A preliminary report *NOAA Technical Memorandum NWS NSSFC-1* Kansas City, Missouri National Oceanic and Atmospheric Administration.

Li, F. 2010. Air quality prediction in Yinchuan by using neural networks. *In:* Tan, Y., Shi, Y. & Tan, K. (eds.) *Advances in Swarm Intelligence.* Springer Berlin / Heidelberg.

Lim, Y. S., Lim, Y. C. & Pauline, M. J. W. 2008. ARIMA and integrated ARFIMA models for forecasting air pollution index in Shah alam, Selangor. *The Malaysian Journal of Analytical Sciences* 12**,** 257-263.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R. 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting (pre-1986),* 1**,** 111.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K. & Simmons, L. F. 1993. The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting,* 9**,** 5-22.

Makridakis, S. & Hibon, M. 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting,* 16**,** 451-476.

Makridakis, S. & Winkler, R. L. 1983. Averages of forecasts: Some empirical results. *Management Science,* 29**,** 987-996.

Marco, G. & Bo, X. 2013. Air Quality Legislation and Standards in the European Union: Background, Status and Public Participation. *Advances in Climate Change Research,* 4**,** 50-59.

Mariani, S., Casaioli, M., Lanciani, A., Flavoni, S. & Accadia, C. 2015. QPF performance of the updated SIMM forecasting system using reforecasts. *Meteorological Applications,* 22**,** 256-272.

Martins, V. L. M. & Werner, L. 2012. Forecast combination in industrial series: A comparison between individual forecasts and its combinations with and without correlated errors. *Expert Systems with Applications,* 39**,** 11479-11486.

Mintz, D., Fitz-Simons, T. & Wayland, M. Tracking air quality trends with SAS/GRAPH®. *In:* Sas Institute Inc., C., Nc, ed. SAS Conference Proceedings: SAS Users Group International 22, 16-19 March 1997 San Diego, California. 807-812.

Morabito, F. C. & Versaci, M. 2003. Fuzzy neural identification and forecasting techniques to process experimental urban air pollution data. *Neural Networks,* 16**,** 493-506.

Moustris, K., Ziomas, I. & Paliatsos, A. 2010. 3-day-ahead forecasting of regional pollution index for the pollutants NO2, CO, SO2, and O3 using artificial neural networks in Athens, Greece. *Water, Air, & Soil Pollution,* 209**,** 29-43.

Ohno, N. & Kageyama, A. 2007. Scientific visualization of geophysical simulation data by the CAVE VR system with volume rendering. *Physics of the Earth and Planetary Interiors,* 163**,** 305-311.

Palli, D., Sera, F., Giovannelli, L., Masala, G., Grechi, D., Bendinelli, B., Caini, S., Dolara, P. & Saieva, C. 2009. Environmental ozone exposure and oxidative DNA damage in adult residents of Florence, Italy. *Environmental Pollution,* 157**,** 1521-1525.

Palmer, A., José Montaño, J. & Sesé, A. 2006. Designing an artificial neural network for forecasting tourism time series. *Tourism Management,* 27**,** 781-790.

Papanastasiou, D. K., Melas, D. & Kioutsioukis, I. 2007. Development and assessment of neural network and multiple regression models in order to predict PM10 levels in a medium-sized mediterranean city. *Water, Air, and Soil Pollution,* 182**,** 325-334.

Paulhus, J. L. H. & Kohler, M. A. 1952. Interpolation of missing precipitation records. *Monthly Weather Review,* 80**,** 129-133.

Perez, P. & Salini, G. 2008. PM2.5 Forecasting in a large city: Comparison of three methods. *Atmospheric Environment,* 42**,** 8219-8224.

Platt, K. H. 2007. Chinese air pollution deadliest in world, report says. Available: http://news.nationalgeographic.com [Accessed November 2010].

Pope, I. C., Burnett, R. T., Thun, M. J. & Et Al. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA,* 287**,** 1132-1141.

Prybutok, V. R., Yi, J. & Mitchell, D. 2000. Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *European Journal of Operational Research,* 122**,** 31-40.

Raji, U., Mashor, M. Y., Ali, A. N., Adom, A. H. & Sadullah, A. F. 2005. HMLP,MLP and recurrent networks for carbon monoxide concentrations forecasting: A comparison studies. *WSEAS Transactions on Systems,* 4**,** 812-820.

Rizzo, A. & Glasson, J. 2011. Iskandar Malaysia. *Cities*.

Ruiz-Aguilar, J. J., Turias, I. J. & Jiménez-Come, M. J. 2014. Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting. *Transportation Research Part E: Logistics and Transportation Review,* 67**,** 1-13.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986. Learning representations by back-propagating errors. *Nature,* 323**,** 533-536.

Sansuddin, N., Ramli, N., Yahaya, A., Yusof, N., Ghazali, N. & Madhoun, W. 2011. Statistical analysis of PM10 concentrations at different locations in Malaysia. *Environmental Monitoring and Assessment,* 180**,** 573-588.

Sarle, W. S. Neural networks and statistical models. Proceedings of the Nineteenth Annual SAS Users Group International Conference, 1994 Cary, NC, USA. SAS Institute.

Sastry, N. 2002. Forest fires, air pollution, and mortality in Southeast Asia. *Demography,* 39**,** 1-23.

Schaefer, J. T. 1990. The critical success index as an indicator of warning skill. *Weather and Forecasting,* 5**,** 570-575.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M. & Doyle, M. 2003. A rigorous inter-comparison of ground-level ozone predictions. *Atmospheric Environment,* 37**,** 3237-3253.

Sfetsos, A. & Vlachogiannis, D. 2010. Time series forecasting of hourly PM10 using localized linear models. *Software Engineering & Applications,* 3**,** 374-383.

Shen, S., Li, G. & Song, H. 2011. Combination forecasts of International tourism demand. *Annals of Tourism Research,* 38**,** 72-89.

Shi, J. J. 2000. Reducing prediction error by transforming input data for neural networks. *Journal of Computing in Civil Engineering,* 14**,** 109-116.

Singh, R. & Singh, K. 2010. A descriptive classification of causes of data quality problems in data warehousing *International Journal of Computer Science Issues,* 7**,** 41-50.

Slini, T., Karatzas, K. & Moussiopoulos, N. 2002. Statistical analysis of environmental data as the basis of forecasting: An air quality application. *The Science of The Total Environment,* 288**,** 227-237.

Song, Q. 1999. Seasonal forecasting in fuzzy time series. *Fuzzy Sets and Systems,* 107**,** 235-236.

Song, Q. & Chissom, B. S. 1993a. Forecasting enrollments with fuzzy time series — Part I. *Fuzzy Sets and Systems,* 54**,** 1-9.

Song, Q. & Chissom, B. S. 1993b. Fuzzy time series and its models. *Fuzzy Sets and Systems,* 54**,** 269-277.

Song, Q. & Chissom, B. S. 1994. Forecasting enrollments with fuzzy time series — part II. *Fuzzy Sets and Systems,* 62**,** 1-8.

Suhaila, J., Sayang, M. D. & Jemain, A. A. 2008. Revised spatial weighting methods for estimation of missing rainfall data. *Asia-Pacific Journal of Atmospheric Sciences,* 44**,** 93-104.

Suhartono 2011. Time series forecasting by using seasonal autoregressive integrated moving average: Subset, multiplicative or additive model. *Journal of Mathematics and Statistics,* 7**,** 20-27.

Suhartono, Lee, M. & Javedani, H. 2011. A weighted fuzzy integrated time series for forecasting tourist arrivals. *In:* Abd Manaf, A., Zeki, A., Zamani, M., Chuprat, S. & El-Qawasmeh, E. (eds.) *Informatics Engineering and Information Science.* Springer Berlin Heidelberg.

Suhartono & Lee, M. H. 2011. A hybrid approach based on winter's model and weighted fuzzy time series for forecasting trend and seasonal data. *Journal of Mathematics and Statistics,* 7**,** 177-183.

Sunaryo, S., Suhartono, S. & Endharta, A. J. 2011. Double seasonal recurrent neural networks for forecasting short term electricity load demand in Indonesia. *In:* Prof. Hubert Cardot (ed.) *Recurrent Neural Networks for Temporal Data Processing.* InTech.

Tabios, G. Q. & Salas, J. D. 1985. A comparative analysis of techniques for spatial interpolation of precipitation. *JAWRA Journal of the American Water Resources Association,* 21**,** 365-380.

Tang, W. Y., Kassim, A. H. M. & Abubakar, S. H. 1996. Comparative studies of various missing data treatment methods - Malaysian experience. *Atmospheric Research,* 42**,** 247-262.

Taskaya-Temizel, T. & Casey, M. C. 2005. A comparative study of autoregressive neural network hybrids. *Neural Networks,* 18**,** 781-789.

Teegavarapu, R. S. V. & Chandramouli, V. 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology,* 312**,** 191-206.

Tronci, N., Molteni, F. & Bozzini, M. 1986. A comparison of local approximation methods for the analysis of meteorological data. *Archives for meteorology, geophysics, and bioclimatology, Series B,* 36**,** 189-211.

Tseng, F.-M., Yu, H.-C. & Tzeng, G.-H. 2002. Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting and Social Change,* 69**,** 71-87.

Turrado, C. C., López, M. D. C. M., Lasheras, F. S., Gómez, B. a. R., Rollé, J. L. C. & De Cos Juez, F. J. 2014. Missing Data Imputation of Solar Radiation Data under Different Atmospheric Conditions. *Sensors (Basel, Switzerland),* 14**,** 20382-20399.

United State Environmental Protection Agency 2014. Air quality index - A guide to air quality and your health North Carolina, US: USEPA.

Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L. J., Guillen, A., Marquez, L. & Pasadas, M. 2008. Hybridization of intelligent techniques and ARIMA models for time series prediction. *Fuzzy Sets and Systems,* 159**,** 821-845.

Vautard, R., Beekmann, M., Roux, J. & Gombert, D. 2001. Validation of a hybrid forecasting system for the ozone concentrations over the Paris area. *Atmospheric Environment,* 35**,** 2449-2461.

Vidal, J. 2010. London Air Pollution 'Worst in Europe'. Available: http://www.guardian.co.uk [Accessed 12 September 2010].

Videnova, I., Nedialkov, D., Dimitrova, M. & Popova, S. 2006. Neural networks for air pollution nowcasting. *Applied Artificial Intelligence,* 20**,** 493-506.

Wang, D. & Lu, W.-Z. 2006a. Ground-level ozone prediction using multilayer perceptron trained with an innovative hybrid approach. *Ecological Modelling,* 198**,** 332-340.

Wang, X.-K. & Lu, W.-Z. 2006b. Seasonal variation of air pollution index: Hong Kong case study. *Chemosphere,* 63**,** 1261-1272.

Wei, W. W. S. 2006. *Time series analysis: Univariate and multivariate methods,* Pearson Addison Wesley.

Westerlund, J., Urbain, J.-P. & Bonilla, J. 2014. Application of air quality combination forecasting to Bogota. *Atmospheric Environment,* 89**,** 22-28.

Who 2006. WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulphur dioxide - global update 2005. Geneva, Switzerland: World Health Organization.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'donnell, J. & Rowe, C. M. 1985. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans,* 90**,** 8995-9005.

World Resources Institute 2002. Rising energy use: Health effects of air pollution. August 29,2002 ed.: World Resources Institute.

Xia, Y., Fabian, P., Stohl, A. & Winterhalter, M. 1999. Forest climatology: Estimation of missing values for Bavaria, Germany. *Agricultural and Forest Meteorology,* 96**,** 131-144.

Yi, J. S., Kang, Y. A., Stasko, J. T. & Jacko, J. A. 2007. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on,* 13**,** 1224-1231.

Youden, W. J. 1950. Index for rating diagnostic tests. *Cancer,* 3**,** 32-35.

Young, K. C. 1992. A three-way model for interpolating for monthly precipitation values. *Monthly Weather Review,* 120**,** 2561-2569.

Yu, H.-K. 2005. Weighted fuzzy time series models for TAIEX forecasting. *Physica A: Statistical Mechanics and its Applications,* 349**,** 609-624.

Zdeb, M. & Allison, R. Stretching the bounds of SAS/GRAPH® software. SAS Conference Proceedings: SAS Users Group International 30, 10-13 April 2005 Philadelphia, Pennsylvania. SAS Institute Inc., Cary, NC, 1-20.

Zhang, G., Eddy Patuwo, B. & Y. Hu, M. 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting,* 14**,** 35-62.

Zhang, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing,* 50**,** 159-175.

Zhang, G. P., Patuwo, B. E. & Hu, M. Y. 2001. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers & Operations Research,* 28**,** 381-396.

Zhang, G. P. & Qi, M. 2005. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research,* 160**,** 501-514.