Sains Malaysiana 49(1)(2020): 201-209
http://dx.doi.org/10.17576/jsm-2020-4901-24

Air Pollutant Index Calendar-Based Graphics for Visualizing Trends Profiling and Analysis

(Indeks Pencemaran Udara berdasarkan Kalendar Grafik untuk Pemprofilan Tren Visualisasi dan Analisis)

Nur Haizum Abd Rahman* & Muhammad Hisyam Lee

ABSTRACT

Detection of air quality abnormality is important as an early warning system for air quality control and management. The information can raise citizens' awareness towards current air quality status. By using time series plot, the data pattern can be identified but not able to exactly determine the abnormality due to overcrowded plot. Therefore, visualization data profiling was presented in this study by using seven years Malaysia daily air pollutant index to improve the detection. Result shown, the developed approach can simply identify the poor air quality across the month and year. Malaysia air quality was good and consistent between November and May. However, upward trend existed between June and October due to the forest fire happened in Sumatra. This visualization approach improved air pollution detection profiling and it is useful for related agencies to guide the control actions to be taken. This approach can be applied to any countries and data set to give more competent information.

Keywords: Air pollutant index; calendar; data visualization; profiling

ABSTRAK

Pengesanan kelainan kualiti udara adalah penting sebagai sistem amaran awal untuk kawalan dan pengurusan kualiti udara. Maklumat ini dapat meningkatkan kesedaran masyarakat terhadap status kualiti udara semasa. Dengan menggunakan plot siri masa, corak data dapat dikenal pasti tetapi tidak dapat menentukan secara tepat kelainan akibat plot yang sesak. Oleh itu, untuk meningkatkan pengesanan, pemprofilan data visualisasi telah dibincangkan dalam kajian ini dengan menggunakan indeks pencemaran udara harian di Malaysia selama tujuh tahun. Keputusan menunjukkan pendekatan yang digunakan dapat mengenal pasti kualiti udara yang tidak baik sepanjang bulan dan tahun. Kualiti udara di Malaysia adalah baik dan konsisten antara November dan Mei. Bagaimanapun, aliran menaik wujud antara bulan Jun dan Oktober akibat kebakaran hutan di Sumatra. Pendekatan profil visualisasi dapat mengesan pencemaran udara dan berguna kepada agensi berkaitan untuk membimbing tindakan kawalan yang akan diambil. Pendekatan ini boleh digunakan untuk mana-mana negara dan set data untuk memberikan maklumat yang lebih cekap.

Kata kunci: Indeks pencemaran udara; kalendar; pemprofilan; visualisasi data

Introduction

Clean air is considered as crucial necessity for human health and well-being. However, the air pollution continues to pose a major threat to health globally (Moustris et al. 2010). In this study, we develop a method for organizing and visualizing temporal data, collected at daily intervals into a calendar graphics layout. The objective to conduct this calendar-based graphics was to provide understanding into people's daily activities especially for sensitive groups. This sensitive group with greater risk includes people with lung and heart disease, older adults, children and people who are active outdoors (United States Environmental Protection Agency 2014).

This works was motivated by studying the air pollutant data in Malaysia. There have been 52 continuous ambient air quality monitoring (CAQM) stations that continuously recorded ambient air for gaseous pollutants such as; sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and carbon monoxide (CO), particulate matters diameter 10 (PM_{10}) and diameter 2.5 ($PM_{2.5}$) and also meteorological parameters like wind speed, wind direction and temperature (Ahmat et al. 2015).

Air pollutants directly affects our quality of life. The air quality status may change from day to day or even hour to hour. Therefore, the detection of poor air quality is important to provide an early warning system for air quality control and management including to the public understanding. The air pollutant index (API) is designed as referential parameter in describing air quality status so that everyone able to realize and understand easily current situations like weather forecast.

The API in Malaysia was developed based on the API introduced by the United States Environmental Protection Agency (USEPA) and it is determined by the calculation of sub-indexes of five main pollutants. The pollutants known as particulate matter (PM_{10}), ozone (O_3), carbon dioxide (CO_2), sulphur dioxide (SO_2) and nitrogen dioxide (NO_2).

It can be used by the government agencies to characterize the status of air quality since API provide an easy assessment and harmonization at the given location through out the countries (Sansuddin et al. 2011). API also helps the general public to understand the air quality status easily for their own health precaution. This 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)' (Leh et al. 2012). These ranges reflected as benchmark for air quality management or data interpretation for decision making processes (Afroz et al. 2003).

As mention before, there have been 52 CAQM in Malaysia and a few has been monitored since 1996. Hence, large amount of data available. Usually, the examination construction based on conventional time series plots to detect any abnormality of temporal patterns. However, this approach is insufficient with the measurements are made at different locations at a given time point (Wang et al. 2018). Thus, it is needed to create calendar-based graphics for comparing and contrasting between locations. Thus, data visualization using calendar-based graphics can help to provide timely air quality status information to the public, government officials or administrative users. These graphics presentations describe the data within the range indicating different health status used to give visual information. Besides, it is also a great instrument to highlight the polluted area and time information to improve the actions that should be taken.

The term data visualization means that the data is supported by the computer in order to present interactive visual representations of data to amplify cognition and the use of knowledge (Dilla & Raschke 2015). Data visualizations are important to help people to see things that do not normally occurred. In addition, this approach can convey the information universally and make it simple to share the ideas with others.

Depending on the purpose of the study, visualization can be achieved in many ways. Thus, studied that use data visualization have been described and applied in many areas of studies. Hugine et al. (2014) used tree map as the visualization approach in order to represent the surgical data to allow quick visual judgements for both surgeons and health care administrators. Lee et al. (2015) have use visualization technique in electronic health records (EHRs) to summarize the patient data for clinicians in order to get responds from public health nurses. Positive feedback from public health nurses and improvement were provided with the help of well-designed data visualization.

Another study by Dilla and Raschke (2015) showed the advantage of data visualization since it can help in detecting fraud transactions. Such visualization analysis supports the different cognitive processes required as it allows large data sets navigation and it recognizes the change in data representation. Gualtieri et al. (2014) with 20 years data and Lanzafame et al. (2014) with two years data, visualize air quality in Italy. The main atmospheric pollutants were examined by both researchers. They visualize the data profile in order to show data characteristics mainly in concentrations peak especially during commute times which are early morning and afternoon.

This paper is organized as follows. The second section demonstrates the material and methods including the chosen monitoring station and calendar-based graphic construction. The third section will be discussed on result of obtained calendar-based graphic. Finally, in section four the conclusion of this approach.

Materials and Methods

The Monitoring Stations

Malaysia is a part of Southeast Asia countries with a federation of 13 states and three federal territories. It divided into two regions, Peninsular Malaysia which consists of 11 states and two federal territories and East Malaysia which consists of two states and one federal territory. Malaysia has been regularly affected by the trans-boundary pollution from neighbouring countries (particularly from biomass burning in Sumatra, Indonesia), which has been usually the main factor behind hazardous occurrences. It is shown by the dataset obtained from the Department of the Environment (DoE) Malaysia through Alam Sekitar Malaysia Sdn. Bhd. (ASMA). ASMA is a private company which is responsible for monitoring and managing the ambient air quality in Malaysia. It offers variety of air quality monitoring locations background; urban, suburban and rural area.

In this study, 46 continuous monitoring stations were used and the remaining five stations that did not contain the data within the same period of time were excluded. The study used the API daily data set for seven years, which covered the period from 2005 until 2011. The details of these 46 continuous monitoring stations located in both region, Peninsular Malaysia (32 stations) and East Malaysia (14 stations) are shown in Table 1. Meanwhile, the locations are shown in Figure 1.

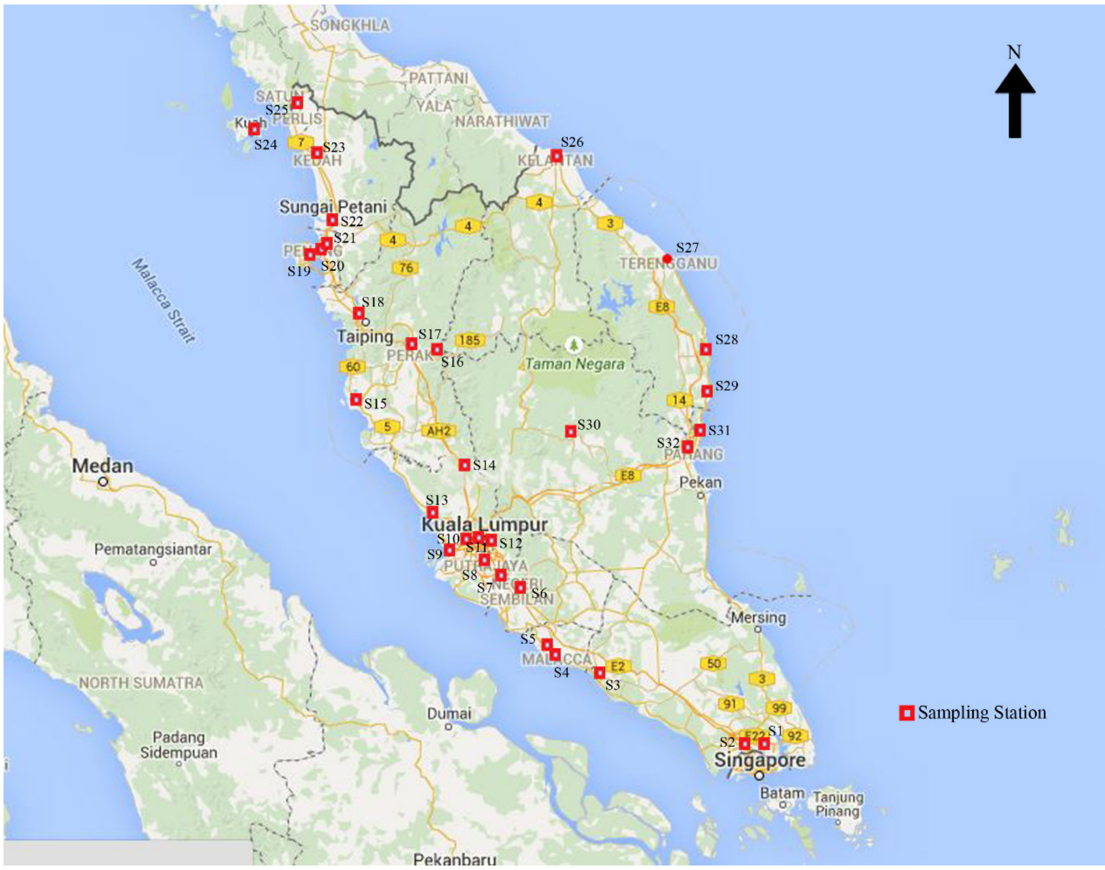
Construction

The approach to build calendar-based graphic is achieved by PROC GMAP function from SAS/GRAPH software. PROC GMAP function create customize map without representing the real geographical areas. It is based on X and Y coordinates where a variable will identify each map area. The map area used is a geometry that represent the data grid with calendar representation. Thus, it is very useful in displaying information with time dimension.

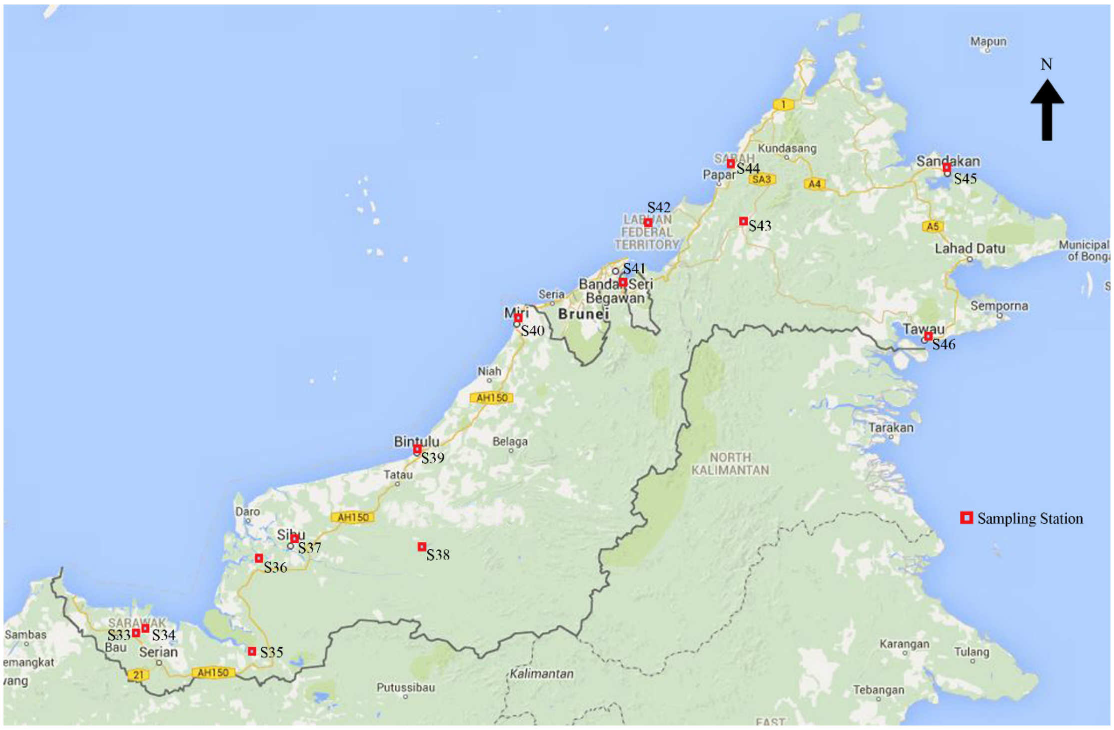
The PROC GMAP function outline months and label the calendar using the Annotate facility. To make a yearlong calendar, it requires cells for days, block according to months and will be organized into a grid layout for a year. The available block for each day follows the real general calendar. The first day of the month could be any for example from Sunday to Saturday, which it is determined

TABLE 1. Air quality monitoring stations in Malaysia

Site	Site State	Location	Latitude	Longitude
S1	Johor	SM Pasir Gudang 2, Johor	1.493	103.890
S2	Johor	SM Vok. Perdagangan, Johor Bahru	1.496	103.726
S3	Johor	SM Teknik, Muar	2.058	102.597
S4	Melaka	SM Tun Tuah, Bachang	2.212	102.236
S5	Melaka	SMK Bukit Rambai, Melaka	2.259	102.173
S6	N. Sembilan	SM Teknik Tuanku Jaafar, Seremban	2.723	101.967
S7	N. Sembilan	Taman Semarak (Phase II), Nilai	2.821	101.812
S8	Putrajaya	SK Putrajaya 8(2), Jln P8/E2, Presint 8	2.940	101.685
S9	Selangor	SM(P) Raja Zarina, Klang	3.012	101.410
S10	Selangor	SK TTDI Jaya, Shah Alam	3.107	101.557
S11	Selangor	SK Sri Petaling, Petaling Jaya	3.109	101.640
S12	Kuala Lumpur	SMK Seri Permaisuri, Cheras	3.107	101.718
S13	Selangor	SM Sains, Kuala Selangor	3.319	101.274
S14	Perak	UPSI, Tanjung Malim	3.686	101.525
S15	Perak	Kompleks Pentadbiran Seri Manjung	4.200	100.664
S16	Perak	SM Pegoh, Ipoh	4.593	101.304
S17	Perak	SM Jalan Tasek, Ipoh	4.630	101.118
S18	Perak	SK Air Puteh, Taiping	4.882	100.683
S19	Pulau Pinang	USM, Penang	5.356	100.303
S20	Pulau Pinang	SK Cenderawasih, Perai	5.389	100.386
S21	Pulau Pinang	SK Seberang Jaya II, Perai	5.397	100.405
S22	Kedah	SK Bakar Arang, Sungai Petani	5.628	100.472
S23	Kedah	SMK Agama Kedah, Mergong, Alor Setar	6.140	100.351
S24	Kedah	Kompleks Sukan Langkawi, Langkawi	6.332	99.861
S25	Perlis	MADA, Behor Temak, Kangar	6.534	100.196
S26	Kelantan	Maktab Sultan Ismail, Kota Baru	6.108	102.252
S27	Terengganu	SK Pusat Chabang Tiga, Kuala Terengganu	5.307	103.122
S28	Terengganu	TNB Quarters, Paka-Kertih	4.603	103.433
S29	Terengganu	SRK Bukit Kuang, Teluk Kalung, Kemaman	4.266	103.431
S30	Pahang	Pej. Kajicuaca, Batu Embun, Jerantut	3.956	102.363
S31	Pahang	SK Balok Baru, Kuantan	3.961	103.381
S32	Pahang	SK Indera Mahkota, Kuantan	3.819	103.292
S33	Sarawak	Medical Store, Kuching, Sarawak	1.418	110.354
S34	Sarawak	Pejabat Daerah Samarahan, Kota Samarahan	1.461	110.435
S35	Sarawak	Pejabat Residen, Sri Aman	1.240	111.462
S36	Sarawak	Balai Polis Sentral, Sarikei	2.133	111.522
S37	Sarawak	Ibu Pejabat Polis Sibu	2.319	111.868
S38	Sarawak	Stadium Tertutup, Jalan Bleteh, Kapit	2.243	113.094
S39	Sarawak	Balai Polis Pusat, Bintulu	3.177	113.041
S40	Sarawak	SM Dato Permaisuri, Miri	4.425	114.014
S41	Sarawak	Dewan Suarah, Limbang	4.760	115.014
S42	Labuan	Taman Perumahan MPL, Labuan	5.332	115.238
S43	Sabah	SMK Gunsanad, Keningau	5.338	116.164
S44	Sabah	SM Putatan, Kota Kinabalu	5.893	116.043
S45	Sabah	Pejabat JKR Sandakan, Sandakan	5.856	118.109
S46	Sabah	Pejabat JKR, Tawau	4.251	117.936



(a) Peninsular Malaysia



(b) East Malaysia

FIGURE 1. Locations of sampling stations for air quality monitoring in Malaysia: (a) Peninsular Malaysia; (b) East Malaysia

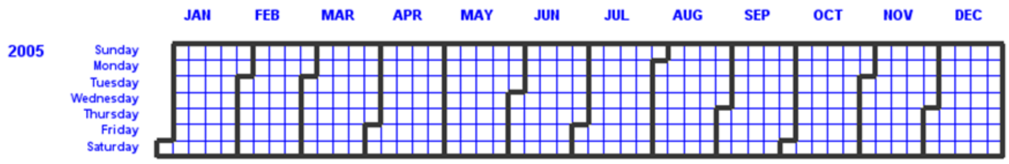


FIGURE 2. PROC GMAP output for one year grid in 2005

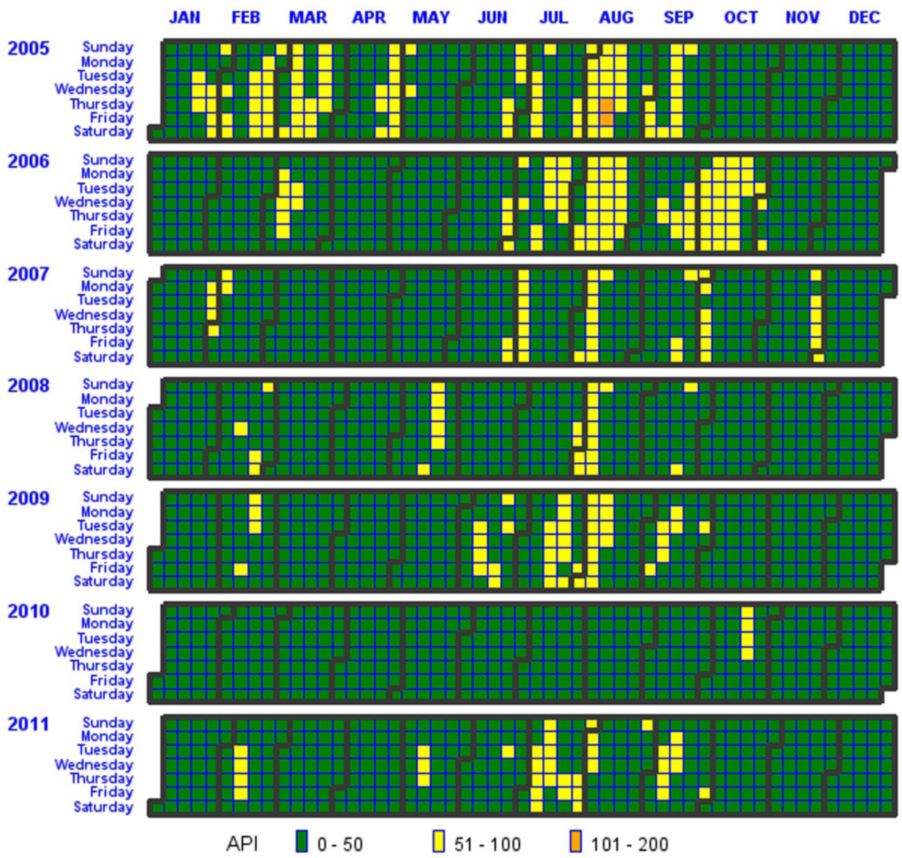


FIGURE 3. Air pollutant index (API) trends in Malaysia from 2005 to 2011

by the year of the calendar. To separate between months, the bold line will be used to differentiate between months. The steps in obtaining for a one-year cell is as follows:

Create the minimum year (*min_year*) and the maximum year (*max_year*); Create a data set with 365 observations and two variables, day and year; Create a grid to be used as a map data set; Add the variable *yr_mon* to the data set GRID; and Create an Annotate data set from the month boundaries.

Thus, from these five steps, calendar-based grid graphics can be a map data set by outlining months and labeling the calendar. The coding can be extended to create more than a year calendar by changing the macro variables, *min_year* and *max_year*. For this study, the displayed data are the air pollutant index (API) observation for each day in the seven-year period. Two variables to map the observation which are day, to place the observation in the right rectangle of the calendar and the API, variable to be mapped in the calendar. The example for one year grid calendar shown in Figure 2.

Results and Discussion

The air quality status in Malaysia is defined from an indicator known as air pollutant index (API). Malaysia's air quality system has been recorded since 1996 and currently the total number of stations available is 52. Considering the same time interval from 2005 until 2011, only 46 stations were left. An overview of this seven years air quality status was presented in this study. Initially by averaging the daily scale at all stations. The obtained profiling for whole Malaysia is shown in Figure 3.

Based on Figure 3, most of the data recorded were within good (2231 days) and moderate (323 days) conditions. However, two days were recorded with unhealthy API in 2005 where the readings recorded were

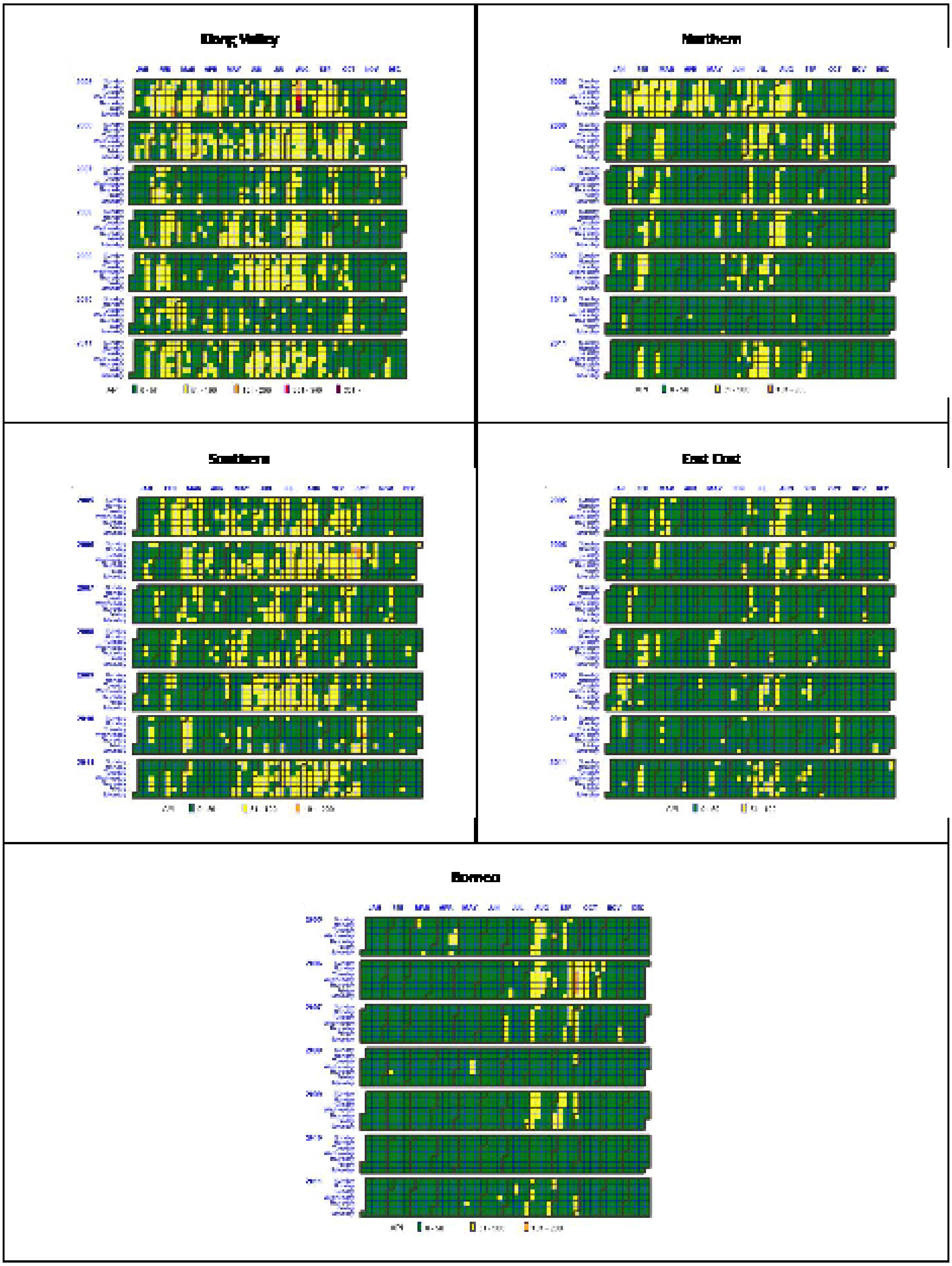


FIGURE 4. Air pollutant index (API) trends from 2005 to 2011in Klang Valley, Northern region, Southern region, East Cost and Borneo

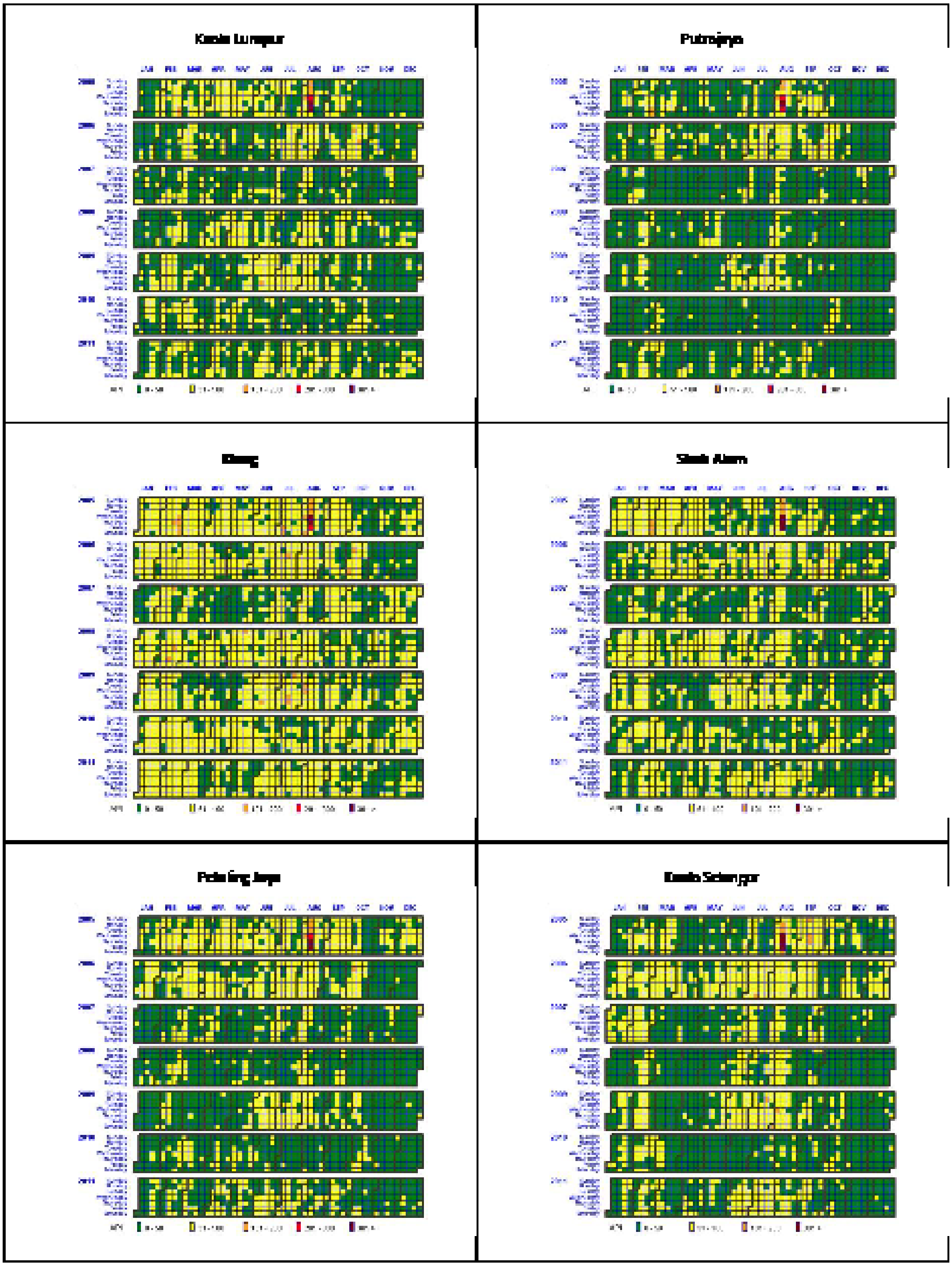


FIGURE 5. Air pollutant index (API) trends for each station in Klang Valley

117 on 11th August and 103 on 12th August. The haze in 2005 is one of the worst haze events reported and some areas received severe haze, thus affecting the overall air quality status in Malaysia. Based on the figure, the trend of API was good between November and May, and worst from Jun to October. This is due to the monsoon factor in which from Jun to October, the wind blow from southwest Malaysia and within this period the forest fire happening in Sumatra brings the polluted air containing small particles into Malaysia's airspace.

West and east are the two sub-regions in Malaysia. West Malaysia or Peninsular Malaysia can be divided into four main sub-areas; Klang Valley, Northern region, Southern region, and East Cost. Meanwhile, east Malaysia which consists of two states and one federal territory is usually defined as one area. The Klang Valley is known as the city centre of Malaysia where a large scale of industrial and commercial activities is located here. Besides, it is densely populated with high vehicular traffic. Thus, the profiling shown in Figure 4 is true indicating that the Klang Valley is more prone to air pollution compared to other areas. This is shown from year 2005 to 2011 in which most of the data were moderate with unhealthy, very unhealthy, and hazardous air quality were also recorded. The Southern region was shown to be the second most polluted area as a whole, followed by the Northern region. The East Coast areas remained good most of the time throughout the years. Meanwhile, for east Malaysia or known as Borneo, the air quality status was good in general except for some days which unhealthy API was detected in 2006.

The Klang Valley is an area centred in Kuala Lumpur. It includes its adjoining cities in the state of Selangor and the Federal Territory of Putrajaya. Since Klang Valley is the most predominant occurrences of worst air quality, profiling for each station is presented here. A total of six monitoring stations are located in this area; one in Federal Territory of Kuala Lumpur, four in the state of Selangor (Klang, Shah Alam, Petaling Jaya, and Kuala Selangor), and one in Federal Territory of Putrajaya. The data profiling are shown in Figure 5. Comparing all six stations, Putrajaya showed the best air quality status while Klang showed the worst air quality value. However, the haze reported in 2005 also affected Putrajaya as shown in the figure.

Hazardous API was recorded in all stations in Klang Valley. The highest was recorded in Kuala Selangor with the reading of 519, followed by Klang, Shah Alam, Putrajaya, Petaling Jaya, and Kuala Lumpur where the API values recorded were 495, 440, 372, 353, and 326 respectively. The hazy condition reported was more severe in intensity compared to the haze event recorded in September. Thus, haze emergency was declared on 11th August 2005. In details, the overall number of air quality status in the Klang Valley is shown in Table 2.

TABLE 2. The frequency of air quality status in Klang Valley

Year	Good (0-50)	Moderate (51-100)	Unhealthy (101-200)	Very Unhealthy (201-300)	Hazardous (>300)
2005	174 (47.6)	183(50.1)	5(1.4)	1(0.3)	2(0.5)
2006	177 (48.5)	187 (51.2)	1(0.3)	-	-
2007	281 (77.0)	84 (23.0)	-	-	-
2008	254 (69.6)	112 (30.7)	-	-	-
2009	227 (62.2)	138 (37.8)	-	-	-
2010	287 (78.6)	78 (21.4)	-	-	-
2011	223 (61.1)	142 (38.9)	-	-	-

*The percentage in bracket

Conclusion

The air quality status in Malaysia is determined by air pollutant index (API). A seven years daily API analysis was conducted in which the monitoring stations are located in different backgrounds; urban, suburban and rural areas. The analysis was done by using data visualization approach to obtain the API data profiling. Normally, the data stored in contexts enable us to determine the status of air quality effectively. However, based on the performed profiling, the data was more presentable and provided excellent search. The purpose of conducting data profiling in this study was to help us to understand and characterize the trend of air quality status throughout the years and to compare globally with all available monitoring stations dataset. In addition, it helps to easily detect abnormal data. In conclusion, the interactive data visualization tool for air pollution detection profiling is useful for related agencies or government to guide the control actions to be taken. Moreover, this approach can be applied to any countries and data set to give more competent information.

Acknowledgements

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

Afroz, R., Hassan, M.N. & Ibrahim, N.A. 2003. Review of air pollution and health impacts in Malaysia. *Environmental Research* 92(2): 71-77.

Ahmat, H., Yahaya, A.S. & Ramli, N.A. 2015. PM_{10} analysis for three industrialized areas using extreme value. *Sains Malaysiana* 44(2): 175-186.

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.

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(P3): 530-549.

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

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(2): 681-690.

Lee, S., Kim, E. & Monsen, K.A. 2015. Public health nurse perceptions of Omaha System data visualization. *International Journal of Medical Informatics* 84(10): 826-834.

Leh, O.L.H., Ahmad, S., Aiyub, K., Jani, Y.M. & Hwa, T.K. 2012. Urban air environmental health indicators for Kuala Lumpur city. *Sains Malaysiana* 41(2): 179-191.

Moustris, K.P., Ziomas, I.C. & Paliatsos, A.G. 2010. 3-day-ahead forecasting of regional pollution index for the pollutants NO_2, CO, SO_2, and O_3 using artificial neural networks in Athens, Greece. *Water, Air, and Soil Pollution* 209(1-4): 29-43.

Sansuddin, N., Ramli, N.A., Yahaya, A.S., Yusof, N.F.F.M., Ghazali, N.A. & Madhoun, W.A. 2011. Statistical analysis of PM10 concentrations at different locations in Malaysia. *Environmental Monitoring and Assessment* 180(1-4): 573-588.

United States Environmental Protection Agency. 2014. Air quality index: A guide to air quality and your health. In *EPA*. https://doi.org/10.1023/A:1020370119096.

Wang, E., Cook, D. & Hyndman, R. 2018. Calendar-based graphics for visualizing people's daily schedules. *Journal of Statistical Software* 10(2): 1-19. doi:10.18637/jss.v000.i00.

Nur Haizum Abd Rahman*
Department of Mathematics
Faculty of Science
Universiti Putra Malaysia
43400 UPM Serdang, Selangor Darul Ehsan
Malaysia

Muhammad Hisyam Lee
Department of Mathematical Sciences
Faculty of Science
Universiti Teknologi Malaysia
81310 Johor Bahru, Johor Darul Takzim
Malaysia

*Corresponding author; email: nurhaizum_ar@upm.edu.my

Received: 3 July 2019
Accepted: 17 October 2019