Poll Res. 41 (3) : 1016-1021 (2022)

ISSN 0257–8050
DOI No.: http://doi.org/10.53550/PR.2022.v41i03.037

AN ANALYSIS OF PM2.5 USING THE DIRECTED ACYCLIC GRAPH: A CASE STUDY OF SOUTHEAST ASIA

BUSARAKAM PRATANSUP[1] AND PANNAPA CHANGPETCH[2]

Department of Mathematics, Faculty of Science, Mahidol University, Bangkok, Thailand
Postal address: 272 Rama VI Road, Bangkok 10400, Thailand

(Received 8 January, 2022; Accepted 4 February, 2022)

ABSTRACT

In this study, we used the directed acyclic graph (DAG) to analyze a dataset of PM2.5 concentration levels in 11 cities in Southeastern Asian countries for the period of March 18 to April 11, 2021. This is the period that PM2.5 posted a major problem for several cities in Southeast Asia. We analyzed the data to determine the effect of the PM2.5 concentration in each city on that of each of the other cities. The results show that the PM2.5 concentration in Yangon, Myanmar, had a direct effect on the PM2.5 concentration in Chiang Mai, Thailand, and that the PM2.5 concentration in Kuala Lumpur, Malaysia, had a direct effect on the PM2.5 concentration in Singapore, Republic of Singapore. Where one country has a direct effect on another, a reduction or increase of PM2.5 in the former will reduce or increase, respectively, the PM2.5 concentration in those on which it has an effect. Given this relationship, if the concentration of PM2.5 is to be reduced either overall or in any specific country is to be reduced, international cooperation will be essential. Our application of DAG to this environmental issue exemplifies the viability of this technique for exploring and indicating solutions for real-world problems more generally.

KEY WORDS: Directed acyclic graph, PM2.5, Southeast Asia, Air Pollution.

INTRODUCTION

Fine particulate matter with a diameter of no more than 2.5 micrometers, also referred to as particle pollution, PM2.5 has become a major health concern in many countries. High concentrations of PM2.5 are linked to a variety of other significant health problems, including harmful cardiovascular events such as heart attack and stroke, both of which can lead to premature death (https://www.epa.gov/sites/default/files/2016-04/documents/2012_aqi_factsheet.pdf, accessed on August 14, 2012).

Air pollution related to PM2.5 is the focus of a growing literature with relevance to countries worldwide. Among the works in this area in which researchers implement statistical methods and data mining techniques are the following: Yuchi *et al.* (2019) evaluated the relative efficacy of multiple linear regression, random forest regression, and a mixed model to predict indoor PM2.5 in Ulaanbaatar, Mongolia. Roy *et al.* (1997) used regression to study the causes of PM2.5 in Birmingham, England. Karimian *et al.* (2019) predicted the effects of PM2.5 concentration on long short-term memory (LSTM) using three data mining methods: multiple additive regression tree, a deep feedforward neural network, and a new hybrid model base. Chao *et al.* (2016) implemented a land use regression method to predict the PM2.5 concentration in Shanghai, China.

We advance the research in this area by using the directed acyclic graph (DAG) data mining technique as a way to visualize the effects between countries. To be specific, we use DAG to explore the effects of PM2.5 concentration for 11 cities in Southeastern Asian countries, as PM2.5 has become a known problem in this part of the world.

Given its efficacy in visualizing the effects between objects of interest, DAG has been used in a wide variety of fields. For example, using a DAG approach in conjunction with the bootstrap technique, Chowdhury *et al.* (2020) identified a biomarker of a protein set in ovarian cancer, whereas Lal *et al.* (2020) drew on DAG to develop and examine a digital twin model of critical patients to determine the causal relationship between organ systems and treatment methods. Yang and Zhao (2014) used DAG to investigate a causal pattern between India's economic growth, energy use, and carbon emissions, and Haigh and Bessler (2004) used DAG to generate causal messages between three related markets and then applied the results to an error-correction model to address problems relating to the dynamic model of price discovery. Using data from Nigerian insurance companies, Changpetch and Akinyemi (2020) applied DAG to the total amount of the claims for 89 diagnoses to determine the direct and indirect effects between those diagnoses and Williams *et al.* (2018) used the DAG approach to interpret diagnoses related to causal associations in epidemiological clinical trials. Based on a socioeconomic survey of 43,844 Thai households, Changpetch and Haughton (2018) investigated direct and indirect effects between the proportion of household expenditure on alcohol, on tobacco, and on gambling and fourteen demographic factors.

The paper is organized as follows: In Section 2, we present the dataset. In Section 3, we describe the method used and the results derived from the analysis. In Section 4, we provide a discussion and conclusion.

DATASET

The dataset covers daily PM2.5 concentration data from March 18 to April 11, 2021, from11 major cities in Southeast Asia: (i) Bangkok, Thailand, (ii) Chiang Mai, Thailand, (iii) Yangon, Myanmar, (iv) Vientiane, Laos, (v) Hanoi, Vietnam, (vi) Phnom Penh, Cambodia, (vii) Kuala Lumpur, Malaysia, (viii) Singapore, Republicof Singapore, (ix) Jakarta, Indonesia, (x) Manila, Philippines, (xi) Dili, Timor-Leste (https://www.iqair.com/th-en/, accessed on April 28, 2021). Notethat daily PM concentration data for Brunei Darussalam were not available such that we could not include this country in the analysis. All the cities covered in the analysis are currently the capital cities of their respective countries with two exceptions: Yangon, included because it is Myanmar's most populous city and most important commercial center and the countrys capital until 2006, and Chiang Mai, included because its PM2.5 concentr atio became the highest in the world during March 2 21 The distances between the 11 cities are shown in Table 1.

The original PM2.5 data constitute the quantitative variable such that it was necessary to discretize the variable for use with DAG. We discretized the data based on the air-quality categories defined by the U.S. Environmental Protection Agency (EPA) (www.epa.gov, accessed on August 14, 2021), as shown in Table 2.

Table 2. Six air-quality categories based on PM2.5 level

Level	Category	PM2.5 (μg/m3)
L1	Good	0.0–12.0
L2	Moderate	12.1–35.4
L3	Unhealthy for sensitive groups	35.5–55.4
L4	Unhealthy	55.5–150.4
L5	Very unhealthy	150.5–250.4
L6	Hazardous	250.5 and above

We discretized the original data by the levels defined in Table 2 as preparation for utilizing DAG. A summary of the discretized data is shown in Table

Table 1. Distance (km) between 11 cities in Southeast Asia

	Bangkok	Chiang Mai	Yangon	Vientiane	Hanoi	Phnom Penh	Kuala Lumpur	Singapore	Jakarta	Manila	Dili
Bangkok	0.00										
Chiang Mai	582.89	0.00									
Yangon	583.25	371.21	0.00								
Vientiane	519.46	393.66	699.39	0.00							
Hanoi	988.01	760.01	1123.17	481.35	0.00						
Phnom Penh	531.28	1019.55	1109.94	749.32	1055.28	0.00					
Kuala Lumpur	1186.12	1763.75	1642.13	1649.44	2037.33	999.68	0.00				
Singapore	1434.09	2016.94	1927.15	1858.85	2205.60	1150.32	315.07	0.00			
Jakarta	2323.35	2905.97	2819.07	2723.49	3026.71	1986.70	1182.49	893.36	0.00		
Manila	2208.39	2386.69	2670.51	1994.80	1752.56	1775.72	2464.77	2392.52	2785.90	0.00	
Dili	3717.49	4213.92	4299.74	3880.79	3933.49	3204.20	2949.49	2645.90	2084.69	2623.56	0.00

Table 3. Percentages of L1–L4 of discretized data for 11 cities

Level	Bangkok	Chiang Mai	Yangon	Vientiane	Hanoi	Phnom Penh	Kuala Lumpur	Singapore	Jakarta	Manila	Dili
L1	7.69%	3.85%	0.00%	11.54%	0.00%	0.00%	19.23%	38.46%	7.69%	84.62%	100.00%
L2	69.23%	19.23%	19.23%	50.00%	23.08%	65.38%	65.38%	61.54%	61.54%	15.38%	0.00%
L3	23.08%	15.38%	42.31%	19.23%	50.00%	30.77%	15.38%	0.00%	26.92%	0.00%	0.00%
L4	0.00%	61.54%	38.46%	19.23%	26.92%	3.85%	0.00%	0.00%	3.85%	0.00%	0.00%

3. Note that there are no records for the "very unhealthy" and "hazardous" categories. Therefore, L5 and L6 are not included in Table 3.

The percentages for L1–L4 are shown in Table 3. For example, for Bangkok, 2 of the 26 records (7.69%) belong to L1 ("Good"). In Deli, all the records belong to L1 (100%). Figure 1 summarizes these percentages via a stack plot. As shown in this Figure, Chaing Mai suffered the unhealthy level (L4) of PM2.5 for more than half of the study period.

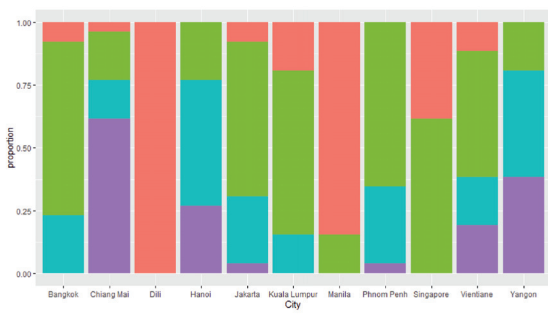

Fig. 1. Stack plot representing percentages of L1–L4 records of the discretized data for 11 cities

METHODOLOGY AND RESULTS

As a tool for visualization, DAG represents the relationship between nodes through arrows. For example, $A \rightarrow B$ indicates that A affects B. In this study, we use DAG to explore the relationships between PM2.5 concentrations in 11 major cities in Southeast Asia. We employed an algorithm implemented by the Bayesialab software http://www.bayesia.com, accessed on April 31, 2021)—i.e., a Taboo search algorithm (a greedy score–based algorithm), which makes it possible to temporarily iterate to less optimal solutions with a smaller score to avoid becoming stuck near a local optimum in the search space.

There are two relationships, as shown in Figure 2: First, the PM2.5 concentration in Yangon, Myanmar, directly affected the PM2.5 concentration in Chiang Mai, Thailand. Second, the PM2.5 concentration in Kuala Lumpur, Malaysia, directly affected the PM2.5 concentration in Singapore.

Figure 3 shows the impact of the PM2.5 concentration in Yangon, Myanmar, on the PM2.5 concentration in Chiang Mai, Thailand. If we force the PM2.5 concentration in Yangon to reach 100% at L2, the PM2.5 concentration in Chiang Mai will be 100% at L2 as well (Figure 3(a)). If we force the PM2.5 concentration in Yangon to reach 100% at L3, the PM2.5 concentration in Chiang Mai will be 9.09% at L1, 27.7% at L3, and 63.64% at L4 (Figure 3(b)). If we force the PM2.5 concentration in Yangon to reach 100% at L4, the PM2.5 concentration in Chiang Mai will be 10% at L3 and 90% at L4 (Figure 3(c)). Based on Figures 3(a), 3(b), and 3(c), we can conclude that if we force the PM2.5 concentration in Yangon higher, the PM2.5 concentration in Chiang Mai will increase as well.

Figure 4 shows the impact of the PM2.5 concentration in Kuala Lumpur, Malaysia, on the PM2.5 concentration in Singapore, Republic of Singapore. If we force the PM2.5 concentration in Kuala Lumpur to reach 100% at L1, the PM2.5 concentration in Singapore will reach 100% at L1 as well (Figure 4(a)). If we force the PM2.5 concentration in Kuala Lumpur to reach 100% at L2, the PM2.5 concentration in Singapore will be 29.41% at L1 and 70.59% at L2 (Figure 4(b)). If we force the PM2.5 concentration in Kuala Lumpur to reach 100% at L3, the PM2.5 concentration in Singapore

Fig. 2. Directed acyclic graph indicating the direction of effects for 11 Southeast Asian cities

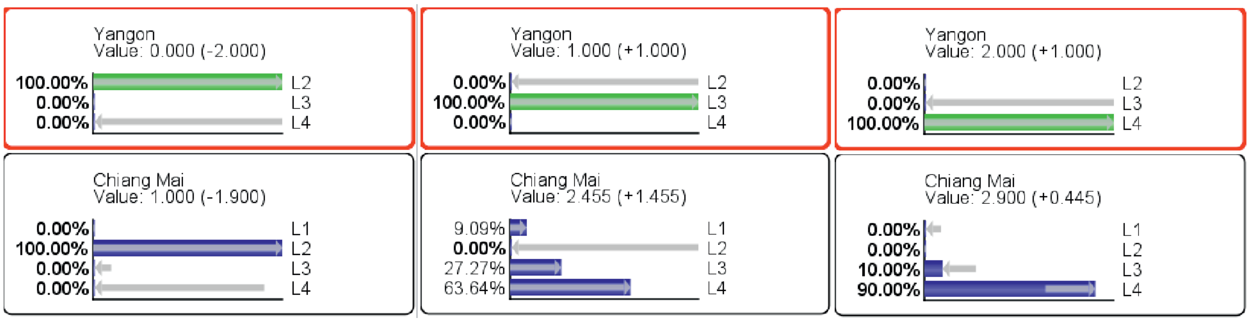

Figure 3(a) Figure 3(b) Figure 3(c)

Fig. 3. Impact of multiple levels of PM2.5 in Yangon, Myanmar, on the PM2.5 level in Chiang Mai, Thailand

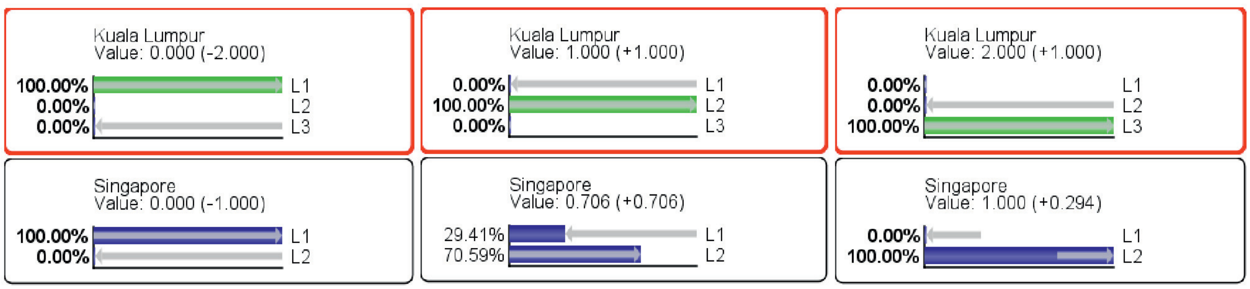

Figure 4(a) Figure 4(b) Figure 4(c)

Fig. 4. Impact of multiple levels of PM2.5 in Kuala Lumpur, Malaysia, on the PM2.5 level in Singapore, Republic of Singapore

will reach 100% at L2 (Figure 4(c)). Based on Figures 4(a), 4(b), and 4(c), we can conclude that if we force the PM2.5 concentration in Kuala Lumpur higher, the PM2.5 concentrations in Singapore will increase as well.

DISCUSSION

We used the directed acyclic graph (DAG) technique to explore the relationships between the PM2.5 concentrations for 11 major cities in Southeastern Asian countries. We based our analysis of these relationships on data for the period of March 18 to April 11, 2021.

The first relationship observed from DAG is that the PM2.5 concentration in Yangon, Myanmar, directly affected the PM2.5 concentration in Chiang Mai, Thailand. Of all the cities, these two cities are the closest together in terms of distance (Figure 1). We also observed the direction of the wind during the period of March 18 to April 11, 2021, and found that the wind blew from Yangon, Myanmar, to Northern Thailand where Chiang Mai on most of the days during this period (http://asmc.asean.org/home/, accessed on August 15, 2021). This air mass movement direction from the southwest direction where Yangon, Myanmar, is located, which affected the air quality of Chiang Mai, Thailand, is in keeping with observations made in other research studies.

The second relationship observed from DAG is that the PM2.5 concentration in Kuala Lumpur, Malaysia, directly affected the PM2.5 concentration in Singapore, Republic of Singapore. Of all the cities, these two are the second closest together in terms of distance (Figure 1). As indicated in previous studies, Singapore is affected by the activities taking place in neighboring countries. For example, Reddington *et al.* (2014) showed that fires occurring in the Peninsular Malaysia account for some fraction of the total effect of fire on increasing PM2.5 in Singapore, Republic of Singapore.

Chiang Mai, Thailand, showed a very high level of PM2.5 in the focal period mainly due to burning forests in the area but also affected by neighboring Myanmar due to the wind blowing principally from that country towards Chiang Mai. A direct effect of a given country on another or others indicates that a reduction or increase of PM2.5 in the former will reduce or increase, respectively, the PM2.5 concentration in those on which it has an effect. It is evident, therefore, that if we are to reduce the concentration of PM2.5 in Chiang Mai, Thailand, it will be necessary to secure international cooperation from the neighboring country, Myanmar. Similarly, the Republic of Singapore will need cooperation from Malaysia to mitigate the effect from that country.

This study in which we used DAG to visualize and analyze important environmental effects between countries offers a viable way to use this technique to consider and indicate solutions to consequential real-world issues. The technique can be used in similar applications worldwide to yield results that are accurate, meaningful, and actionable.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

Abba, E. J., Unnikrishnan, S., Kumar, R., Yeole, B. and Chowdhury, Z. 2012. Fine aerosol and PAH carcinogenicity estimation in outdoor environment of Mumbai City, India. *Int. J. Environ. Health Res.* 22: 134-149.

Betha, R., Behera, S.N. and Balasubramanian, R. 2014. Southeast Asian smoke haze: Fractionation of particulate-bound elements and associated health risk. *Environ. Sci. Technol.* 48: 4327-4335.

Chang, C. C., Chiu, H. F. and Yang, C. Y. 2015. Fine particulate air pollution and outpatient department visits for headache in Taipei, Taiwan. *J. Toxicol. Environ. Health.* A. 78: 506-515.

Changpetch, P. and Akinyemi, M. I. 2020. An analysis of health insurance data using the directed acyclic graph: An application in Nigeria. *Open Public Health J.* 13: 718-722.

Changpetch, P. and Haughton, D. 2018. Sociological mechanisms underlying alcohol, tobacco and gambling: A causal mediation analysis. *Thailand Statistician.* 16: 56-63.

Chantara, S., Sillapapiromsuk, S. and Wiriya, W. 2012. Atmospheric pollutants in Chiang Mai (Thailand) over a five-year period (2005-2009), their possible sources and relation to air mass movement. *Atmospheric Environ.* 60: 88-98.

Chen, C. C., Tsai, S. S. and Yang, C. Y. 2015. Association between fine particulate air pollution and daily clinic visits for migraine in a subtropical city: Taipei, Taiwan. *Int. J. Environ. Res. Public Health.* 12: 4697-4708.

Chowdhury, S., Wang, R., Yu, Q., Huntoon, C. J., Karnitz, L. M., Kaufmann, S. H., Gygi, S. P., Birrer, M. J., Paulovich, A. G., Peng, J. and Wang, P. 2020. DAGBagM: Learning directed acyclic graphs of mixed variables with an application to identify prognostic protein biomarkers in ovarian cancer. *BioRxiv.*

Haigh, M. S. and Bessler, D. A. 2004. Causality and price discovery: An application of directed acyclic graphs. *Journal of Business.* 77: 1099-1121.

Harrison, R. M., Deacon, A. R., Jones, M. R. and Appleby, R. S. 1997. Sources and process affecting concentrations of PM10 and PM2.5 particulate matter in Birmingham (U.K.). *Atmospheric Environ.* 31: 4103-4117.

Hassija, V., Chamola, V., Han, G., Rodrigues, J. J. P. C. and Guizani, M. 2020. DAGIoV: A framework for vehicle to vehicle communication using directed acyclic graph and game theory. *IEEE Transactions on Vehicular Technology.* 69: 4182-4191.

Jeensorn, T., Apichartwiwat, P. and Jinsart, W. 2018. PM10 and PM2.5 from haze smog and visibility effect in Chiang Mai Province Thailand. *App. Env. Res.* 40: 1-10.

Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Zhang, X. and Sachdeva, S. 2019. Evaluation of different machine learning approaches to forecasting PM2.5 mass concentrations. *Aerosol Air Qual. Res.* 19: 1400-1410.

Khan, M. F., Latif, M. T., Saw, W. H., Amil, N., Nadzir, M. S. M., Sahani, M., Tahir, N. M. and Chung, J. X. 2016. Fine particulate matter in the tropical environment: Monsoonal effects, source apportionment, and health risk assessment. *Atmos. Chem. Phys.* 16: 597-617.

Lal, A., Li, G., Cubro, E., Chalmers, S., Li, H., Herasevich, V., Dong, Y., Pickering, B. W., Kilickaya, O. and Gajic, O. 2020. Development and verification of a digital twin patient model to predict specific treatment response during the first 24 hours of sepsis. *Crit. Care. Explor.* 16: e0249.

Liu, C., Henderson, B. H., Wang, D., Yang, X. and Peng Z. 2016. A land use regression application into assessing spatial variation of intra-urban fine particulate matter (PM2.5) and nitrogen dioxide concentrations in city of Shanghai, China. *Sci. Total Environ.* 565: 607-615.

Marlier, M. E., DeFries, R. S., Voulgarakis, A., Kinney, P. L., Randerson, J. T., Shindell, D. T., Chen, Y. and Faluvegi, G. 2013. El Niño and health risks from landscape fire emissions in Southeast Asia. *Nat. Clim. Chang.* 3: 131-136.

Pavagadhi, S., Betha, R., Venkatesan, S., Balasubramanian, R. and Hande, M.P. 2013. Physicochemical and toxicological characteristics of urban aerosols during a recent Indonesian biomass burning episode. *Environ. Sci. Pollut. Res.* 20: 2569-2578.

Reddington, C. L., Yoshioka, M., Balasubramanian, R., Ridley, D., Toh, Y. Y., Arnold, S. R. and Spracklen, D. V. 2014. Contribution of vegetation and peat fires to particulate air pollution in Southeast Asia. *Environ. Res. Lett.* 9: 094006.

Szyszkowicz, M., Kaplan, G. G., Grafstein, E. and Rowe,

B. H. 2009. Emergency department visits for migraine and headache: A multi-city study. *Int. J. Occup. Med. Environ. Health.* 22: 235-242.

Thepnuan, D., Chantara, S., Te Lee, C., Lin, N. H. and Tsai, Y. I. 2019. Molecular markers for biomass burning associated with the characterization of PM2.5 and component sources during dry season haze episodes in Upper South East Asia. *Sci. Total Environ.* 658: 708-722.

Velasco, E. and Roth, M. 2012. A Review of Singapore's air quality and greenhouse gas emissions: Current situation and opportunities. *J. Air & Waste Management Assn.* 62: 625-641.

Williams, T. C., Bach, C. C., Matthiesen, N. B., Henrikson, T. B. and Gagliardi, L. 2018. Directed acyclic graphs: A tool for causal studies in pediatrics. *Pediatric Res.* 84: 487-493.

Yang, C., Peng, X., Huang, W., Chen, R., Xu, Z., Chen, B. and Kan, H. 2012. A time-stratified case-crossover study of fine particulate matter air pollution and mortality in Guangzhou, China. *Int. Arch. Occup. Environ. Health.* 85: 579-585.

Yang, Z. and Zhao, Y. 2014. Energy consumption, carbon emissions and economic growth in India: Evidence from directed acyclic graphs. *Economic Modelling.* 38: 533-540.

Yap, J., Ng, Y., Yeo, K. K., Sahlén, A., Lam, C. S. P., Lee, V. and Ma, S. 2019. Particulate air pollution on cardiovascular mortality in the Tropics: Impact on the elderly. *Environ. Health.* 18: 34-39.

Yuchi, W., Gombojav, E., Boldbaatar, B., Galsuren, J., Enkhmaa, S., Beejin, B., Naidan, G., Ochir, C., Legtseg, B., Byambaa, T., Barn, P., Henderson, S. B., Janes, C. R., Lanphear, B. P., McCandless, L. C., Takaro, T. K., Venners, S. A., Webster, G. M. and Allen, R. W. 2019. Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. *Environ. Pollut.* 1: 746-753.