

PERBANDINGAN KAEDAH DALAM PENGANGGARAN NILAI LENYAP BAGI DATA ZARAH PENCEMAR PM₁₀ DAN O₃ (Comparing Methods on Estimating the Missing Values of PM₁₀ and O₃ Data)

SALWA ISMAIL & ZAMIRA HASANAH ZAMZURI

ABSTRAK

Pencemaran udara secara amnya disebabkan oleh kehadiran sekurang-kurangnya satu bahan kimia di atmosfera dalam jumlah yang boleh membahayakan kesihatan. Nilai lenyap sering terjadi dalam data kualiti udara yang biasanya digunakan semasa menjalankan kajian mengenai alam sekitar. Nilai lenyap biasanya berlaku akibat pensampelan tidak mencukupi, ralat dalam ukuran atau kesalahan semasa pengumpulan data. Keadaan ini memberi kesan pada data dan seterusnya menyebabkan analisis statistik diragui. Justeru, kajian ini dijalankan bagi menyelesaikan masalah dan membuat perbandingan kaedah untuk menaksir nilai lenyap dengan menggunakan kaedah kesan tempat bersandar, min baris dan min jam. Lokasi kajian ini adalah Sek. Men. Raja Perempuan Zarina, Klang; Sek. Keb. Indera Mahkota, Pahang; Sek. Men. Pagoh, Ipoh; dan Sek. Men. Dato' Permaisuri, Miri. Manakala zarah pencemar yang digunakan adalah PM₁₀ dan O₃. Pemilihan kaedah terbaik dinilai menggunakan tiga penunjuk prestasi, iaitu indeks persetujuan, punca kuasa dua min ralat dan min ralat mutlak. Ketiga-tiga prestasi penunjuk memilih kaedah kesan tempat bersandar sebagai kaedah terbaik untuk menganggar nilai lenyap.

Kata kunci: nilai lenyap; penunjuk prestasi; kaedah kesan tempat bersandar

ABSTRACT

Air pollution is generally caused by the presence of detrimental amounts of chemicals in the atmosphere. Nonetheless, the missing value is now a problem in air quality data which is important in environmental research. Usually, missing value is caused by inadequate sampling as well as error in taking measurement and data collection. Hence, this study is performed with the objective to impute missing value and to compare between methods; specifically the site-dependent effect method, row mean and hour mean methods. The selected areas were Sek. Men. Raja Perempuan Zarina, Klang; Sek. Keb. Indera Mahkota, Pahang; Sek. Men. Pagoh, Ipoh; and Sek. Men. Dato' Permaisuri, Miri. Meanwhile, the air pollution data used in this study were PM₁₀ and O₃. Selection of the best methods is assessed using three performance indicators, namely index of agreement, root mean square error and the mean absolute deviation. At the end of this study, the three performance indicators show that the site-dependent effect method is the best for estimating missing values.

Keywords: missing value; performance indicator; site-dependent effect method

1. Pendahuluan

Nilai lenyap terjadi apabila terdapat data tidak lengkap semasa pengumpulan data dilakukan. Pengumpulan data biasanya dilakukan dengan pelbagai cara; antaranya kaji selidik, pengukuran, pemerhatian dan sebagainya. Nilai lenyap mampu memberi masalah bagi penyelidik atau pengkaji untuk membuat kajian sepenuhnya tentang data. Masalah bagi penyelidikan berlaku kerana ketidakcukupan pensampelan, ralat dalam ukuran atau kesalahan semasa memperoleh data. Contoh keadaan berlakunya nilai lenyap adalah pengumpulan data kaji selidik apabila adanya responden yang tidak menjawab semua soalan dalam soal selidik.

Manakala nilai lenyap semasa proses kemasukan data secara manual mungkin terjadi apabila kelalaian pengkaji, atau pengukuran yang tidak tepat. Selain itu, eksperimen gagal juga merupakan salah satu faktor nilai lenyap berlaku (Kaiser 2014).

Pencemaran udara secara amnya disebabkan kehadiran satu atau lebih bahan kimia di atmosfera dalam jumlah yang membahayakan kesihatan. Antara bahan pencemar udara adalah PM_{10} dan O_3 , yang PM_{10} adalah zarah pencemar yang lebih kecil daripada 10 mikron (Noor *et al.* 2013). Manakala O_3 pula adalah gas bertoksik dengan pengoksidaan tinggi yang mana merbahaya kepada kesihatan (Sheffield *et al.* 2015). Data lenyap atau tidak lengkap adalah masalah yang kerap berlaku dalam kajian alam sekitar (Nurulkamal *et al.* 2013).

Corak data lenyap pada setiap data kualiti udara pada setiap lokasi atau zarah pencemar biasanya berbeza (Nuryazmin *et al.* 2015). Junninen *et al.* (2004) telah mengelaskan data lenyap kepada tiga selang, iaitu data lenyap tidak lebih dari tiga jam, data lenyap antara tiga jam hingga 24 jam dan data lenyap lebih dari 24 jam. Manakala dalam kajian Plaia dan Bondi (2006), pembahagian panjang jurang adalah empat bahagian, iaitu lenyap satu nilai, lenyap satu hingga tiga nilai, lenyap lebih dari tiga nilai hingga 12 nilai dan lenyap lebih dari 12 nilai. Kajiannya telah menggunakan data PM_{10} bagi lapan buah stesen di kawasan metropolitan Palermo, Sicily bagi tahun 2003. Kaedah yang telah dicadangkan dalam kajian terbabit adalah kaedah kesan tempat yang bersandar (SDEM, daripada *site-dependent effect method*) yang telah menggunakan maklumat ruang dan masa dalam setiap kepekatan PM_{10} dalam lapan tempat pemantauan kajian. Kaedah ini adalah kaedah terbaik antara kaedah lain di dalam kajian dengan setiap indikator menunjukkan SDEM adalah terbaik yang tidak mengambil kira panjang jurang dan stesen berbanding dengan kaedah lain, iaitu taksiran min yang menggunakan min jam, min mengikut baris, taksiran berganda dan min yang lepas dan akan datang. Tambahan pula, SDEM menunjukkan perubahan yang sedikit pada penunjuk indikator di setiap stesen.

Beberapa indeks penunjuk dinilai untuk menentukan kebagusan taksiran (Junninen *et al.* 2004). Kebiasaannya, indeks yang digunakan adalah pekali korelasi (R) dan kuasa dua kepada pekali korelasi, iaitu pekali penentuan (R^2) yang merupakan varians yang diterangkan berjulat 0 sehingga satu. Pekali persetujuan (d) juga digunakan kerana nilai indikator korelasi tidak berhubung dengan saiz perbezaan antara nilai cerapan dan nilai anggaran. Punca kuasa dua min ralat (RMSE, daripada *root mean square error*) juga digunakan dalam kajian ini yang menerangkan perbezaan nilai cerapan dan nilai taksiran yang digunakan untuk menghasilkan purata ralat model. Satu lagi ukuran yang diambil kira adalah min ralat mutlak (MAD, daripada *mean absolute deviation*) yang menunjukkan purata perbezaan antara nilai cerapan dan nilai taksiran.

2. Data dan Skop Kajian

Dua zarah pencemar yang berbeza dipilih bagi membezakan kaedah yang lebih sesuai digunakan bagi setiap zarah udara. Pelaksanaan sesuatu kaedah mungkin berbeza bagi setiap masalah yang ingin diselesaikan dan corak kepekatan zarah pencemar perlu difahami dengan baik. Bagi permasalahan kajian ini, rekod data PM_{10} dan O_3 mengikut jam dan bulan pada tahun 2011 daripada setiap pusat cerapan kaji cuaca diperolehi dari Jabatan Alam Sekitar Malaysia. Empat lokasi telah dipilih, iaitu Sek. Men. Perempuan Raja Zarina, Klang; Sek.

Keb. Indera Mahkota, Pahang; Sek. Men. Pagoh, Ipoh; dan Sek. Men. Dato' Permaisuri, Miri. Pemilihan lokasi adalah secara rawak, iaitu setiap stesen dilabelkan dengan nombor dan menggunakan perisian R, empat nombor rawak dijana mewakili empat buah stesen terpilih. Data keseluruhan tahun 2011 bagi setiap tempat adalah 8760, namun terdapat nilai lenyap bagi setiap tempat dan nilai lenyap adalah direkodkan mengikut jam. Taburan nilai lenyap bagi PM₁₀ dan O₃ ditunjukkan dalam Jadual 1.

Jadual 1: Maklumat nilai lenyap zarah pencemar PM₁₀ dan O₃

TEMPAT	PM ₁₀		O ₃	
	Bilangan Nilai Lenyap	Peratus Nilai Lenyap (%)	Bilangan Nilai Lenyap	Peratus Nilai Lenyap (%)
Sek. Men. Perempuan Raja Zarina, Klang	134	1.53	529	6.04
Sek. Keb. Indera Mahkota, Pahang	340	3.88	577	6.59
Sek. Men. Dato' Permaisuri, Miri	560	6.28	555	6.22
Sek Men. Pagoh, Ipoh	271	3.09	474	5.41

Corak data telah dibahagikan kepada empat selang, lenyap pada satu nilai, lenyap antara dua nilai hingga lima nilai secara berturut-turut, lenyap antara enam nilai hingga 10 nilai berturut-turut dan lenyap lebih dari 10 nilai berturut-turut. Corak data lenyap yang telah dibahagikan dalam selang telah dirumuskan dalam Jadual 2 bagi zarah pencemar PM₁₀ dan Jadual 3 bagi zarah pencemar O₃. Secara keseluruhannya dapat dilihat satu nilai lenyap sahaja adalah paling banyak peratusannya dengan min 72.53% bagi zarah pencemar PM₁₀ dan 89.57% bagi zarah pencemar O₃. Peratus data lenyap kedua tertinggi pula diikuti dengan nilai lenyap antara 2 sehingga 5, iaitu 19.19% bagi zarah pencemar PM₁₀ dan 9.37% bagi zarah pencemar O₃. Manakala bagi peratus data lenyap yang ketiga tertinggi bagi zarah pencemar PM₁₀ adalah lebih dari 10 nilai dan peratus terendah adalah data lenyap antara nilai 6 sehingga 10 nilai berturut-turut. Bagi zarah pencemar O₃ pula, peratus ketiga tertinggi adalah nilai antara enam hingga sepuluh nilai dengan peratus 0.59% dan peratus paling rendah lenyap adalah 0.48% bagi nilai lenyap lebih dari sepuluh nilai. Maka, dapat disimpulkan di sini bagi kedua-dua zarah pencemar nilai lenyap paling tinggi adalah pada selang satu sahaja.

Jadual 2: Peratusan data lenyap dan statistik deskriptif PM₁₀ mengikut empat kawasan pemerhatian

Lokasi	Peratus Lenyap %	Selang Kekosongan Data Secara Berturutan (l)				Data Mentah	
		l=1	2 <=l<=5	6<=l<=10	l>10	Min	Sisihan Piawai
Kawasan 1	1.53	71.70	16.98	5.66	5.66	63.00	33.32
Kawasan 2	3.88	61.76	29.41	3.92	4.90	38.00	16.59
Kawasan 3	3.09	74.65	14.08	0.00	11.27	48.90	23.13
Kawasan 4	6.39	82.00	16.29	0.57	1.14	32.19	18.82
Min	3.72	72.53	19.19	2.54	5.74		

Jadual 3: Peratusan data lenyap dan statistik deskriptif O₃ mengikut empat kawasan pemerhatian

Lokasi	Peratus Lenyap	Selang Kekosongan Data Secara Berturutan (l)				Data Mentah	
	%	l=1	2 <=l<=5	6<=l<=10	l>10	Min	Sisihan Piawai
Kawasan 1	6.04	89.98	8.62	1.17	0.23	0.01902	0.01679
Kawasan 2	6.59	89.93	8.92	0.46	0.69	0.01789	0.01311
Kawasan 3	5.41	90.17	9.34	0.49	0.00	0.01706	0.01794
Kawasan 4	6.34	88.18	10.59	0.25	0.99	0.01436	0.00937
Min	6.10	89.57	9.37	0.59	0.48		

3. Kaedah Kajian

3.1 Penggantian nilai tunggal

Pelbagai kaedah telah digunakan oleh penyelidik untuk menangani masalah data lenyap ini; antaranya adalah penggantian nilai min dalam data lenyap yang lazimnya digunakan dan ditawarkan dalam banyak pakej perisian statistik (Luengo *et al.* 2011). Penggantian min biasanya akan mengganggu struktur yang sedia ada pada data dengan ketara dan membawa kepada ralat yang besar dalam matriks kovarians atau korelasi dan ia secara tidak langsung melemahkan prestasi sesuatu pemodelan statistik. Pemansuhan nilai lenyap juga adalah salah satu kaedah yang sering digunakan terutamanya dalam bidang psikologi dan pendidikan, namun kaedah ini tidak begitu digalakkan kerana ia memberi kesan anggaran kurang tepat dalam analisis.

Kaedah penggantian juga adalah salah satu kaedah untuk menangani masalah data lenyap dan pelbagai kaedah penggantian yang biasa dilakukan melibatkan nilai lenyap ini (Fang 2010). Kaedah penggantian untuk memenuhi data matriks secara umumnya boleh dibahagikan kepada dua kategori utama, penggantian nilai tunggal dan penggantian pelbagai nilai (Little & Rubin 1987). Kaedah penggantian nilai tunggal menganggar nilai tunggal pada setiap data yang lenyap. Ia mempunyai ciri-ciri menarik, iaitu sangat mudah untuk melengkapkan data secara langsung dan ia dapat mencipta penggantian dengan melakukannya hanya sekali sahaja. Kaedah Penggantian Pelbagai menjana pelbagai nilai simulasi bagi setiap nilai lenyap untuk memeriksa ketidakpastian yang dikaitkan dengan nilai lenyap tersebut. Dalam kajian ini, kaedah penggantian nilai tunggal telah digunakan untuk menganggarkan nilai lenyap. Tiga kaedah dipilih daripada kajian Plaia dan Bondi (2006), iaitu kaedah kesan tempat bersandar, kaedah min baris (RM, daripada *row mean*) dan kaedah min jam (HM, daripada *hour mean*).

(i) Kaedah Kesan Tempat Bersandar (SDEM)

Kaedah SDEM merupakan kaedah baharu yang dicadangkan oleh Plaia dan Bondi (2006). Kaedah ini menggunakan maklumat masa dan ruang bagi tahap konsentrasi pada lapan tempat pemerhatian. Kajian ini menggunakan maklumat ruang dan masa berdasarkan data tahun 2011 yang dibahagikan kepada empat kategori, iaitu empat sukuan tahunan. Set data bagi setiap kawasan dan nilai zarah pencemar diambil untuk dibuat analisis dan dilabel sebagai x_{swdh} dengan s adalah kawasan pemerhatian, $s = 1, 2, 3, \dots, s$ dan jumlah s yang digunakan adalah

mengikut suku tahun, iaitu empat suku tahun 2011. Manakala w adalah minggu dan minggu yang terdapat dalam setiap suku adalah sebanyak 13, $w = 1, 2, 3, \dots, 13$. Hari dalam seminggu pula diwakili oleh d , iaitu $d = 1, 2, 3, \dots, 7$ dan h pula adalah jam dalam sehari, iaitu $h = 1, 2, 3, \dots, 24$. Maka, singkatan-singkatan ini digunakan dalam mengira min-min berikut:

- a) Matriks min kawasan dalam minggu, iaitu matriks 13×4 . Matriks ini berunsur $\bar{x}_{sw..}$, iaitu min kepada cerapan pada minggu w dalam kawasan s . Perbezaan antara $\bar{x}_{sw..}$ dengan baris marginalnya, $\sum_{s=1}^S \frac{\bar{x}_{sw..}}{S}$ (plot garis min minggu pada Rajah 1 dan 2) boleh dianggap sebagai kesan khusus pada minggu untuk kawasan s .
- b) Matriks min kawasan harian: Matriks ini adalah matriks 7×4 yang melibatkan hari dan kawasan. Unsur matriks ini ialah $\bar{x}_{s.d.}$, iaitu min kepada nilai cerapan pada hari dalam seminggu d untuk kawasan s . Kesan khusus hari dalam seminggu berdasarkan kepada kawasan diperoleh dengan melihat perbezaan di antara $\bar{x}_{s.d.}$ dengan min baris marginalnya $\sum_{s=1}^S \frac{\bar{x}_{s.d.}}{S}$ (plot garis min harian pada Rajah 1).
- c) Matriks min kawasan dalam jam: Matriks ini ialah matriks 24×4 yang menandakan unsur $\bar{x}_{s..h}$, iaitu min nilai cerapan pada setiap jam dalam sehari h ($h = 1, 2, 3, \dots, 24$) untuk kawasan s . Kesan jam tertentu pada setiap kawasan dapat dilihat melalui perbandingan antara $\bar{x}_{s..h}$ dengan min baris marginalnya $\sum_{s=1}^S \frac{\bar{x}_{s..h}}{S}$ dan boleh dilihat (plot garis min jam pada Rajah 1).

Kaedah ini mengambil kira kesan minggu, kesan hari dan kesan masa setiap jam (untuk semua kawasan), dan menganggarkan suatu nilai lenyap sebagai:

$$\hat{x}_{swdh} = \bar{x}_{wdh} + \frac{1}{2} \left(\bar{x}_{sw..} - \sum_{s=1}^S \frac{\bar{x}_{sw..}}{S} \right) + \frac{1}{2} \left(\bar{x}_{s.d.} - \sum_{s=1}^S \frac{\bar{x}_{s.d.}}{S} \right) + \frac{1}{2} \left(\bar{x}_{s..h} - \sum_{s=1}^S \frac{\bar{x}_{s..h}}{S} \right) \quad (1)$$

(ii) Kaedah Min Masa (HM)

Kaedah ini menggunakan maklumat jam pada nilai PM_{10} dan O_3 dalam kawasan yang sama. Menurut Li *et al.* (1999), kaedah ini menggantikan kelenyapan data secara per jam berdasarkan kepada min keseluruhan pada jam tertentu dengan kawasan yang tertentu untuk keseluruhan tahun. Penganggaran nilai untuk digantikan kepada data yang lenyap adalah seperti berikut:

$$\hat{x}_{swdh} = \bar{x}_{s..h}, \quad (2)$$

dengan $s = 1, 2, \dots, S$, $h = 1, 2, 3, \dots, 24$.

(iii) Kaedah Min Baris (RM)

Kaedah ini menggunakan maklumat nilai PM_{10} atau O_3 pada kawasan yang berlainan (Engels & Diehr 2003). Kaedah ini menggantikan nilai yang lenyap berdasarkan min semua cerapan yang diketahui pada baris yang sama untuk data matriks tersebut, iaitu menjanakan nilai min semua cerapan yang diketahui pada masa yang sama untuk kawasan yang berlainan:

$$\hat{x}_{swdh} = \bar{x}_{.wdh}, \quad (3)$$

dengan $w = 1, 2, \dots, 13$, $d = 1, 2, 3, \dots, 7$ $h = 1, 2, 3, \dots, 24$

3.2 Teknik penggantian

Kajian ini menggunakan data set lengkap dan sebahagian nilai dilenyapkan. Kemudian, kaedah penggantian nilai lenyap digunakan dan seterusnya perbandingan dengan nilai cerapan atau nilai asal dibuat. Indeks penunjuk akan digunakan bagi menentukan kaedah terbaik bagi kajian ini.

3.3 Indeks penunjuk

Tiga indeks penunjuk telah diputuskan untuk menunjukkan penggantian terbaik nilai lenyap. Setiap kaedah itu harus diuji kecekapannya supaya dapat dilihat sebagai terbaik di antara semua kaedah. Kajian ini menggunakan tiga indeks penunjuk, iaitu indeks persetujuan, punca kuasa dua ralat dan min ralat mutlak. Cerapan O_i ialah data ke- i , \bar{O} ialah min kepada data yang dicerap, P_i ialah data ke- i yang diganti, \bar{P} ialah purata data yang diganti, σ_o ialah sisihan piawai data yang dicerap, manakala σ_p ialah sisihan piawai bagi data yang digantikan dengan teknik penggantian, serta N ialah bilangan penggantian yakni bilangan data lenyap (Junninen *et al.* 2004).

(i) Indeks Persetujuan (d)

Rumus untuk pengiraan indeks persetujuan (Willmot *et al.* 2012) adalah seperti berikut:

$$d = 1 - \left[\frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \right]. \quad (4)$$

(ii) Punca Kuasa Dua Min Ralat (RMSD)

Rumus pengiraan Punca Kuasa Dua Min Ralat (Junninen *et al.* 2004) adalah seperti berikut:

$$\text{RMSD} = \left(\frac{1}{N} \sum_{i=1}^N [O_i - P_i]^2 \right)^{1/2}. \quad (5)$$

(iii) Min Ralat Mutlak (MAD)

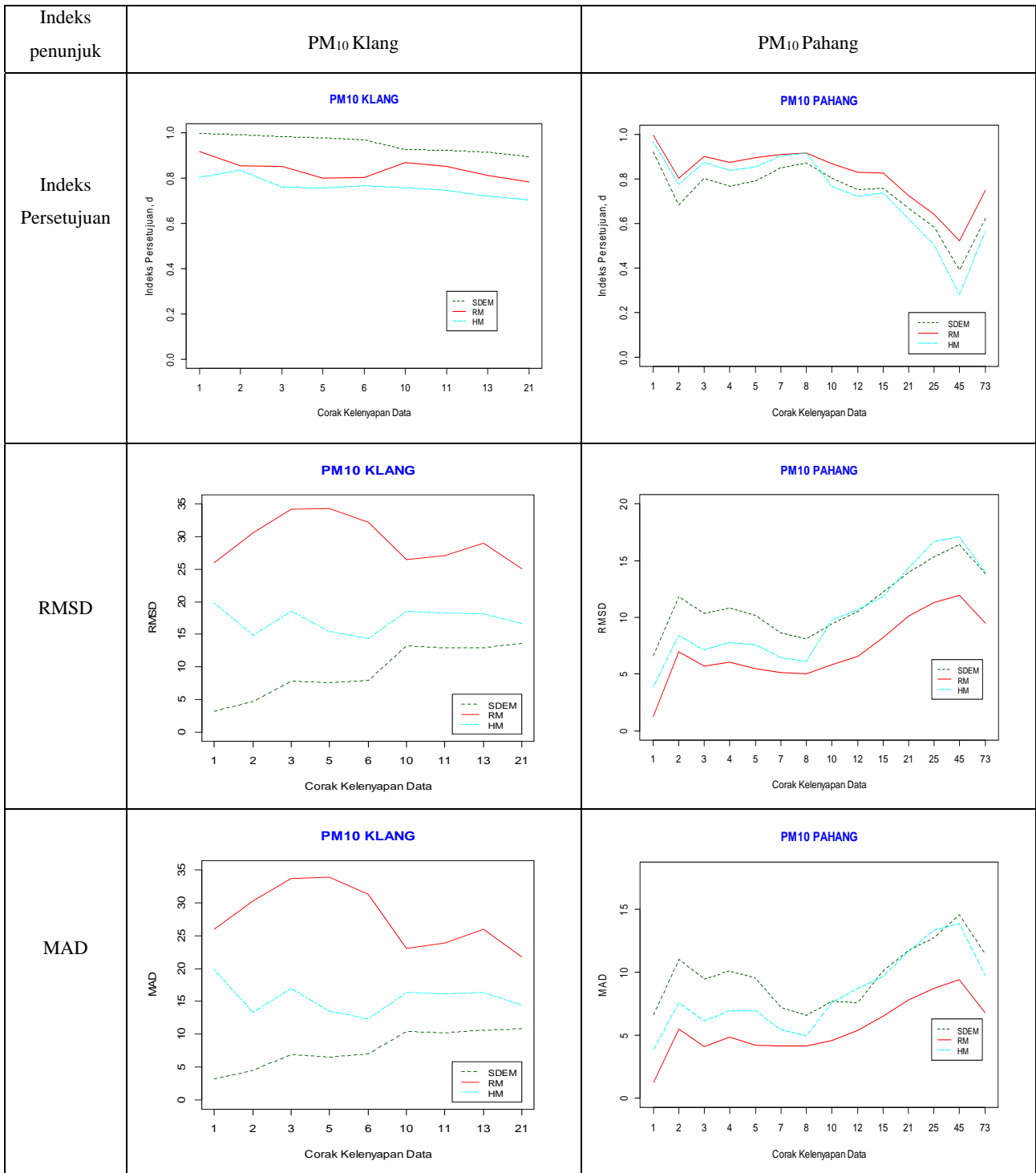
Rumus pengiraan Min Ralat Mutlak (Junninen *et al.* 2004) adalah seperti berikut:

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^N |O_i - P_i|. \quad (6)$$

4. Hasil Kajian

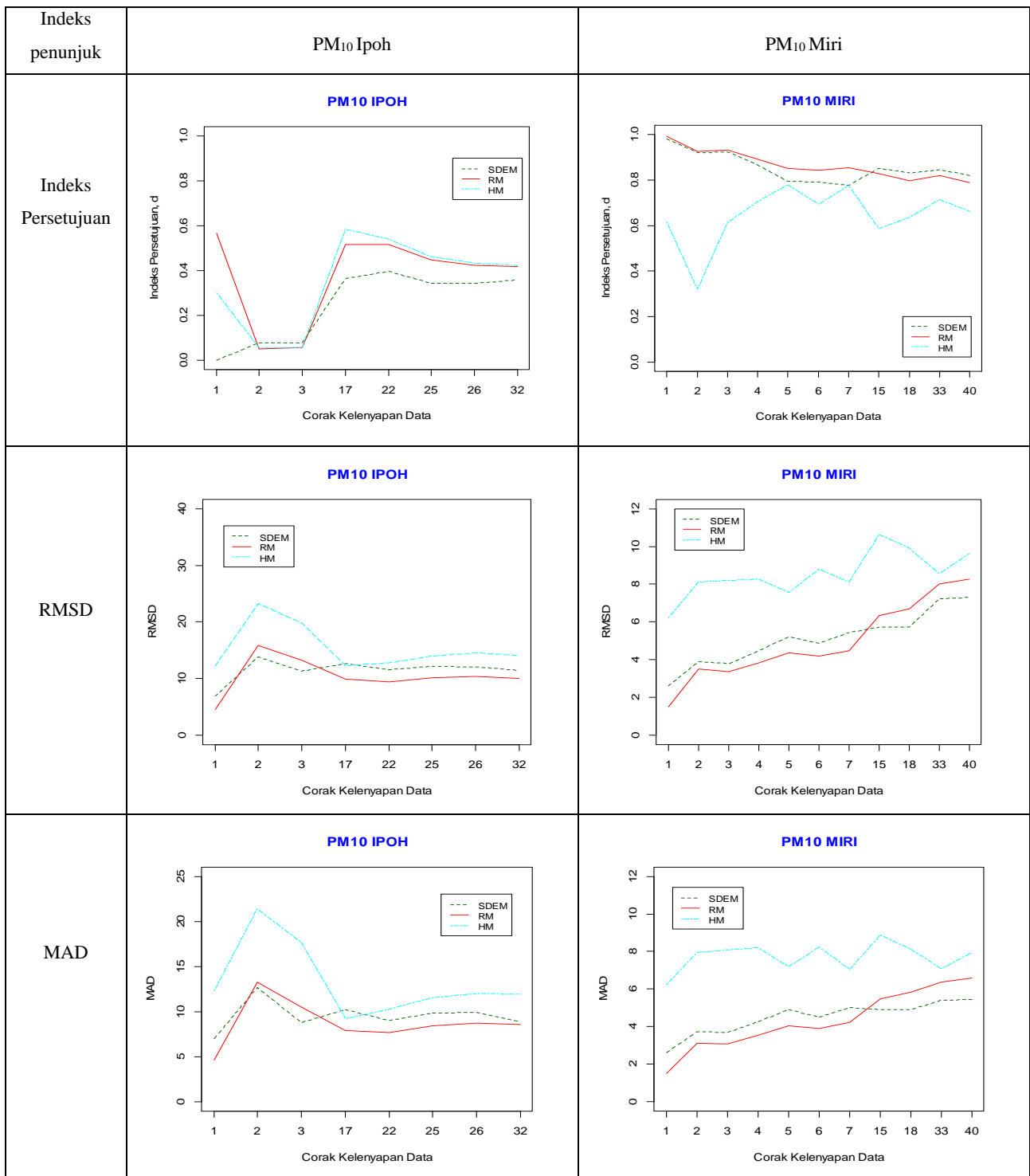
Tiga kaedah yang telah dinyatakan, iaitu Kaedah Kesan Tempat Bersandar (SDEM), Min Baris (RM) dan Min Jam (HM) telah digunakan untuk menganggar nilai lenyap bagi data PM_{10} dan O_3 . Perbandingan antara tiga kaedah ini dilakukan dengan melihat kepada nilai indeks penunjuk. Semakin besar nilai indeks persetujuan (d), semakin baik nilai anggaran tersebut, yakni menghampiri nilai sebenar. Manakala bagi RMSD dan MAD, kaedah dengan nilai yang paling kecil adalah lebih baik. Prestasi ketiga-tiga kaedah dalam menganggarkan nilai lenyap bagi elemen pencemar PM_{10} di empat buah stesen yang berbeza ditunjukkan dalam Rajah 1 dan 2. Corak kelenyapan (paksi- x) dalam kedua-dua rajah ini merujuk kepada bilangan nilai lenyap berturutan dalam data bagi lokasi tersebut. Bilangan nilai lenyap berturutan adalah berbeza mengikut lokasi, sebagai contoh, bagi Klang, terdapat 21 nilai lenyap manakala untuk Pahang, bilangannya lebih tinggi, iaitu 73.

Keputusan secara keseluruhan bagi perbandingan kaedah ditunjukkan dalam Jadual 4. Jurang pendek merujuk kepada purata jurang kurang dari 12 jam dan sebaliknya untuk jurang panjang. Bagi zarah pencemar PM_{10} dan O_3 di Klang, SDEM dilihat sesuai kerana ia mempunyai d yang tinggi, RMSD dan MAD yang rendah bagi setiap corak nilai lenyap. Manakala RM dilihat terbaik bagi dua tempat, iaitu Kuantan dan Ipoh bagi PM_{10} dan HM bagi O_3 pada kawasan tersebut. Dapat dilihat bahawa penentuan kaedah terbaik adalah berbeza-beza yang mana bergantung kepada faktor lokasi stesen dan jurang bagi nilai lenyap tersebut. Dalam kajian ini, RM dan SDEM dilihat terbaik secara keseluruhannya kerana mempunyai jumlah yang sama bagi keputusan taksiran jika tidak melihat pada jurang namun SDEM mengambil kira min jam, min baris dan min hari bagi setiap stesen dan ini tentunya memberi lebih pengaruh dalam penganggaran nilai lenyap.



Rajah 1: Keputusan PM₁₀ bagi kawasan Klang dan Pahang

Perbandingan kaedah dalam penganggaran nilai lenyap bagi data zarah pencemar PM_{10} dan O_3



Rajah 2: Keputusan PM_{10} bagi kawasan Ipoh dan Miri

Jadual 4: Keputusan taksiran

Tempat	PM ₁₀		O ₃	
	Jurang pendek	Jurang panjang	Jurang pendek	Jurang panjang
Klang	SDEM	SDEM	SDEM	HM
Kuantan	RM	RM	HM	HM
Ipoh	RM	RM	RM	HM
Miri	RM	SDEM	SDEM	SDEM

5. Kesimpulan

Kajian ini membandingkan tiga kaedah dalam menganggarkan nilai lenyap bagi data zarah pencemar. Dua zarah pencemar diambil kira, iaitu PM₁₀ dan O₃ daripada empat buah stesen terpilih. Data dilenyapkan dan kemudian dianggarkan menggunakan tiga kaedah di atas. Penentuan kaedah terbaik dilihat melalui tiga indeks, iaitu indeks persetujuan (*d*), RMSD dan MAD. Keputusan menunjukkan kaedah terbaik dalam menganggarkan nilai lenyap adalah berbeza antara stesen dan juga jurang bagi corak nilai lenyap. Secara keseluruhannya, kaedah SDEM merupakan kaedah terbaik yang mana ia mengambil kira lebih banyak maklumat berbanding dua kaedah lain. Secara amnya, dalam menganggarkan nilai lenyap bagi data zarah pencemar, lebih dari satu kaedah perlu diambil kira agar anggaran yang lebih tepat dapat diperolehi.

Penghargaan

Penghargaan kepada Jabatan Alam Sekitar Malaysia kerana membekalkan data yang digunakan dalam kajian ini.

Rujukan

- Engels J.M. & Diehr P. 2003. Imputation of missing longitudinal data: a comparison of methods. *Journal of Clinical Epidemiology* **10**: 968-976
- Fang L.P. 2010. Perawatan data lenyap dengan kaedah penggantian bagi kepekatan karbon monoksida di Sri Petaling pada tahun 2006. Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia.
- 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 .
- Kaiser J. 2014. Dealing with missing values in data. *Journal of System Integration* **5**(1): 1804-2724.
- Li K.H., Le N.D., Sun, L. & Zidek J.V. 1999. Spatial-temporal models for ambient hourly PM₁₀ in Vancouver. *Environmetrics* **10**: 321-328.
- Little R.J.A. & Rubin D.B. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.
- Luengo J., Garcia S. & Herrera F. 2011. *On the Choice of The Best Imputation Methods for Missing Values Considering Three Groups of Classifications Methods*. London: Springer-Verlag.
- Noor M.N., Bakri A. M., Yahaya A.S., Ramli N.A. & Fitri N.F.M.Y. 2013. Estimation of Missing Values in Environmental data Set using Interpolation Technique: Fitting on Lognormal Distribution. *Australian Journal of Basic and Applied Sciences* **7**(5): 336-341.
- Nurulkamal M., Razali A.M. & Ibrahim K. 2013. Application of the single imputation method to estimate missing wind speed data in Malaysia. *Research Journal of Applied Sciences, Engineering and Technology* **6**(10): 1780-1784.

- Nuryazmin A.Z., Jemain A.A. & Muda N. 2015. A comparison of various imputation methods for missing values in air quality data. *Sains Malaysiana* **44**:449-456.
- Plaia A. & Bondi A.L. 2006. Single Imputation method of missing values in environmental pollution data sets. *Atmospheric Environment* **40**: 7316-7330 .
- Sheffield P. E., Zhou J., Shmool J.L.R. & Clougherty J.E. 2015. Ambient ozone exposure and children's acute asthma in New York city: a case-crossover analysis. *Environmental Health* **14**:25
- Willmott C.J., Robeson S.M. & Matsuura K. 2012. Short Communication: A Refined Index of Model Performance. *International Journal of Climatology* **32**: 2088-2094.

*Pusat Pengajian Sains Matematik
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor DE, MALAYSIA
Mel-e: salwaismail91@gmail.com* , zamira@ukm.edu.my*

*Penulis untuk dihubungi