

Sains Malaysiana 49(4)(2020): 963-970  
<http://dx.doi.org/10.17576/jsm-2020-4904-25>

## Pendekatan Baharu untuk Mengelompok Stesen Pengawasan Kualiti Udara menggunakan Homologi Gigih

(A New Approach to Cluster Air Quality Monitoring Stations using Persistent Homology)

NUR FARIHA SYAQINA ZULKEPLI\*, MOHD SALMI MD NOORANI, FATIMAH ABDUL RAZAK, MUNIRA ISMAIL & MOHD ALMIE ALIAS

### ABSTRAK

Isu pencemaran udara merupakan masalah global yang terus dibincangkan sehingga kini. Seringkali penggunaan pendekatan kuantitatif seperti analisis kelompok, analisis korelasi dan analisis komponen prinsipal digunakan untuk menganalisis keserupaan pencemaran udara antara stesen. Walau bagaimanapun, kajian berkaitan dengan pendekatan kualitatif khususnya pendekatan topologi untuk menganalisis keserupaan pencemaran udara tidak lagi dipelopori dengan meluas di Malaysia. Oleh itu, penyelidikan ini adalah kajian rintis yang dijalankan untuk mengkaji keserupaan pencemaran udara antara beberapa stesen di Malaysia menggunakan teknik dalam analisis data bertopologi yang dikenali sebagai homologi gigih. Sifat topologi pencemaran udara diperihalkan oleh ciri-ciri topologi seperti komponen berkait, lubang dan lompong. Habuk halus ( $PM_{10}$ ) yang diketahui sebagai pencemar utama digunakan untuk memperihalkan perilaku pencemaran udara di stesen pengawasan kualiti udara Klang, Petaling Jaya dan Shah Alam. Ciri-ciri topologi yang diperolehi daripada  $PM_{10}$  dianalisis menggunakan ukuran jarak (jarak Wasserstein) untuk mendapatkan keserupaan topologi. Darjah keserupaan dicirikan oleh nilai jarak yang kecil dan sebaliknya. Hasil daripada ukuran jarak menunjukkan Petaling Jaya dan Shah Alam adalah stesen yang paling serupa dan Klang adalah stesen yang paling tak serupa. Penentusahan untuk hasil tersebut dijalankan melalui analisis kelompok aglomeratif berhierarki yang mengelompokkan stesen mengikut jarak ketakserupaan dan hasilnya adalah konsisten dengan keputusan kajian ini. Melalui penemuan ini, kajian yang lebih mendalam dengan menggunakan jentera lain dalam bidang analisis data bertopologi boleh dilakukan sebagai kaedah alternatif dalam menganalisis masalah pencemaran udara di Malaysia.

Kata kunci: Analisis data bertopologi; homologi gigih; jarak Wasserstein; keserupaan topologi;  $PM_{10}$

### ABSTRACT

The issue of air pollution is a global problem that continues to be discussed today. Often, the use of quantitative approaches such as cluster analysis, correlation analysis and principal component analysis is used to analyze the similarity of air pollution between stations. However, studies related to qualitative approaches, especially topological approaches to analyzing the similarity of air pollution, have not been widely popularized in Malaysia. Therefore, this study is a pilot study conducted to investigate the similarity of air pollution between several stations in Malaysia using a technique in the topological data analysis known as persistent homology. The topological properties of air pollution are described by topological features such as connected components, holes, and void. The particulate matter ( $PM_{10}$ ) known as the main pollutant is used to describe the air pollution behavior at Klang, Petaling Jaya and Shah Alam air quality monitoring stations. The topological features obtained from the  $PM_{10}$  is analyzed using distance measure (Wasserstein distance) to obtain topological similarity. The degree of similarity is characterized by a small distance value and vice versa. Results of distance measure show that Petaling Jaya and Shah Alam are the most similar stations and Klang is the most dissimilar station. The validation of these results was carried out by analysis of hierarchical agglomerative cluster analysis that grouped the stations according to their dissimilarity distance and the results are consistent with the findings of this study. From this finding, a more in-depth study using other machinery in the field of topological data analysis can be done as an alternative method for analyzing air pollution problems in Malaysia.

Keywords: Persistent homology;  $PM_{10}$ ; topological data analysis; topological similarity; Wasserstein distance

### PENGENALAN

Pencemaran udara adalah isu yang biasa berlaku sehingga kini di Malaysia. Isu yang berterusan ini meningkatkan lagi

risiko untuk manusia terdedah kepada penyakit-penyakit berbahaya seperti asma, bronkitis kronik dan kanser paru-paru (Afroz et al. 2003; Cohen et al. 2017). Tahap

pencemaran udara bergantung kepada sesuatu lokasi dan pemerhatian yang lebih perlu dilakukan di lokasi yang menunjukkan tahap pencemaran yang tinggi. Jenis pencemar yang diawasi secara berterusan oleh stesen pengawasan kualiti udara yang menyumbang kepada peningkatan tahap pencemaran udara adalah terdiri daripada habuk halus ( $PM_{10}$ ), ozon ( $O_3$ ), karbon monoksida (CO), nitrogen dioksida ( $NO_2$ ) dan sulfur dioksida ( $SO_2$ ) (JASM 2019a). Walau bagaimanapun,  $PM_{10}$  dilaporkan sebagai pencemar utama kerana kepekatan  $PM_{10}$  adalah melebihi kepekatan pencemar yang lain (Azmi et al. 2010; JASM 2019a). Oleh itu, kajian ini dijalankan ke atas pencemar  $PM_{10}$  untuk mengkaji tahap pencemaran udara di kawasan Klang, Petaling Jaya dan Shah Alam.

Kebiasaannya, keserupaan pencemaran udara antara stesen pengawasan kualiti udara dianalisis menggunakan pendekatan statistik seperti analisis kelompok, analisis korelasi dan analisis komponen prinsipal (AKP) (Spangl et al. 2007). Melalui analisis kelompok iaitu analisis kelompok agglomeratif berhierarki (AKAB), ciri-ciri keserupaan di antara stesen dihitung menggunakan ukuran jarak berdasarkan kepekatan pencemar dan stesen tersebut akan dikelompokkan berdasarkan nilai jarak tersebut (Austin et al. 2013; Azid et al. 2015; Hua 2018; Iizuka et al. 2014; Isiyaka & Azid 2015). Teknik AKP pula memperihalkan keserupaan stesen melalui kekuatan nilai faktor (Lu et al. 2011) manakala analisis korelasi menentukan kekuatan perhubungan antara stesen melalui nilai korelasi (Jamalani et al. 2016; Zainordin et al. 2017).

Analisis data bertopologi (ADB) diketahui sebagai bidang yang semakin terus berkembang dalam menganalisis set data. Penerokaan analisis data berdasarkan topologi digunakan dalam pelbagai bidang seperti pengenalpastian kumpulan kanser payudara (Nicolau et al. 2011), bidang rangkaian otak (Petri et al. 2014) dan bidang kewangan (Gidea & Katz 2018). Homologi gigih merupakan teknik utama yang digunakan dalam bidang ADB dan telah dikaji oleh ramai penyelidik (Otter et al. 2017). Melalui teknik ini, pengenalpastian bentuk data dicirikan oleh ciri-ciri topologi seperti komponen berkait, lubang dan lompong. Dalam kajian kualiti udara, penerapan homologi gigih dalam menganalisis episod jerebu telah dikaji oleh Zulkepli et al. (2019a, 2019b). Kajian ini adalah rentetan daripada kajian lepas yang menumpukan kepada penggunaan teknik topologi dalam menganalisis keserupaan pencemaran udara antara stesen. Maklumat topologi direkodkan dalam rajah kegigihan dan setiap rajah kegigihan mewakili ringkasan maklumat topologi bagi setiap satu set data (Kerber et al. 2017). Ukuran jarak khususnya ukuran jarak Wasserstein diperlukan untuk mengenal pasti perbezaan antara satu rajah kegigihan dengan rajah kegigihan yang lain. Kaedah ukuran jarak antara rajah kegigihan telah dikaji oleh ramai penyelidik seperti yang dilakukan oleh Seversky et al. (2016) dalam menganalisis data siri masa aktiviti manusia, Kramár et al. (2016) mengkaji siri masa imej

medan aliran dan Hajij et al. (2018) mengkaji evolusi struktur graf masa yang berbeza-beza. Oleh itu, dalam kajian ini pendekatan ini akan diterapkan dalam mengkaji masalah pencemaran udara khususnya di Malaysia dan seterusnya dapat memperluaskan lagi bidang kajian secara kualitatif khususnya pendekatan topologi.

Pendekatan topologi iaitu homologi gigih diterapkan ke atas siri masa purata bulanan kepekatan  $PM_{10}$  di tiga stesen pengawasan kualiti udara, Klang, Petaling Jaya dan Shah Alam daripada bulan Januari 2000 hingga bulan Disember 2015. Ciri-ciri topologi seperti komponen berkait, lubang dan lompong bagi setiap stesen direkodkan dalam rajah kegigihan dan keserupaan antara rajah kegigihan dikira menggunakan ukuran jarak Wasserstein. Nilai jarak yang terendah menunjukkan keserupaan antara stesen adalah tinggi antara stesen yang lain. Penentuan bagi hasil kajian ini dilakukan menggunakan analisis kelompok, AKAB dan perbandingan antara dua pendekatan ini dilakukan untuk mengenalpasti keberkesanan teknik yang dikaji dalam kajian ini.

#### DATA

Siri masa purata bulanan data  $PM_{10}$  dari bulan Januari 2000 hingga Disember 2015 di stesen pengawasan kualiti udara Klang, Petaling Jaya dan Shah Alam diperoleh daripada Jabatan Alam Sekitar Malaysia (JASM). Ketiga-tiga stesen itu dipilih kerana terletak dalam kawasan Lembah Klang yang terkenal sebagai kawasan pembangunan ekonomi utama di Malaysia (Abdullah et al. 2012). Stesen pengawasan kualiti udara Klang terletak di kawasan Pelabuhan Klang yang merupakan salah satu kawasan perindustrian utama di Malaysia. Begitu juga dengan stesen Shah Alam yang terletak di kawasan perindustrian tetapi aliran trafik di kawasan tersebut kurang berbanding dengan kawasan stesen di Petaling Jaya dan Klang (Ling et al. 2014). Di kawasan Petaling Jaya, stesen pengawasan berada di kawasan perumahan dan aktiviti industri di situ lebih ringan berbanding dua stesen yang lain. Stesen Klang kerap kali dilaporkan sebagai kawasan yang mempunyai tahap pencemaran yang lebih tinggi daripada kawasan stesen yang lain (JASM 2019a, 2019b). Semua nilai data yang hilang digantikan dengan nilai min data.

#### KAEDAH TAKENS

Disebabkan keadaan siri masa  $PM_{10}$  yang berada dalam bentuk 1-matra, penurasan ciri-ciri topologi yang wujud dalam matra tinggi seperti lubang dan lompong tidak dapat dilakukan. Pengubahan keadaan data kepada matra yang lebih tinggi dilakukan dengan menggunakan teknik Takens. Merujuk kepada kaedah Takens (Takens 1981), siri masa  $x_0, x_1, \dots, x_{n-1}$  dibina semula dalam ruang fasa dengan setiap titik diwakili vektor  $x_n(m, \tau) = x_n, x_{n+\tau}, \dots, x_{n+(m-1)\tau}$  dengan  $\tau$  sebagai masa tunda dan  $m$  sebagai matra penbenaman. Penulis menggunakan nilai  $\tau = 1$

dan  $m = 3$  untuk setiap set data dalam kajian ini. Hal ini disebabkan penggunaan nilai  $\tau$  dan  $m$  yang berbeza adalah tidak sesuai untuk membuat perbandingan antara tiga stesen. Atas sebab ini juga, Umeda (2017) telah menggunakan nilai  $\tau = 1$  dan  $m = 3$  untuk set datanya. Nilai  $\tau = 1$  merupakan nilai remeh dan  $m = 3$  digunakan untuk mendapatkan ciri-ciri topologi lubang dan lompong. Kajian-kajian yang menggunakan pendekatan kaedah Takens untuk penerapan homologi gigih boleh dirujuk dalam Khasawneh dan Munch (2014), Khasawneh et al. (2018), Maletić et al. (2016) dan Pereira dan Mello (2015). Selain itu, kaedah Takens juga digunakan untuk peramalan siri masa seperti yang dilakukan Adenan dan Noorani (2016) untuk siri masa aliran sungai dan Hamid dan Noorani (2017) untuk siri masa ozon.

#### HOMOLOGI GIGIH

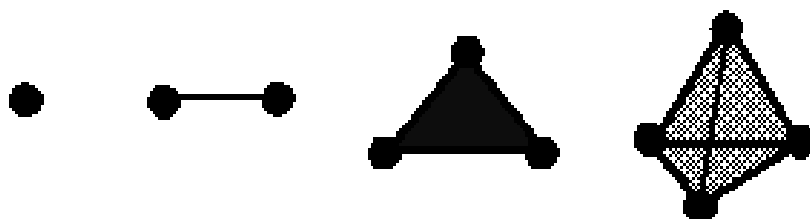
Homologi gigih diterapkan melalui pembinaan kompleks simpleks pada siri masa yang telah dibenamkan melalui kaedah Takens dan dikenali sebagai titik awan. Secara asasnya, titik awan ialah gugusan titik-titik yang berada dalam ruang Euklidian  $m$ -matra,  $\mathbb{R}^m$  (Ghrist 2008) dan dalam kajian ini titik awan adalah dalam matra 3 (iaitu  $m = 3$ ). Simpleks-simpleks adalah terdiri daripada titik atau bucu (0-simpleks), garisan atau tepi (1-simpleks), segi tiga (2-simpleks) dan tetrahedron (3-simpleks). Secara umumnya,  $k$ -simpleks,  $k = 0, 1, 2, \dots$ , ditakrifkan sebagai hul cembung bagi  $k+1$  bucu-bucunya (Hatcher 2002). Rajah 1 menunjukkan contoh simpleks-simpleks yang digunakan untuk membina kompleks simpleks. Nilai penurasan,  $\varepsilon$  digunakan untuk membina kompleks simpleks iaitu dengan melukis bulatan pada setiap titik dalam titik awan dengan  $\varepsilon$  sebagai jejari. Pembesaran bulatan berlaku selaras dengan peningkatan  $\varepsilon$  dan dua titik digabungkan dengan satu garisan jika bulatan tersebut bersilang.

Dalam kajian ini, kompleks Rips dibina pada titik awan. Berdasarkan kompleks Rips, satu  $k$ -simpleks,  $k = 0, 1, 2, \dots$ , dengan bucu  $\{q_1, q_2, \dots, q_n\}$  adalah sebahagian kompleks Rips jika jarak bersama antara sebarang pasangan bucu adalah kurang daripada  $\varepsilon$  (Gidea & Katz 2018). Gabungan simpleks pada titik awan berdasarkan nilai penurasan,  $\varepsilon$  membentuk kompleks simpleks. Pada setiap nilai penurasan, ciri-ciri topologi (komponen berkait, lubang, lompong) dikenal pasti dan maklumat tersebut direkodkan dalam kod bar dan rajah kegigihan. Kod bar adalah perwakilan bergraf ciri-ciri topologi yang mengandungi garisan mengufuk dengan setiap kiri dan kanannya mewakili titik lahir,  $b$  dan mati,  $d$  (Ghrist 2008). Rajah kegigihan merekodkan ciri-ciri topologi sebagai titik  $(b, d)$  dalam satah nyata terpeluas,  $\bar{\mathbb{R}}^2$  dengan  $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$  (Otter et al. 2017). Nilai kelahiran dan kematian ciri-ciri topologi masing-masing berada pada paksi- $x$  dan paksi- $y$  rajah kegigihan.

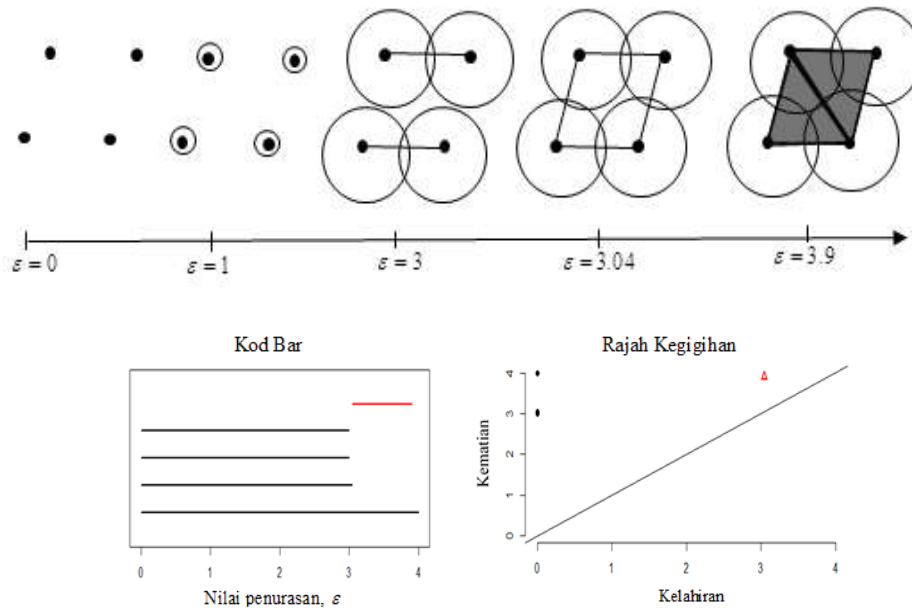
Rajah 2 menunjukkan proses pembinaan kompleks simpleks dan ciri-ciri topologi yang direkodkan sepanjang proses penurasan pada setiap nilai  $\varepsilon$ . Pada peringkat permulaan penurasan,  $\varepsilon = 0$  simpleks yang terhasil adalah empat 0-simpleks dan kelahiran ciri topologi yang dapat dilihat adalah empat komponen berkait. Empat bulatan mula kelihatan selaras dengan peningkatan nilai penurasan kepada  $\varepsilon = 1$ . Bulatan tersebut bersilang pada  $\varepsilon = 3$  dan dua 1-simpleks terhasil mematikan dua komponen berkait. Ciri topologi lubang mula lahir pada  $\varepsilon = 3.04$  dan mati pada  $\varepsilon = 3.9$  selaras dengan kemunculan 2-simpleks yang menutup lubang tersebut. Proses penurasan dihentikan pada nilai penurasan  $\varepsilon = 4$  apabila semua bucu telah dihubungkan dan membentuk satu komponen berkait yang besar. Semua maklumat topologi tersebut direkodkan dalam kod bar dan rajah kegigihan Rajah 2.

#### UKURAN JARAK

Perbandingan antara satu rajah kegigihan dengan satu lagi rajah kegigihan dihitung menggunakan



RAJAH 1. 0-simpleks, 1-simpleks, 2-simpleks dan 3-simpleks (kiri ke kanan)



RAJAH 2. Ilustrasi pembentukan ciri-ciri topologi daripada pembinaan kompleks simpleks. Garisan mengufuk dan titik hitam dalam rajah kegigihan mewakili ciri topologi komponen berkait dan garisan mengufuk dan segitiga merah mewakili ciri topologi lubang

ukur jarak Wassestein (Edelsbrunner & Harer 2010). Definisi ukur jarak diberi seperti berikut:

$$W_p(X, Y) = \inf_{\varphi: X \rightarrow Y} \sum_{x \in X} \left[ \|x - \varphi(x)\|_{\infty}^p \right]^{1/p} \quad (1)$$

dengan  $p$  adalah darjah Wasserstein yang ingin dihitung manakala  $X$  dan  $Y$  adalah dua rajah kegigihan. Infimum bagi pemetaan bijeksi  $\varphi: X \rightarrow Y$  diambil kira dahulu sebelum jarak antara titik-titik yang mewakili ciri-ciri topologi dalam dua rajah tersebut dihitung. Unsur  $x \in X$  dan  $y \in Y = (b_j, d_j)$  dan  $y_j = (b_j, d_j)$  mewakili ciri-ciri topologi dengan titik kelahiran,  $b$  dan titik kematian,  $d$ , dan jarak antara mereka dihitung sebagai  $\|x_i - y_j\|_{\infty} = \max\{|b_i - b_j|, |d_i - d_j|\}$ .

KEPUTUSAN DAN PERBINCANGAN

Siri masa purata bulanan  $PM_{10}$  dari Januari 2000 hingga Disember 2015 dibina semula menggunakan kaedah Takens dengan pemilihan masa tunda  $\tau = 1$  dan matra pembenam  $m = 3$  menjadikan data titik awan dalam matra tiga,  $\mathbb{R}^3$ . Tiga data titik awan terhasil daripada setiap stesen pengawasan kualiti udara (Klang, Petaling Jaya, Shah Alam). Homologi gigih diterapkan pada data titik awan tersebut, dimulakan dengan pembinaan kompleks Rips daripada nilai penurasan,  $\epsilon_{\min} = 0$  hingga  $\epsilon_{\max} = 100$ . Pemilihan nilai penurasan tersebut adalah untuk mengkaji perkembangan topologi daripada pembentukan kompleks simpleks paling mudah hingga menjadi satu bentuk kompleks simpleks yang telah bersambung semua

bucunya. Penerapan homologi gigih adalah melalui perisian pakej R-TDA (Fasy et al. 2017). Peringkat pertama, sifat kualitatif dilihat melalui rajah kegigihan yang merekodkan perkembangan ciri-ciri topologi, komponen berkait, lubang dan lompong. Kemudian, analisis kuantitatif dilakukan untuk menghitung persamaan topologi tersebut menggunakan darjah  $p = 1$  ukur jarak Wasserstein. Penggunaan  $p = 1$  digunakan kerana pengiraan jarak Wasserstein untuk darjah  $p = 2$  dan  $p = 3$  tidak mengubah hasil kajian ini. Keseluruhan proses bagi teknik kajian ini ditunjuk melalui carta aliran di Rajah 3.

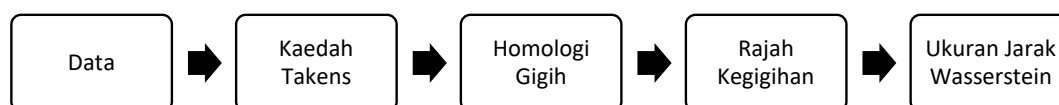
Rajah 4 menunjukkan rajah kegigihan yang mengandungi maklumat topologi untuk setiap stesen pengawasan kualiti udara yang dikaji dalam penyelidikan ini. Melalui pemerhatian, kebanyakan komponen berkait (titik hitam) untuk stesen Petaling Jaya dan Shah Alam mempunyai titik kematian pada nilai penurasan  $\epsilon$  kurang daripada atau sama dengan 20. Berbeza dengan stesen Klang, banyak komponen berkait mempunyai nilai kematian antara  $\epsilon$  lebih daripada 20 hingga kurang daripada atau sama dengan 40. Bagi ciri-ciri topologi yang lain, lubang (segi tiga merah) dan lompong (rombus biru) di stesen Petaling Jaya dan Shah Alam mempunyai nilai kelahiran dan kematian pada nilai penurasan  $\epsilon$  kurang daripada atau sama dengan 30 kecuali satu lubang di rajah kegigihan Petaling Jaya. Terdapat beberapa ciri topologi lubang dan satu lompong pada stesen Klang yang wujud pada nilai penurasan  $\epsilon$  lebih daripada 30 dan ini membezakan stesen Klang dengan dua stesen yang lain.

Berdasarkan kepada keseluruhan pemerhatian pada Rajah 4, stesen Klang mempunyai jangka hayat ciri-

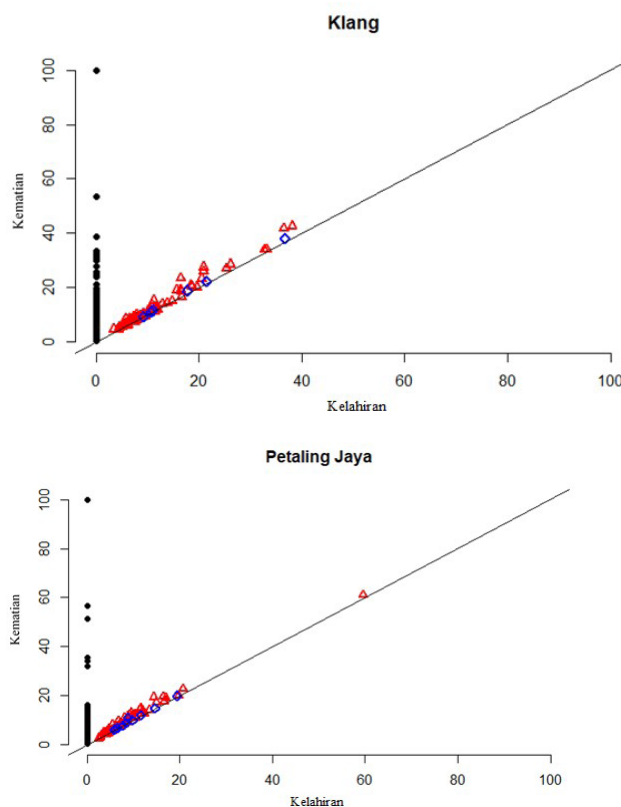
ciri topologi yang lebih lama berbanding dengan stesen yang lain. Hal ini disebabkan seringkali stesen Klang mempunyai bacaan  $PM_{10}$  yang lebih tinggi berbanding dengan stesen-stesen yang lain (JASM 2019a; Rahman et al. 2015). Perbezaan perilaku pencemar  $PM_{10}$  dicirikan melalui kewujudan ciri-ciri topologi berdasarkan nilai kematian dan nilai kelahiran ciri-ciri tersebut. Berdasarkan nilai-nilai tersebut stesen Petaling Jaya dan Shah Alam menunjukkan perilaku ciri topologi yang seakan serupa berbanding stesen Klang. Analisis kuantitatif diperlukan untuk menganalisis perbezaan ciri-ciri kualitatif ini dan ukur jarak Wasserstein digunakan untuk menganalisis keserupaan setiap ciri topologi tersebut.

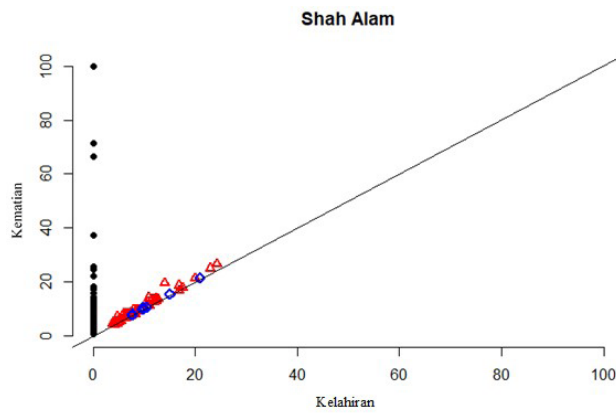
Jadual 1-3 menunjukkan hasil kiraan ukur jarak Wasserstein bagi ciri topologi komponen berkait, lubang dan lompong antara rajah kegigihan bagi tiga stesen iaitu Klang, Petaling Jaya dan Shah Alam. Jarak yang terendah menunjukkan dua stesen adalah lebih serupa antara stesen yang lain. Berdasarkan ukuran jarak untuk komponen berkait, lubang dan lompong, stesen Petaling Jaya dan Shah Alam adalah yang paling serupa berdasarkan jarak yang terendah manakala stesen Klang adalah yang paling tidak sama antara yang lain.

Penentusahan hasil ini dilakukan dengan menggunakan kaedah analisis kelompok aglomeratif berhierarki (AKAB) yang mengelompokkan stesen berikut mengikut jarak ketakserupaan yang dihitung menggunakan ukuran jarak Euklidan (Everitt et al. 2011). Rajah 5 menunjukkan hasil AKAB untuk tiga stesen tersebut. Stesen Petaling Jaya dan Shah Alam menunjukkan jarak ketakserupaan yang paling minimum dan dikelompokkan dalam satu kelompok manakala stesen Klang dikelompokkan dalam kelompok yang lain. Pengelompokan ini memberi kesimpulan yang stesen Klang adalah stesen yang paling tak serupa berbanding dua stesen yang lain. Hal ini demikian kerana, kedudukan stesen-stesen tersebut dengan Klang di kawasan perindustrian yang utama manakala Petaling Jaya dan Shah Alam yang terletak di kawasan berdekatan dengan aktiviti perindustrian tidak begitu berat berbanding Klang (Ling et al. 2014). Oleh kerana keputusan melalui pendekatan topologi memberi hasil yang sama dengan penentusahan menggunakan AKAB, kajian ini membuktikan keberkesanan pendekatan topologi dalam masalah kualiti udara dan seterusnya mampu dikembangkan lagi teknik kajian dalam isu pencemaran udara khususnya di Malaysia.

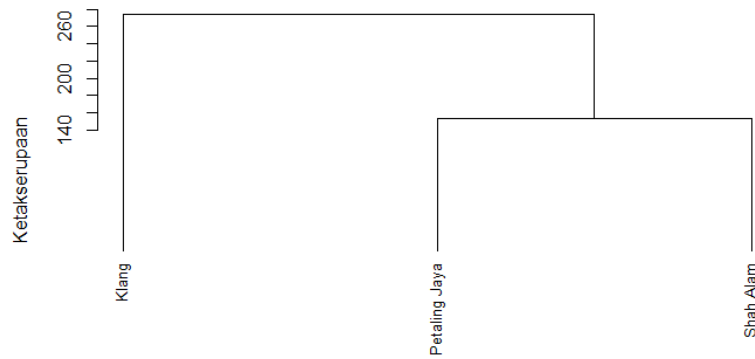


RAJAH 3. Carta aliran bagi keseluruhan kajian





RAJAH 4. Rajah kegigihan bagi setiap stesen pengawasan kualiti udara Klang, Petaling Jaya dan Shah Alam. Titik hitam, segi tiga merah dan rombus biru mewakili ciri-ciri topologi komponen berkait, lubang dan lompong



RAJAH 5. Dendrogram persamaan antara stesen pengawasan kualiti udara Klang, Petaling Jaya dan Shah Alam menggunakan kaedah AKAB

JADUAL 1. Ukur jarak Wasserstein untuk ciri topologi komponen berkait

	Klang	Petaling Jaya	Shah Alam
Klang	0		
Petaling Jaya	263.8945	0	
Shah Alam	248.9466	160.7275	0

JADUAL 2. Ukur jarak Wasserstein untuk ciri topologi lubang

	Klang	Petaling Jaya	Shah Alam
Klang	0		
Petaling Jaya	49.04755	0	
Shah Alam	45.2415	37.2712	0

JADUAL 3. Ukur jarak Wasserstein untuk ciri topologi lompong

	Klang	Petaling Jaya	Shah Alam
Klang	0		
Petaling Jaya	4.538522	0	
Shah Alam	4.064433	3.406084	0

## KESIMPULAN

Penerapan homologi gigih dalam menganalisis keserupaan pencemaran udara berdasarkan pencemar  $PM_{10}$  antara stesen pengawasan kualiti udara Klang, Petaling Jaya dan Shah Alam membuahkan hasil yang positif. Melalui pendekatan tersebut, stesen Klang mempunyai sifat topologi yang tak serupa berbanding dengan dua stesen yang lain. Persamaan ciri-ciri topologi yang lebih serupa diperhatikan antara stesen Petaling Jaya dan Shah Alam. Penentusahan keputusan tersebut melalui kaedah pengelompokan juga memberi hasil yang sama. Oleh itu, selain menggunakan pendekatan kuantitatif (analisis kelompok), pendekatan topologi juga dapat digunakan sebagai kaedah alternatif dan seterusnya membuka ruang yang luas bagi penggunaan jentera-jentera lain dalam bidang analisis data bertopologi dalam menganalisis kualiti udara di Malaysia.

## PENGHARGAAN

Penulis mengucapkan terima kasih kepada Jabatan Alam Sekitar Malaysia (JASM) dan Prof. Dr. Mohd Talib Latif daripada Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia untuk keizinan dan kerjasama dalam menggunakan data. Setinggi-tinggi penghargaan juga kepada Universiti Kebangsaan Malaysia untuk geran universiti penyelidikan DIP-2017-011 serta geran Kementerian Pendidikan Malaysia FRGS/1/2019/STG06/UKM/01/3.

## RUJUKAN

- Abdullah, A.M., Samah, M.A. & Jun, T.Y. 2012. An overview of the air pollution trend in Klang Valley, Malaysia. *Open Environmental Sciences* 6(1): 13-19.
- Adenan, N.H. & Noorani, M.S.M. 2016. Multiple time-scales nonlinear prediction of river flow using chaos approach. *Jurnal Teknologi* 78: 1-7.
- Afroz, R., Hassan, M.N. & Ibrahim, N.A. 2003. Review of air pollution and health impacts in Malaysia. *Environmental Research* 92(2): 71-77.
- Austin, E., Coull, B.A., Zanobetti, A. & Koutrakis, P. 2013. A framework to spatially cluster air pollution monitoring sites in US based on the  $PM_{2.5}$  composition. *Environment International* 59: 244-254.
- Azid, A., Juahir, H., Ezani, E., Toriman, M.E., Endut, A., Rahman, M.N.A., Yunus, K., Kamarudin, M.K.A., Hasnam, C.N.C., Saudi, A.S. & Umar, M.R. 2015. Identification source of variation on regional impact of air quality pattern using chemometric. *Aerosol and Air Quality Research* 15: 1545-1558.
- 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(1): 53-64.
- Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R. & Feigin, V. 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *The Lancet* 389(10082): 1907-1918.
- Edelsbrunner, H. & Harer, J. 2010. *Computational Topology: An Introduction*. Applied Math Textbook.
- Everitt, B.S., Landau, S., Leese, M. & Stahl, D. 2011. *Cluster Analysis*. 5th ed. Chichester, UK: John Wiley & Sons.
- Fasy, B.T., Kim, J., Lecci, F., Maria, C. & Rouvreau, V. 2017. Statistical tools for topological data analysis. <https://cran.r-project.org/web/packages/TDA/TDA.pdf>. Diakses pada 2 Februari 2018.
- Ghrist, R. 2008. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society* 45(1): 61-75.
- Gidea, M. & Katz, Y. 2018. Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications* 491: 820-834.
- Hajij, M., Wang, B., Scheidegger, C. & Rosen, P. 2018. Visual detection of structural changes in time-varying graphs using persistent homology. 2018 *IEEE Pacific Visualization Symposium (PacificVis)*. hlm. 125-134.
- Hamid, N.Z.A. & Noorani, M.S.M. 2017. Aplikasi model baharu penambahbaikan pendekatan kalut ke atas peramalan siri masa kepekatan ozon. *Sains Malaysiana* 46(8): 1333-1339.
- Hatcher, A. 2002. *Algebraic Topology*. Cambridge: Cambridge University Press.
- Hua, A.K. 2018. Applied chemometric approach in identification sources of air quality pattern in Selangor, Malaysia. *Sains Malaysiana* 47(3): 471-479.
- Iizuka, A., Shirato, S., Mizukoshi, A., Noguchi, M., Yamasaki, A. & Yanagisawa, Y. 2014. A cluster analysis of constant ambient air monitoring data from the Kanto Region of Japan. *International Journal of Environmental Research and Public Health* 11(7): 6844-6855.
- Isiyaka, H.A. & Azid, A. 2015. Air quality pattern assessment in Malaysia using multivariate techniques. *Malaysian Journal of Analytical Sciences* 19(5): 966-978.
- Jamalani, M.A., Abdullah, A.M., Azid, A., Ramli, M.F., Baharudin, M.R., Bose, M. MELhadi, R.E., Youssef,

- K.A.A.B., Gnadimzadeh, A. & Gumel, D.Y. 2016. Monthly analysis of  $PM_{10}$  in ambient air of Klang Valley, Malaysia. *Malaysian Journal of Analytical Sciences* 20(5): 1159-1170.
- JASM. 2019a. *Laporan Kualiti Alam Sekeliling* 2015. <https://enviro.doe.gov.my/>. Diakses pada 24 Februari 2019.
- JASM. 2019b. *Laporan Kualiti Alam Sekeliling* 2013. <https://enviro.doe.gov.my/>. Diakses pada 24 Februari 2019.
- Kerber, M., Morozov, D. & Nigmatov, A. 2017. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics* 22: 1-4.
- Khasawneh, F.A. & Munch, E. 2014. Stability determination in turning using persistent homology and time series analysis. *Prosiding ASME 2014 International Mechanical Engineering Congress and Exposition*, November 14-20, Montreal, Canada.
- Khasawneh, F.A., Munch, E. & Perea, J.A. 2018. Chatter classification in turning using machine learning and topological data analysis. *IFAC-PapersOnLine* 51(14): 195-200.
- Kramár, M., Levanger, R., Tithof, J., Suri, B., Xu, M., Paul, M., Schatz, M.F. & Mischaikow, K. 2016. Analysis of Kolmogorov flow and Rayleigh-Bénard convection using persistent homology. *Physica D: Nonlinear Phenomena* 334: 82-98.
- Ling, O.H.L., Siti Nur, A.M. & Noralizawati, M. 2014. Air quality and land use in urban region of Petaling Jaya, Shah Alam and Klang, Malaysia. *Environment Asia* 7(1): 134-144.
- Lu, W.Z., He, H.D. & Dong, L.Y. 2011. Performance assessment of air quality monitoring networks using principal component analysis and cluster analysis. *Building and Environment* 46(3): 577-583.
- Maletić, S., Zhao, Y. & Rajković, M. 2016. Persistent topological features of dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 26(5): 053105.
- Nicolau, M., Levine, A.J. & Carlsson, G. 2011. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences* 108(17): 7265-7270.
- Otter, N., Porter, M.A., Tillmann, U., Grindrod, P. & Harrington, H.A. 2017. A roadmap for the computation of persistent homology. *EPJ Data Science* 6(1): 17.
- Pereira, C.M. & Mello, R.F.D. 2015. Persistent homology for time series and spatial data clustering. *Expert Systems with Applications* 42(15-16): 6026-6038.
- Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P.J. & Vaccarino, F. 2014. Homological scaffolds of brain functional networks. *Journal of the Royal Society Interface* 11(101): 20140873.
- Rahman, S.R.A., Ismail, S.N.S., Raml, M.F., Latif, M.T., Abidin, E.Z. & Praveena, S.M. 2015. The assessment of ambient air pollution trend in Klang Valley, Malaysia. *World Environment* 5(1): 1-11.
- Seversky, L.M., Davis, S. & Berger, M. 2016. On time-series topological data analysis: New data and opportunities. *Prosiding IEEE Conference on Computer Vision and Pattern Recognition Workshops*. hlm. 59-67.
- Spangl, W., Schneider, J., Moosmann, L. & Nagl, C. 2007. Representativeness and classification of air quality monitoring stations. *Final Report: Umweltbundesamt*.
- Takens, F. 1981. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics Dynamical Systems and Turbulence, Warwick* 1980: 366-381.
- Umeda, Y. 2017. Time series classification via topological data analysis. *Information and Media Technologies* 12: 228-239.
- Zainordin, N.S., Ramli, N.A. & Elbayoumi, M. 2017. Distribution and temporal behaviour of O<sub>3</sub> and NO<sub>2</sub> near selected schools in Seberang Perai, Pulau Pinang and Parit Buntar, Perak, Malaysia. *Sains Malaysiana* 46(2): 197-207.
- Zulkepli, N.F.S., Noorani, M.S.M., Razak, F.A., Ismail, M. & Alias, M.A. 2019a. Topological characterization of haze episodes using persistent homology. *Aerosol and Air Quality Research* 19: 1614-1624.
- Zulkepli, N.F.S., Noorani, M.S.M., Razak, F.A., Ismail, M. & Alias, M.A. 2019b. Haze detection using persistent homology. *AIP Conference Proceedings* 2111: 020012.

Fakulti Sains dan Teknologi  
Universiti Kebangsaan Malaysia  
43600 UKM Bangi, Selangor Darul Ehsan  
Malaysia

\*Pengarang untuk surat-menyurat; email: farihasyaqina@yahoo.com

Diserahkan: 28 Ogos 2019  
Diterima: 14 Disember 2019