

Pemodelan Taburan Kebarangkalian Zarah Terampai Melampau di Lembah Klang (Modelling of Probability Distributions of Extreme Particulate Matter in Klang Valley)

MUHAMMAD ASLAM MOHD SAFARI* & WAN ZAWIAH WAN ZIN

ABSTRAK

Kajian ini bertujuan untuk mengenal pasti model statistik terbaik bagi mewakili set data melampau untuk salah satu bahan pencemaran udara iaitu zarah terampai (PM_{10}). Data dari enam buah stesen pengawasan kualiti udara di sekitar Lembah Klang dari tahun 2009 hingga 2011 digunakan dalam kajian ini. Dalam penentuan taburan terbaik, taburan parametrik dan taburan tak berparameter telah diuji. Dua siri data esktrim yang digunakan ialah siri data maksimum bulanan dan siri data melangkaui ambang bagi PM_{10} . Seterusnya, dua taburan parametrik iaitu Taburan Melampau Teritlak (GEV) dan Taburan Pareto Teritlak (GPD) masing-masing dipadankan kepada siri data maksimum bulanan dan siri data melangkaui ambang. Kaedah penganggaran parameter L-momen dan ujian kebagusan penyuaian Anderson Darling digunakan dalam pemilihan taburan parametrik terbaik yang juga menentukan kaedah pemilihan data melampau yang mana lebih baik. Bagi kaedah tak berparameter, penganggaran fungsi ketumpatan kernel (KDE) digunakan untuk menentukan taburan terbaik PM_{10} melampau. Hasil pengiraan ralat min kuasa dua (MSE) mendapati taburan tak berparameter merupakan taburan terbaik bagi data melampau PM_{10} di kebanyakan stesen kajian. Taburan terbaik bagi setiap stesen kajian seterusnya digunakan bagi menghitung tempoh ulangan PM_{10} yang sangat berguna bagi pihak yang terbabit.

Kata kunci: Fungsi ketumpatan kernel; L-momen; PM_{10} ; taburan Nilai Melampau Teritlak; taburan pareto teritlak; taburan tak berparameter; ujian penyuaian Anderson Darling

ABSTRACT

This study aims to identify the best statistical model to represent the data set for one of the air pollutants that is the particulate matter with diameters smaller than 10 micrometers (PM_{10}). Data from six air quality monitoring stations in the Klang Valley from 2009 to 2011 were used in this study. In determining the more appropriate probability distribution, both parametric and non-parametric approaches were tested. Two series of extreme data for PM_{10} were used, which are the monthly maximum and the Peak over threshold data series. Next, two parametric distributions, which are the Generalized Extreme Value (GEV) and Generalized Pareto (GPD) were fitted to the monthly maximum and the Peak over threshold data series, respectively. L-moment parameter estimation method and Anderson Darling goodness of fit test were used to identify the best parametric distribution as well as the more suitable data series to represent extreme data. For the non-parametric approach, the kernel density estimation (KDE) is used in this study to determine the best distribution for extreme PM_{10} . Based on the mean squared error (MSE) results, it is found that the nonparametric distribution is the best distribution for extreme PM_{10} data from most of the air quality monitoring stations. The best distribution for each air quality monitoring station is then used to estimate several return periods for extreme PM_{10} which are very useful for relevant authorities.

Keywords: Anderson Darling goodness of fit test; generalized extreme value; generalized pareto; kernel density estimation; L-moments; non-parametric distribution; PM_{10}

PENGENALAN

Pencemaran udara merupakan salah satu jenis pencemaran utama yang berlaku di Malaysia. Menurut Pertubuhan Kesihatan Dunia (WHO 2015), pencemaran udara ditakrifkan sebagai satu keadaan yang melibatkan kehadiran sebarang bahan kimia, jirim zarah atau bahan biologi di ruang udara yang boleh memberi mudarat, ketidakselesaan kepada manusia atau organisma hidup lain serta merosakkan alam sekitar. Afroz et al. (2003) menyatakan bahawa kenderaan bermotor merupakan punca utama di Malaysia yang menyumbang kepada peningkatan

bahan pencemar di udara terutama zarah terampai (PM_{10}). Pertambahan kenderaan bermotor telah menyebabkan peningkatan aras kepekatan bahan pencemar di udara berdasarkan kajian Awang et al. (2000).

Zarah terampai (PM_{10}) yang mempunyai saiz diameter kurang atau sama dengan 10 (μm) merupakan salah satu bahan pencemar utama di Malaysia (Dominick et al. 2012). Mohd Yusoff et al. (1989) mentakrifkan PM_{10} sebagai pepejal yang terapung di udara yang boleh mengganggu kehidupan manusia, tumbuhan, binatang dan alam sekitar. PM_{10} terhasil daripada debu dan asap yang berpunca

daripada kerja pembinaan, asap kenderaan serta aktiviti perindustrian. Selain itu, menurut Brook et al. (2004), PM_{10} ialah bahan bawaan udara yang mengandungi bahan pepejal campuran yang heterogen dan zarah cecair yang terperangkap di udara serta mempunyai saiz dan tindak balas kimia yang berbeza mengikut tempat dan masa. Kepekatan PM_{10} yang tinggi boleh memberi kemudaratan terhadap kesihatan manusia. Kostova et al. (2012) mendapati bahawa terdapat hubungan yang bererti antara PM_{10} dengan kemasukan pesakit ke hospital akibat penyakit respiratori. Manakala Santus et al. (2012) membuktikan, terdapat hubungan yang bererti antara penyakit paru-paru obstruktif kronik (COPD) dengan paras kepekatan PM_{10} . PM_{10} juga merupakan bahan pencemar yang boleh meningkatkan kadar risiko kematian akibat penyakit respiratori sebanyak 3.63% (Wan Mahiyudin et al. 2013). Selain daripada itu, Sahani et al. (2014) mendapati kejadian jerebu ($PM_{10} > 100 \mu\text{g}/\text{m}^3$) boleh meningkatkan kadar risiko kematian sebanyak 19% bagi pesakit yang menghidap penyakit respiratori.

Terdapat beberapa kajian berkenaan pemodelan taburan melampau menggunakan data pencemaran udara dilakukan di dalam dan luar negara. Di Malaysia, Hurairah et al. (2012) dan Masseran et al. (2016) telah melakukan pemodelan taburan melampau bagi data Indeks Pencemaran Udara (IPU) dan data karbon monoksida (CO). Pemodelan taburan melampau data pencemaran udara juga dilakukan di luar negara seperti di Munich, United Kingdom dan Barcelona (Kuchenhoff & Thamerus 1995; Quantela-del-Rio & Francisco-Fernandez 2011; Tobias & Scotto 2003; Zhou et al. 2012). Kaedah yang berbeza telah digunakan oleh para pengkaji ini dalam usaha memodelkan taburan melampau data pencemaran udara. Lazimnya, pemodelan taburan melampau dilakukan menggunakan model parametrik dan seterusnya taburan terbaik yang berjaya dikenal pasti akan digunakan bagi menghitung tempoh ulangan atau nilai ulangan sesuatu data yang dikaji. Walau bagaimanapun, Quantela-del-Rio dan Francisco-Fernandez (2011) dalam kajian mereka telah berjaya membuktikan bahawa pemodelan taburan melampau data Ozon (O_3) dengan menggunakan taburan tak berparameter memberikan penyuaian yang lebih baik berbanding taburan Nilai Melampau Teritlak (GEV). Oleh itu, kajian seperti ini adalah penting kerana model yang terbaik berupaya menerangkan ciri-ciri data pencemaran udara dengan lebih tepat dan seterusnya membolehkan ramalan terhadap data pencemaran udara dilakukan dengan lebih baik.

Kajian ini bertujuan bagi menentukan taburan yang lebih baik dalam memodelkan siri data melampau PM_{10} di Lembah Klang berdasarkan kaedah taburan parametrik dan taburan tak berparameter. Langkah pertama dalam mencapai tujuan ini ialah penentuan set data yang dikatakan sebagai data melampau. Seterusnya, siri data melampau yang dijana akan dipadankan dengan taburan yang bersesuaian dan perbandingan antara taburan parametrik dengan taburan tak berparameter akan dilakukan. Setelah model terbaik untuk setiap stesen kajian

berjaya diperoleh, tempoh ulangan akan dihitung bagi mengenal pasti anggaran selang masa kebolehdajian bagi sesuatu kepekatan PM_{10} untuk berlaku atau berulang.

DATA KAJIAN DAN SKOP KAJIAN

Data PM_{10} di sekitar Lembah Klang yang direkodkan dalam unit jam diperoleh dari Jabatan Alam Sekitar (JAS) digunakan dalam kajian ini bermula dari tahun 2009 hingga 2011. Melalui kaedah blok maksima dan puncak melangkaui ambang, data maksimum bulanan dan data melangkaui ambang bagi zarah terampai dikenal pasti dan digunakan bagi pepadanan taburan melampau. Blok maksimum bulanan digunakan bagi PM_{10} berpandukan garis panduan kepekatan PM_{10} yang direkodkan mengikut purata 24 jam.

Data PM_{10} yang direkodkan setiap jam dari enam stesen pengawasan kualiti udara di sekitar Lembah Klang (Jadual 1) digunakan dalam kajian ini. Stesen-stesen dari Lembah Klang dipilih dalam kajian ini memandangkan kawasan ini merupakan kawasan yang amat pesat membangun di Malaysia dengan kepadatan penduduk yang tinggi. Pelepasan asap dari kenderaan bermotor merupakan faktor utama pencemaran udara di Lembah Klang (Abdullah et al. 2012). Menurut Keyword et al. (2003), selain daripada habuk dan zarah halus, asap daripada kenderaan berenjin diesel dan asap aerosol hasil pembakaran hutan di Semenanjung Malaysia juga merupakan antara penyumbang kepada berlakunya jerebu di Lembah Klang.

KAEDAH KAJIAN

Secara umumnya, terdapat dua kaedah pemodelan yang digunakan dalam kajian ini. Kaedah yang pertama merupakan pemodelan taburan melampau menggunakan taburan parametrik. Kaedah yang kedua adalah pemodelan taburan melampau menggunakan taburan tak berparameter. Kemudian, perbandingan antara taburan parametrik dan taburan tak berparameter dilakukan bagi memperoleh taburan melampau terbaik di setiap stesen pengawasan kualiti udara di Lembah Klang. Akhir sekali, tempoh ulangan bagi setiap stesen akan dihitung berdasarkan taburan terbaik yang diperoleh.

PEMODELAN TABURAN PARAMETRIK

BLOK MAKSIMA

Kaedah blok maksima merupakan pemilihan nilai atau data melampau di dalam selang tidak bertindih dan mempunyai saiz yang sama. Katakan X_1, X_2, \dots merupakan satu sampel rawak tak bersandar dengan fungsi taburan kebarangkalian F . Blok maksima bagi $m=1, 2, \dots$, dan $j=1, 2, \dots, k$ boleh ditulis sebagai:

$$M_j = \max_{(j-1)m+1 \leq i \leq jm} X_i.$$

JADUAL 1. Enam stesen pengawasan kualiti udara di sekitar Lembah Klang

Lokasi ID	Lokasi Stesen	Longitud	Latitud	Kategori
CA011	Sek. Men. Perempuan Raja Zarina, Klang	N 03 00.597	E 101 24.507	Bandar
CA016	Sek. Ren. Sri Petaling, Petaling Jaya	N 03 06.553	E 101 38.322	Perindustrian
CA025	Sek. Keb. TTDI Jaya, Shah Alam, Selangor	N 03 06.281	E 101 33.375	Bandar
CA053	Sek. Keb. Presint 8(2)	N 02 55.915	E 101 40.910	Bandar
CA054	S.M. Keb. Seri Permaisuri, Cheras	N 03 6.374	E 101 43.068	Bandar
CA058	Sek. Keb. Batu Muda, Batu Muda	N 03 12.748	E 101 40.929	Bandar

Oleh itu, $m \times k$ cerapan dibahagikan kepada k blok bersaiz m . Manakala $n = m \times k$ merupakan jumlah cerapan yang diperolehi iaitu data-data melampau yang telah dicerap. Saiz selang yang digunakan dalam kajian ini ialah selang bulanan ($m=30$ hari) berpandukan garis panduan kepekatan PM_{10} yang direkodkan mengikut purata 24 jam. Seterusnya data PM_{10} maksimum bulanan akan dipadankan terhadap taburan Nilai Melampau Teritlak (GEV) kerana Coles (2001) membuktikan taburan GEV merupakan taburan yang sesuai bagi pemodelan data blok maksima. Fungsi ketumpatan kebarangkalian dan fungsi taburan kumulatif bagi taburan GEV boleh ditulis sebagai:

$$f(x) = \frac{1}{\alpha} \left(1 - \frac{\kappa}{\alpha}(x - \varepsilon)\right)^{\frac{1}{\kappa} - 1} \exp\left(-\left[1 - \frac{\kappa}{\alpha}(x - \varepsilon)\right]^{\frac{1}{\kappa}}\right) \quad (1)$$

$$F(x) = \exp\left(-\left[1 - \frac{\kappa}{\alpha}(x - \varepsilon)\right]^{\frac{1}{\kappa}}\right). \quad (2)$$

Fungsi kuantil bagi taburan GEV boleh ditulis sebagai:

$$Q(f) = \varepsilon + \frac{\alpha}{\kappa} \left[1 - (-\ln(F))^{\kappa}\right], \quad (3)$$

dengan ε merupakan parameter lokasi; α adalah parameter skala; dan κ sebagai parameter bentuk.

PUNCAK MELANGKAUI AMBANG

Model puncak melangkaui ambang merupakan salah satu kaedah yang digunakan dalam teori nilai melampau. Data yang melebihi satu ambang (u) yang cukup tinggi, akan dipadankan dengan taburan Pareto Teritlak (GPD) (Coles 2001).

Katakan X_1, X_2, \dots, X_n merupakan satu cerapan rawak yang bertabur secara secaman dan tak bersandar, dengan fungsi taburan F . Teori bagi kaedah puncak melangkaui ambang mengandaikan, bersyaratkan suatu ambang u , taburan kumulatif untuk nilai keberlangkauan x melebihi u ditakrifkan sebagai:

$$F_u(x) = P(X \leq x | X > u) = \frac{F(x) - F(u)}{1 - F(u)}, x \geq u. \quad (4)$$

Menurut Coles (2001), jika ambang u , adalah cukup tinggi, taburan GPD merupakan taburan asimptot bagi $F_u(x)$. Fungsi ketumpatan kebarangkalian, fungsi taburan kumulatif dan fungsi kuantil bagi taburan GPD boleh ditulis:

$$f(x) = \frac{1}{\alpha} \left(1 - \frac{\kappa}{\alpha}(x - \varepsilon)\right)^{\frac{1}{\kappa} - 1}. \quad (5)$$

$$F(x) = 1 - \left(1 - \frac{\kappa}{\alpha}(x - \varepsilon)\right)^{\frac{1}{\kappa}}. \quad (6)$$

$$Q(f) = \varepsilon + \frac{\alpha}{\kappa} \left[1 - (1 - F)^{\kappa}\right], \quad (7)$$

dengan ε merupakan parameter lokasi; α adalah parameter skala; dan κ sebagai parameter bentuk.

Pemilihan ambang u , dilakukan dengan menggunakan dua ujian yang berbeza iaitu min frekuensi kejadian melangkaui ambang dan min keberlangkauan melebihi ambang. Melalui ujian min frekuensi kejadian melangkaui ambang, Begueria (2005) berpendapat, nilai min frekuensi yang patut dipilih adalah antara selang [1.2, 5]. Melalui ujian min keberlangkauan melebihi ambang pula, Davison (1984) dan Lang et al. (1999) mencadangkan agar nilai ambang yang dipilih berada dalam dominan yang min keberlangkauan adalah fungsi linear terhadap ambang.

PENGANGGARAN PARAMETER

Penganggaran bagi parameter taburan yang akan digunakan dalam kajian ini ialah penganggaran menggunakan kaedah L-momen. Kaedah ini menggunakan maklumat berdasarkan gabungan linear statistik tertib yang telah dikemukakan oleh Hosking (1990). Menurut Hosking dan Wallis (1997), kaedah L-momen lebih teguh dan mudah pengiraannya dalam menganggar parameter untuk sebarang taburan berbanding kaedah kebolehdajadian maksimum. L-momen merupakan satu kaedah penganggaran parameter yang sesuai bagi menganggar parameter taburan melampau untuk sampel data yang kecil (Vogel & Fennessey 1993). Oleh itu, kaedah ini sesuai digunakan dalam kajian ini disebabkan sampel data PM_{10} maksimum bulanan dan data PM_{10} melangkaui ambang yang tidak begitu besar. L-momen ke- r , ditulis sebagai , bagi populasi ditakrifkan sebagai:

$$\lambda_r = r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-j}{j} E[X_{r-j:r}], r = 1, 2, \dots \tag{8}$$

merupakan pemboleh ubah rawak bagi statistik tertib ke- $(r-j)$ daripada r cerapan.

UJIAN KEBAGUSAN PENYUAIAN

Ujian kebagusan penyuaian Anderson Darling (AD) yang telah dikemukakan oleh Anderson and Darling (1954) digunakan bagi memilih taburan parametrik terbaik dalam kajian ini. Menurut Stephens (1976), ujian kebagusan penyuaian Anderson Darling memberikan penyuaian yang lebih baik berbanding ujian Pearson khi kuasa dua bagi sampel data yang kecil. Ujian statistik AD boleh ditulis sebagai:

$$AD = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln F(y_i) - \ln(1 - F(y_{n+1-i}))],$$

$$i = 1, 2, \dots, n. \tag{9}$$

$F(\cdot)$ merupakan fungsi taburan kumulatif (CDF) bagi taburan tertentu, y_i merupakan cerapan ke- i dan n adalah bilangan cerapan.

PEMODELAN MENGGUNAKAN TABURAN TAK BERPARAMETER

Penganggaran taburan menggunakan kaedah penganggaran tak berparameter tidak memerlukan andaian bahawa data mengikut sebarang taburan tertentu. Ini menjadikan kaedah ini lebih fleksibel dalam menentukan taburan yang sesuai bagi data tertentu. Dalam kajian ini, kaedah penganggaran fungsi ketumpatan kernel merupakan kaedah tak berparameter yang digunakan bagi PM_{10} maksimum bulanan dan PM_{10} melangkaui ambang.

Katakan X merupakan pemboleh ubah rawak selanjur dengan fungsi ketumpatan f dan fungsi taburan F . Diberi satu sampel rawak $X_1, X_2, X_3, \dots, X_n$ dengan setiap X_i mempunyai taburan yang sama dengan X , Parzen (1962) menyatakan bahawa penganggaran kernel bukan parametrik bagi $f(\cdot)$ boleh ditulis sebagai:

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right). \tag{10}$$

Fungsi taburan boleh ditulis sebagai:

$$F_h(x) = \int_{-\infty}^x f_h(t) dt = \frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h}\right), \tag{11}$$

dengan

$$H(x) = \int_{-\infty}^x K(t) dt.$$

K merupakan fungsi kernel dan h adalah parameter pelicin atau lebar jalur. Pemilihan K dalam kaedah ini bukanlah satu perkara yang penting kerana mana-mana fungsi K yang memberikan hasil yang baik boleh digunakan dan tiada cara khusus bagi memilih fungsi K . Sebaliknya, kaedah pemilihan dan penggunaan lebar jalur, h yang tepat adalah lebih penting kerana lebar jalur akan mempengaruhi bentuk taburan yang ingin dikenal pasti. Menurut Silverman (1986), jika nilai h terlalu besar, penganggar yang dihasilkan adalah kurang licin manakala jika nilai h terlalu kecil, penganggar yang terhasil adalah terlalu licin. Kesannya, fungsi ketumpatan yang ingin dianggar gagal diperoleh atau anggaran fungsi ketumpatan yang diperoleh adalah kurang tepat. Fungsi kernel yang digunakan dalam kajian ini adalah fungsi kernel Gaussian.

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) I_{(-\infty, +\infty)}. \tag{12}$$

Antara kaedah yang biasa digunakan dalam pemilihan lebar jalur ialah kaedah silang pengesahan dan *plug in*. Pemilihan lebar jalur haruslah berdasarkan kamiran ralat min kuasa dua, *MISE*.

$$MISE(\hat{F}_h) = \int_{-\infty}^{+\infty} (\hat{F}_h(x) - F(x))^2 dx. \tag{13}$$

MISE merupakan satu ralat kuasa dua antara fungsi sebenar dengan fungsi anggaran. Lebar jalur yang dipilih mestilah lebar jalur yang meminimumkan penghampiran asimptot ralat ini. Lebar jalur tersebut boleh ditulis sebagai:

$$h_{AMISE} = Cn^{-1/3} = \left(\frac{0.5 \int_{-\infty}^{+\infty} V_F^2(x) dx}{\int_{-\infty}^{+\infty} B_F^2(x) dx} \right)^{1/3} n^{-1/3}, \tag{14}$$

dengan

$$B_F(x) = \left(\frac{1}{2}\right) (f'(x))^2 \left(\int_{-\infty}^{+\infty} x^2 K(x) dx \right).$$

$$V_F^2 = 2f(x) \left(\int_{-\infty}^{+\infty} xK(x) H(x) dx \right).$$

Pemalar C dalam rumus () bergantung kepada fungsi kernel dan fungsi taburan teori bagi data. Melalui kaedah *plug in*, penganggaran rumus () boleh ditulis sebagai:

$$\hat{h} = \hat{C}n^{-1/3}. \tag{15}$$

Bagi mencari lebar jalur yang sesuai, kaedah lalaran yang telah diperkenalkan oleh Polansky dan Baker (2000) akan digunakan. Satu kajian simulasi yang telah dilakukan oleh Bowman et al. (1998) menunjukkan bahawa kaedah

silang pengesahan memberikan lebar jalur yang lebih baik berbanding kaedah *plug in* yang telah diperkenalkan oleh Altman & Leger (1995). Walau bagaimanapun kaedah *plug in* dipilih kerana terdapat kelemahan dalam penggunaan kaedah silang pengesahan. Kelemahan kaedah silang pengesahan ialah kaedah ini mengambil masa yang lama dalam menghitung lebar jalur dengan menggunakan perisian tertentu. Selain itu, Quanteladel-Rio dan Francisco-Fernandez (2011) mendapati kaedah *plug in* menghasilkan lebar jalur yang lebih baik berdasarkan kajian simulasi yang mereka lakukan di samping perhitungan yang lebih cepat berbanding kaedah silang pengesahan. Quanteladel-Rio dan Estevez-Perez (2012) telah menunjukkan cara mudah bagi melakukan pemilihan lebar jalur dengan menggunakan perisian *R*.

PERBANDINGAN MODEL DAN PENGHITUNGAN TEMPOH ULANGAN

Perbandingan antara taburan parametrik dan tak berparameter dilakukan bagi memilih model terbaik yang mewakili taburan setiap stesen pengawasan kualiti udara dengan menghitung min ralat kuasa dua, *MSE* bagi kedua-dua taburan. Rumus *MSE* boleh ditulis sebagai:

$$MSE = \frac{1}{n} \sum_{k=1}^n \left(ND(c_k) - E[ND(c_k)] \right)^2, \quad (16)$$

dengan $ND(c_k)$ merupakan bilangan hari data PM_{10} maksimum bulanan dan PM_{10} melangkaui ambang melebihi ambang c_k . $E(ND(c_k))$ ialah jangkaan bilangan hari bagi PM_{10} maksimum bulanan dan PM_{10} melangkaui ambang melebihi ambang c_k berdasarkan taburan parametrik dan tak berparameter. Nilai c_k yang dipilih bagi data PM_{10} maksimum bulanan ialah 50, 75, 100, 120, 150 dan 180. Manakala c_k bagi data PM_{10} melangkaui ambang ialah 75, 90, 100, 120, 150 dan 180.

Setelah taburan terbaik berjaya diperolehi, tempoh ulangan, $T(x)$ akan dihitung bagi mengenal pasti tempoh yang akan diambil bagi sesuatu kejadian untuk berlaku semula dalam selang masa tertentu. Rumus tempoh ulangan, $T(x)$ boleh ditulis sebagai:

$$T(x) = \frac{1}{P(X > x)} = \frac{1}{1 - F(x)}, \quad (17)$$

dengan mengambil kira $F(x)$ sebagai fungsi taburan terbaik bagi setiap stesen pengawasan kualiti udara. Manakala nilai x yang akan digunakan ialah nilai ambang berbahaya bagi PM_{10} mengikut garis panduan yang telah ditetapkan oleh Jabatan Alam Sekitar (JAS). Oleh yang demikian, anggaran masa yang akan diambil untuk keadaan melebihi ambang berbahaya berlaku dapat diketahui dan seterusnya langkah-langkah untuk mengurangkan risiko berkaitan pencemaran dapat diambil.

HASIL KAJIAN

PEMODELAN TABURAN PARAMETRIK

Stesen pengawasan kualiti udara Shah Alam dipilih sebagai contoh untuk menunjukkan cara pemilihan ambang u , dilakukan. Ujian yang pertama adalah berdasarkan min frekuensi kejadian yang melangkaui ambang. Ujian ini memerlukan min frekuensi berada dalam domain [1.2, 5] bagi pemilihan ambang yang sesuai. Jadual 2 menunjukkan nilai-nilai frekuensi pada setiap ambang.

Seterusnya, ujian kedua adalah berdasarkan min keberlangkauan melebihi ambang mensyaratkan nilai ambang dipilih daripada domain yang berfungsi linear. Min keberlangkauan pada setiap ambang dikira dengan mengambil purata PM_{10} melangkaui ambang yang ditolak dengan nilai ambang. Nilai-nilai min keberlangkauan pada setiap ambang ditunjukkan dalam Jadual 3.

Rajah 1 menunjukkan plot min frekuensi dan min keberlangkauan bagi setiap calon ambang untuk stesen pengawasan kualiti udara Shah Alam. Berdasarkan ujian pertama, ambang pada persentil 95% dan 96% (82 dan 85) saja dipilih kerana nilai min frekuensi bagi kedua-dua ambang ini berada dalam selang [1.2, 5]. Melalui ujian kedua, dapat dilihat terdapat fungsi linear min keberlangkauan di beberapa bahagian dalam Rajah 1. Namun berdasarkan kedua-dua ujian ini, nilai ambang 85.00 ug/m^3 pada persentil 96% dipilih kerana memenuhi syarat kedua-dua ujian ini. Langkah-langkah pemilihan ambang yang sama dilakukan bagi data PM_{10} di stesen pengawasan kualiti udara yang berbaki dan hasil pemilihan ambang bagi kesemua stesen di tunjukkan dalam Jadual 4.

Jadual 5 menunjukkan hasil anggaran parameter taburan GEV dan GPD dengan menggunakan kaedah L-momen bagi data PM_{10} maksimum bulanan dan PM_{10} melangkaui ambang. Secara umumnya, parameter lokasi (ϵ) menunjukkan kedudukan atau purata sesuatu taburan. Nilai ϵ yang positif menunjukkan kedudukan taburan itu beralih ke kanan manakala nilai ϵ yang negatif menunjukkan kedudukan taburan beralih ke kiri. Parameter skala (α) pula menunjukkan saiz taburan sama ada sesuatu taburan itu lebih terserak atau tidak. Manakala parameter bentuk (κ) menunjukkan bentuk atau kepencongan sesuatu taburan sama ada pencong ke kiri atau ke kanan. Nilai κ yang besar akan menjadikan sesuatu taburan pencong ke kanan.

Seterusnya, ujian kebagusan penyuaian Anderson Darling (AD) digunakan bagi memilih taburan parametrik terbaik. Taburan terbaik akan dipilih dengan mengenal pasti nilai-p ujian Anderson Darling yang lebih besar daripada aras keertian 0.05 serta nilai-p terbesar bagi setiap taburan. Jadual 6 menunjukkan nilai-p ujian AD bagi taburan GEV dan GPD.

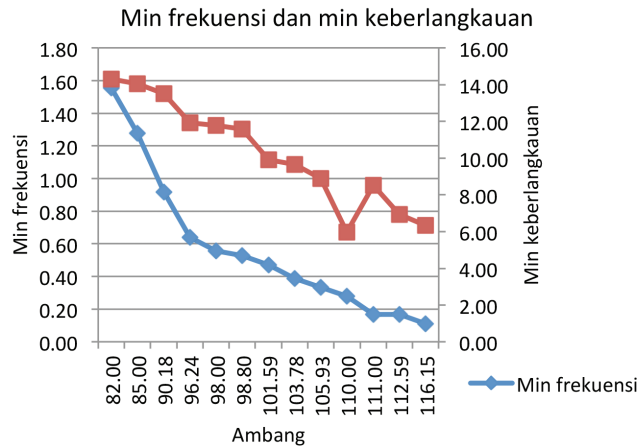
Jadual 6 menunjukkan nilai-p ujian AD bagi taburan GEV dan GPD di setiap stesen kajian. Bagi stesen pengawasan kualiti udara di Klang, Petaling Jaya, Cheras dan Batu Muda, taburan GEV merupakan taburan terbaik bagi menerangkan PM_{10} melampau dengan nilai-p

JADUAL 2. Nilai-nilai min frekuensi di setiap ambang

Persentil	0.95	0.96	0.97	0.98	0.981	0.983	0.985	0.987	0.989	0.991	0.993	0.995	0.997	0.999
Ambang	82	85	90.18	96.24	98	98.8	101.59	103.78	105.93	110	111	112.59	116.15	125.25
Jan-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Feb-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mar-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apr-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0
May-09	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Jun-09	8	7	4	3	3	3	3	3	3	3	2	2	2	2
Jul-09	11	9	7	6	4	4	4	4	3	2	0	0	0	0
Aug-09	6	6	6	4	4	3	2	2	2	2	2	2	1	0
Sep-09	3	2	2	1	1	1	1	0	0	0	0	0	0	0
Oct-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nov-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dec-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jan-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Feb-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mar-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apr-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
May-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jun-10	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-10	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Sep-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Oct-10	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Nov-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dec-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jan-11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Feb-11	1	1	1	1	1	1	1	1	0	0	0	0	0	0
Mar-11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apr-11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
May-11	3	3	3	3	3	3	2	1	1	1	0	0	0	0
Jun-11	4	4	3	1	1	1	1	0	0	0	0	0	0	0
Jul-11	12	10	6	3	3	3	3	3	3	2	2	2	1	0
Aug-11	2	1	0	0	0	0	0	0	0	0	0	0	0	0
Sep-11	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Oct-11	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Nov-11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dec-11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jumlah	56	46	33	23	20	19	17	14	12	10	6	6	4	2
Min frekuensi	1.56	1.28	0.92	0.64	0.56	0.53	0.47	0.39	0.33	0.28	0.17	0.17	0.11	0.06

JADUAL 3. Purata PM₁₀ melangkaui ambang dan min keberlangkauan di Shah Alam

Persentil	Ambang	Purata PM ₁₀ melangkaui ambang	Min keberlangkauan
0.95	82.00	96.30	14.30
0.96	85.00	99.05	14.05
0.97	90.18	103.69	13.51
0.98	96.24	108.17	11.93
0.981	98.00	109.79	11.79
0.983	98.80	110.39	11.59
0.985	101.59	111.52	9.93
0.987	103.78	113.43	9.65
0.989	105.93	114.83	8.90
0.991	110.00	115.96	5.96
0.993	111.00	119.51	8.51
0.995	112.59	119.51	6.92
0.997	116.15	122.50	6.35
0.999	125.25	127.77	2.52



RAJAH 1. Plot min frekuensi dan min keberlangkauan di Shah Alam

JADUAL 4. Keputusan pemilihan ambang bagi semua stesen

Klang	Stesen				
	Petaling Jaya	Shah Alam	Putrajaya	Cheras	Batu Muda
104.24	81.24	85.00	68.00	78.24	76.30

JADUAL 5. Anggaran parameter untuk setiap stesen kajian parameter bagi taburan GEV dan GPD

Stesen	Parameter	Taburan	
		GEV	GPD
Klang	Lokasi($\hat{\epsilon}$)	88.2366	101.5964
	Skala($\hat{\alpha}$)	22.9177	32.3733
	Bentuk($\hat{\kappa}$)	-0.1507	0.0960
Petaling Jaya	Lokasi($\hat{\epsilon}$)	66.3344	82.5313
	Skala($\hat{\alpha}$)	17.2171	10.6983
	Bentuk($\hat{\kappa}$)	-0.0831	0.0264
Shah Alam	Lokasi($\hat{\epsilon}$)	73.4930	85.0782
	Skala($\hat{\alpha}$)	15.3343	20.5382
	Bentuk($\hat{\kappa}$)	-0.1006	0.3700
Putrajaya	Lokasi($\hat{\epsilon}$)	54.4515	68.1411
	Skala($\hat{\alpha}$)	15.1459	15.3028
	Bentuk($\hat{\kappa}$)	-0.1006	0.1485
Cheras	Lokasi($\hat{\epsilon}$)	66.0970	77.5132
	Skala($\hat{\alpha}$)	14.0124	17.7267
	Bentuk($\hat{\kappa}$)	-0.1816	0.0492
Batu Muda	Lokasi($\hat{\epsilon}$)	62.3583	77.1476
	Skala($\hat{\alpha}$)	14.0933	13.3683
	Bentuk($\hat{\kappa}$)	0.1324	-0.0116

0.8570, 0.9960, 0.9721 dan 0.9465. Manakala taburan GPD merupakan taburan parametrik terbaik di Shah Alam dan Putrajaya dengan nilai-p 0.9953 dan 0.9244. Melalui keputusan ujian AD ini, pemodelan PM_{10} melampau dengan menggunakan kaedah blok maksima merupakan kaedah yang sesuai digunakan di stesen pengawasan kualiti udara Klang, Petaling Jaya, Cheras dan Batu Muda. Manakala kaedah puncak melangkaui ambang dilihat lebih sesuai diaplikasikan di Shah Alam dan Putrajaya.

PEMODELAN TABURAN TAK BERPARAMETER

Penganggaran taburan tak berparameter menggunakan fungsi ketumpatan kernel memerlukan pemilihan lebar jalur, h dan fungsi kernel, K yang sesuai. Dalam kajian ini, fungsi kernel yang digunakan adalah fungsi kernel Gaussian. Manakala kaedah pemilihan lebar jalur yang digunakan dalam bab ini adalah dengan menggunakan kaedah *plug-in* yang telah dihasilkan oleh Polansky dan Baker pada tahun

JADUAL 6. Nilai-p ujian statistik Anderson Darling bagi taburan GEV dan GPD

Stesen	Nilai-p Ujian Anderson Darling	
	GEV	GPD
Klang	0.8570*	0.8513
Petaling Jaya	0.9960*	0.0001
Shah Alam	0.8717	0.9953*
Putrajaya	0.8562	0.9244*
Cheras	0.9721*	0.6452
Batu Muda	0.9465*	0.0001

*Taburan terbaik bagi setiap stesen

JADUAL 7. Lebar jalur (*h*) bagi PM₁₀ maksimum bulanan

Stesen	Lebar Jalur
Klang	12.1446
Petaling Jaya	9.62301
Cheras	7.0053
Batu Muda	7.1845

JADUAL 8. Lebar jalur (*h*) bagi PM₁₀ melangkaui ambang

Stesen	Lebar Jalur
Shah Alam	3.945795
Putrajaya	4.398074

2000. Pemilihan lebar jalur dalam kajian ini dilakukan dengan menggunakan pakej *kerdiest* yang terdapat dalam perisian *R*. Jadual 7 dan 8 menunjukkan lebar jalur bagi PM₁₀ maksimum bulanan dan PM₁₀ melangkaui ambang yang telah diperolehi dengan menggunakan kaedah ini.

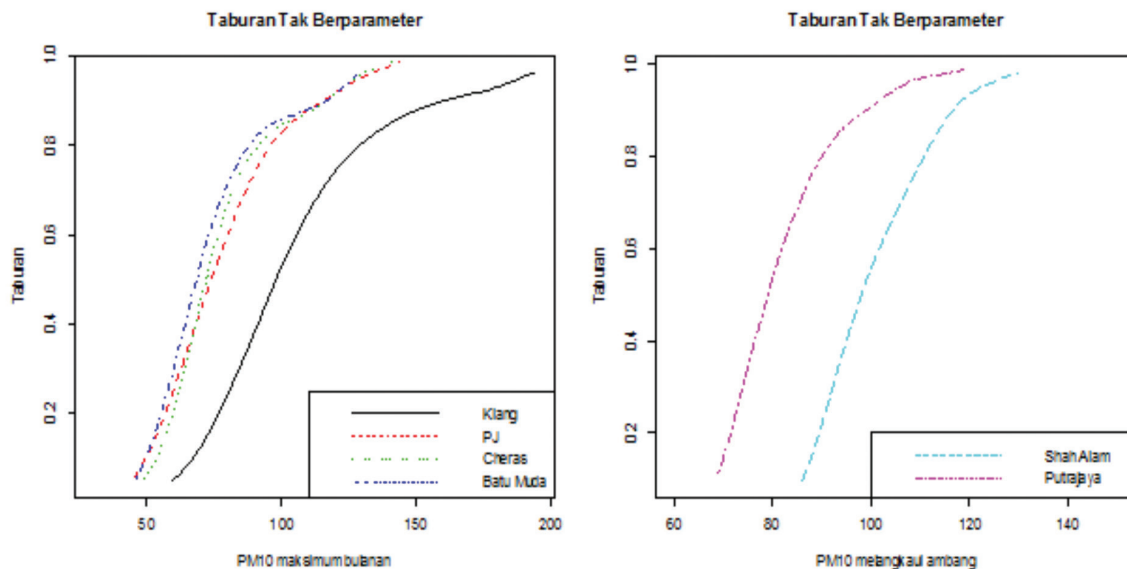
Lebar jalur yang dipilih adalah lebar jalur yang meminimumkan penghampiran asimptot ralat *MISE*.

Lebar jalur yang telah diperolehi akan digunakan dalam menentukan bentuk fungsi taburan tak berparameter dengan bertindak sebagai fungsi pelicin dalam mendapatkan fungsi taburan tak berparameter. Rajah 2 menunjukkan graf fungsi taburan tak berparameter bagi PM₁₀ maksimum bulanan dan PM₁₀ melangkaui ambang berdasarkan lebar jalur yang telah diperolehi mengikut stesen kajian.

Berdasarkan Rajah 2, taburan tak berparameter bagi stesen pengawasan kualiti udara Klang lebih ke kanan berbanding taburan tak berparameter di stesen yang lain kerana kepekatan PM₁₀ maksimum bulanan di Klang adalah lebih tinggi berbanding di kawasan lain. Begitu juga dengan taburan tak berparameter bagi stesen pengawasan kualiti udara Shah Alam, kepekatan PM₁₀ melangkaui ambang di Shah Alam adalah lebih tinggi berbanding di Putrajaya menyebabkan taburan tak berparameter di Shah Alam lebih ke kanan berbanding di Putrajaya.

PERBANDINGAN MODEL TAK BERPARAMETER

Setelah taburan terbaik bagi mewakili data PM₁₀ maksimum bulanan dan PM₁₀ melangkaui ambang dikenal pasti, langkah seterusnya adalah mendapatkan taburan



RAJAH 2. Fungsi taburan tak berparameter PM₁₀ maksimum bulanan dan PM₁₀ melangkaui ambang mengikut stesen kajian

yang lebih baik antara taburan parametrik dengan taburan tak berparameter berdasarkan ujian *MSE*. Nilai *MSE* yang lebih kecil menunjukkan taburan yang lebih baik. Lajur terakhir pada Jadual 9 dan 10 menunjukkan nilai *MSE* taburan parametrik dan tak berparameter untuk semua stesen kajian bagi PM_{10} maksimum bulanan dan PM_{10} melangkaui ambang pada nilai ambang c_k yang berlainan.

Tanda “*” pada nilai *MSE* menandakan nilai *MSE* terkecil dan menunjukkan taburan yang lebih baik antara taburan parametrik dan tak berparameter bagi setiap stesen kajian. Berdasarkan jadual, taburan GEV menunjukkan nilai *MSE* yang lebih kecil berbanding taburan tak berparameter di stesen pengawasan kualiti udara di Petaling Jaya. Manakala nilai *MSE* bagi tiga stesen yang berbaki iaitu stesen Klang, Cheras dan Batu Muda menunjukkan nilai yang lebih kecil bagi taburan tak berparameter berbanding taburan GEV. Dapat disimpulkan bahawa, satu taburan parametrik (GEV) dan tiga taburan tak berparameter merupakan taburan terbaik bagi pemodelan data PM_{10} maksimum bulanan di Lembah Klang.

Berdasarkan Jadual 10, taburan GPD merupakan taburan terbaik di Putrajaya, manakala taburan tak berparameter merupakan taburan terbaik di Shah Alam bagi pemodelan data PM_{10} melangkaui ambang. Secara keseluruhannya, empat taburan tak berparameter merupakan taburan terbaik berbanding taburan parametrik bagi pemodelan data PM_{10} melampaui di Lembah Klang. Seterusnya, pengiraan tempoh ulangan akan dilakukan berdasarkan taburan terbaik yang telah diperoleh daripada hasil pengiraan *MSE*.

TEMPOH ULANGAN

Pengiraan tempoh ulangan, $T(c)$ akan dilakukan bagi setiap stesen berdasarkan taburan terbaik yang telah diperoleh dalam bahagian sebelum ini. Tempoh ulangan akan dihitung dengan menggunakan rumus (17). Fungsi taburan, $F(x)$ terlebih dahulu perlu dihitung dengan menggunakan taburan terbaik yang mewakili stesen kajian. Tempoh ulangan akan dihitung berdasarkan paras berbahaya bagi

JADUAL 9. Nilai *MSE* bagi PM_{10} maksimum bulanan

Stesen		C_k						MSE
		50	75	100	120	150	180	
Klang	<i>ND</i>	36	30	16	9	4	3	
	$E(ND)_{GEV}$	36	30	16	9	4	2	0.4317
	$E(ND)_h$	35	29	17	9	4	3	0.3191*
Petaling Jaya	<i>ND</i>	34	17	6	3	0	0	
	$E(ND)_{GEV}$	35	16	5	2	1	0	0.2915*
	$E(ND)_h$	32	17	6	3	0	0	0.7586
Cheras	<i>ND</i>	34	14	5	3	0	0	
	$E(ND)_{GEV}$	35	15	5	2	1	0	0.7311
	$E(ND)_h$	34	15	6	3	0	0	0.3444*
Batu Muda	<i>ND</i>	34	13	5	4	0	0	
	$E(ND)_{GEV}$	33	13	4	2	1	0	1.2105
	$E(ND)_h$	32	13	5	3	0	0	0.7441*

JADUAL 10. Nilai *MSE* bagi PM_{10} melangkaui ambang

Stesen		C_k						MSE
		75	90	100	120	150	180	
Shah Alam	<i>ND</i>	28	9	3	0	0	0	
	$E(ND)_{GPD}$	27	9	4	0	0	0	0.2532
	$E(ND)_h$	28	9	4	0	0	0	0.2065*
Putrajaya	<i>ND</i>	43	33	18	2	0	0	
	$E(ND)_{GPD}$	43	33	18	3	0	0	0.2176*
	$E(ND)_h$	43	34	19	3	0	0	0.3169

JADUAL 11. Tempoh ulangan bagi PM_{10}

Stesen	Klang	Petaling Jaya	Cheras	Batu Muda	Shah Alam	Putrajaya
$F(150)$	0.8783	0.9832	0.9972	0.9999	0.9999	0.9999
$T(150)$	8.2149	59.76103	360.7823	13886.17	156976	42258.02

PM₁₀ yang telah dikeluarkan oleh Jabatan Alam Sekitar iaitu 150 ug/m³.

Berdasarkan Jadual 11, stesen pengawasan kualiti udara Klang menunjukkan tempoh ulangan paling cepat bagi PM₁₀ berbanding stesen pengawasan lain di sekitar Lembah Klang. Tempoh ulangan bagi stesen Klang ialah 8.2149 bulan atau sekali setahun. Ini bermakna kawasan sekitar Klang lebih kerap tercemar dengan PM₁₀ berbanding kawasan lain di sekitar Lembah Klang. Tempoh ulangan ini adalah relevan kerana hanya data PM₁₀ maksimum bulanan di stesen Klang sahaja yang melebihi nilai garis panduan yang telah ditetapkan oleh JAS iaitu pada 150 ug/m³. Walaupun stesen pengawasan kualiti udara yang lain menunjukkan tempoh ulangan yang lebih lama berbanding Klang, kawalan terhadap pencemaran udara yang mengandungi PM₁₀ perlu dilakukan di sekitar Lembah Klang agar kesan negatif seperti kemudaratan terhadap kesihatan manusia dapat dielakkan.

KESIMPULAN

Hasil kajian mendapati taburan tak berparameter merupakan taburan yang lebih baik bagi mewakili data PM₁₀ melampau di kebanyakan stesen pengawasan kualiti udara di Lembah Klang. Bagi data PM₁₀ maksimum bulanan, satu taburan parametrik (GEV) merupakan taburan yang lebih baik bagi mewakili stesen pengawasan kualiti udara Petaling Jaya. Manakala, tiga taburan tak berparameter merupakan taburan terbaik di stesen pengawasan kualiti udara Klang, Cheras dan Batu Muda. Perbandingan taburan menggunakan data PM₁₀ melangkaui ambang pula mendapati, satu taburan parametrik (GPD) merupakan taburan yang lebih baik bagi mewakili stesen pengawasan kualiti udara Putrajaya dan satu taburan tak berparameter adalah taburan terbaik di stesen pengawasan kualiti udara Shah Alam. Hasil pengiraan tempoh ulangan, stesen pengawasan kualiti udara Klang mempunyai tempoh ulangan yang paling cepat berbanding stesen pengawasan kualiti udara yang lain pada paras kepekatan PM₁₀ 150 ug/m³. Hasil daripada kajian ini mendapati, secara keseluruhannya, paras kepekatan PM₁₀ di sekitar Lembah Klang masih lagi dalam keadaan terkawal. Walau bagaimanapun, kawalan terhadap punca-punca pencemaran udara yang mengandungi PM₁₀ seperti pelepasan asap kenderaan dan kilang serta pembakaran terbuka perlu dilakukan bagi memastikan risiko pendedahan kepada pencemaran udara khususnya kesan negatif daripada PM₁₀ dapat dikurangkan.

PENGHARGAAN

Penghargaan kepada Universiti Kebangsaan Malaysia (UKM) di atas geran DPP-2015-099 dan FRGS/1/2015/ST06/UKM/02/2 serta Jabatan Alam Sekitar Malaysia (JAS) di atas data bahan pencemar iaitu PM₁₀. Tidak lupa kepada Prof. Dr. Mohd Talib Latif di atas bantuan dan sokongan beliau dalam kajian ini.

RUJUKAN

- Abdullah, A.M., Abu Samah, M.A. & Jun, T.Y. 2012. An overview of the air pollution trend in Klang Valley, Malaysia. *Open Environmental Sciences* 6: 13-19.
- Afroz, R., Hassan, M.N. & Ibrahim, N.A. 2003. Review of air pollution and health impacts in Malaysia. *Environmental Research* 92(2): 71-77.
- Altman, N. & Leger, C. 1995. Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference* 46: 195-214.
- Anderson, T.W. & Darling, D.A. 1954. A test for goodness of fit. *The Journal of American Statistical Association* 49: 765-769.
- Awang, M.B., Jaafar, A.B., Abdullah, A.M., Ismail, M.B., Hassan, M.N., Abdullah, R., Johan, S. & Noor, H. 2000. Air quality in Malaysia: Impacts, management issues and future challenges. *Respirology* 5(2): 183-196.
- Beguerra, S. 2005. Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology* 303(1): 215-230.
- Bowman, A., Hall, P. & Prvan, T. 1998. Bandwidth selection for the smoothing of distribution functions. *Biometrika* 85(4): 799-808.
- Brook, R.D., Franklin, B., Cascio, W., Hong, Y., Howard, G., Lipsett, M., Luepker, R., Mittleman, M., Samet, J. & Smith Jr, S.C. 2004. Air pollution and cardiovascular disease: A statement for healthcare professionals from the expert panel on population and prevention science of the American Heart Association. *Circulation* 109(21): 2655-2671.
- Coles, S., Bawa, J., Trenner, L. & Dorazio, P. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Vol. 208. London: Springer.
- Davison, A.C. 1984. Modelling excesses over high thresholds, with an application. In *Statistical Extremes and Applications*. Netherlands: Springer.
- Dominick, D., Juahir, H., Latif, M.T., Zain, S.M. & Aris, A.Z. 2012. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmospheric Environment* 60: 172-181.
- Hosking, J.R. & Wallis, J.R. 1997. *Regional Frequency Analysis*. Cambridge: The press syndicate of the University of Cambridge.
- Hosking, J.R.M. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society* 52: 105-124.
- Hurairah, A., Akma Ibrahim, N., Bin Daud, I. & Haron, K. 2005. An application of a new extreme value distribution to air pollution data. *Management of Environmental Quality: An International Journal* 16(1): 17-25.
- Jabatan Alam Sekitar Malaysia. 2015. Indeks Pencemaran Udara. <http://www.apims.doe.gov.my/>. Diakses pada 8 April 2015.
- Katiman, R. 2006. Perbandaran dan perkembangan wilayah metropolitan lanjutan lembah Klang-Langat, Malaysia. *Jurnal Elektronik Sains Sosial dan Kemanusiaan* 1(1): 1-27.
- Keywood, M.D., Ayers, G.P., Gras, J.L., Boers, R. & Leong, C.P. 2003. Haze in the Klang Valley of Malaysia. *Atmos. Chem. Phys* 3: 615-653.
- Kuchenhoff, H. & Thamerus, M. 1995. Extreme value analysis of Munich air pollution data. *Sonderforschungsbereich* 386(4): 1-24.
- Kostova, S.P., Rumchev, K.V., Vlaev, T. & Popova, S.B. 2012. Using copulas to measure association between air pollution

- and respiratory diseases. *International Scholarly and Scientific Research & Innovation* 6(11): 533-538.
- Lang, M., Ouarda, T.B.M.J. & Bobée, B. 1999. Towards operational guidelines for over-threshold modeling. *Journal of hydrology* 225(3): 103-117.
- Masseran, N., Razali, A.M., Ibrahim, K. & Latif, M.T. 2016. Modeling air quality in main cities of Peninsular Malaysia by using a generalized Pareto model. *Environmental Monitoring and Assessment* 188(1): 1-12.
- Mohd Yusoff, M.R., Hassan, H. & Mohd Zain, I. 1989. Kesan angin dan hujan ke atas penyebaran & kepekatan zarah-han terampai di Kuala Lumpur. *Jurnal Teknologi* 13: 54-65.
- Parzen, E. 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33(3): 1065-1076.
- Polansky, A.M. & Baker, E.R. 2000. Multistage plug-in bandwidth selection for kernel distribution function estimates. *Journal of Statistical Computation and Simulation* 65: 63-80.
- Quintela-del-Rio, A. & Estevez-Perez, G. 2012. Nonparametric kernel distribution function estimation with kerdie: An R package for bandwidth choice and applications. *Journal of Statistical Software* 50(8): 1-20.
- Quintela-del-Rio, A. & Francisco-Fernandez, M. 2011. Analysis of high level ozone concentrations using nonparametric methods. *Science of the Total Environment* 409: 1123-1133.
- Sahani, M., Zainon, N.A., Wan Mahiyudin, W.R., Latif, M.T., Hod, R., Khan, M.D., Mohd Tahir, M. & Chah, C.C. 2014. A case-crossover analysis of forest fire haze events and mortality in Malaysia. *Atmospheric Environment* 96: 257-265.
- Santus, P., Russo, A., Madonini, E., Allegra, L., Blasi, F., Centanni, S., Madonini, A., Schiraldi, G. & Amaducci, S. 2012. How air pollution influences clinical management of respiratory diseases: A case-crossover study in Milan. *Respiratory Research* 13(95): 1-12.
- Silverman, B.W. 1986. *Density Estimation for Statistics And Data Analysis*. London: Chapman & Hall.
- Stephens, M.A. 1976. Asymptotic results for goodness-of-fit statistics with unknown parameters. *Annals of Statistics* 4: 357-369.
- Tobias, A. & Scotto, M.G. 2005. Prediction of extreme ozone levels in Barcelona, Spain. *Environmental Monitoring and Assessment* 100: 23-32.
- Vogel, R.M. & Fennessey, N.M. 1993. L-moment diagrams should replace product moment diagrams. *Water Resources Research* 29: 1745-1752.
- Wan Mahiyuddin, W.R., Sahani, M., Aripin, R., Latif, M.T., Thach, T.Q. & Wong, C.T. 2013. Short-term effects of daily air pollution on mortality. *Atmospheric Environment* 65: 69-79.
- World Health Organization. 2015. Air Pollution. http://www.who.int/topics/air_pollution/en/. Diakses pada 8 April 2015.
- Zhou, S.M., Deng, Q.H. & Liu, W.W. 2012. Extreme air pollution events: Modeling and prediction. *Journal of Central South University* 19: 1668-1672.
- Pusat Pengajian Sains Matematik
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia
- *Pengarang untuk surat-menyurat; email: aslammohdsafari@gmail.com
- Diserahkan: 2 Oktober 2015
Diterima: 7 Disember 2016