

**PENJEJAKAN PANJANG-INGATAN DATA SIRI MASA DENGAN
PENDEKATAN KUANTIL**

MUZIRAH BINTI MUSA

**TESIS YANG DIKEMUKAKAN UNTUK MEMPEROLEH IJAZAH
DOKTOR FALSAFAH**

**FAKULTI SAINS DAN TEKNOLOGI
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI**

2014

ABSTRAK

Kehadiran autokorelasi merupakan isu utama dalam analisis siri masa. Kebiasaannya autokorelasi lat antara data menyusut secara eksponen ke sifar. Apabila penyusutan ke sifar berlaku secara hipobolik, nilai cerapan akan kelihatan terus meningkat untuk suatu tempoh yang agak lama sebelum menyusut ke paras yang baharu. Penyusutan juga berlaku untuk tempoh yang juga agak lama sebelum meningkat semula. Fenomena sebegini dikenali sebagai panjang-ingatan. Kepentingan sifat panjang-ingatan ini banyak diperbincangkan oleh ramai penyelidik dari pelbagai disiplin ilmu. Selain keupayaannya mengambil kira data berkorelasi, pengetahuan tentangnya dilihat sebagai satu landasan ukuran keberkesanan perlaksanaan sesuatu rangka strategi kawalan. Sifat panjang-ingatan yang memperlihatkan ciri-ciri penskalaan boleh membawa ke arah penyediaan laporan pemantauan yang efektif dengan kerangka yang ringkas namun mempunyai maklumat yang menyeluruh serta mudah dicapai, sekali gus dapat meminimumkan kos pemantauannya. Kehadiran panjang-ingatan disukat berasaskan pekali Hurst yang boleh diperoleh dengan pelbagai kaedah. Kajian ini telah mengubahsuai sebahagian daripada kaedah yang ada menerusi pendekatan kuantil. Kaedah Nisbah serakan Berskala (NSB) dan kaedah Nyah-trend Bergerak (NTB) adalah dua kaedah ubahsuai yang dicadangkan dan keupayaannya diselidiki secara simulasi. Penerapannya ke atas data ozon yang di rekod bagi tempoh 9 hingga 12 tahun (1998-2009) untuk 34 buah stesen pemantauan kualiti udara di Semenanjung Malaysia juga turut dilakukan. Anggaran pekali Hurst yang diperoleh menggunakan kaedah ubahsuai memberikan hasil anggaran yang setanding dan lebih baik daripada kaedah sedia ada. Ini bermakna kaedah yang ditawarkan ini boleh dijadikan sebagai alternatif kepada kaedah-kaedah yang sedia ada. Seterusnya menggunakan kaedah yang sedia ada dan kaedah ubahsuai, anggaran nilai pekali Hurst untuk stesen-stesen kajian dilakukan. Indeks Hurst Terlaras (IHT) yang dibina bagi menerangkan fenomena panjang-ingatan secara reruang, kemudiannya dipetakan ke seluruh Semenanjung Malaysia. Pemetaan dilakukan menggunakan kaedah Krige. Kajian ini mendapati pola kehadiran panjang-ingatan berlaku di kebanyakan kawasan bandar dan perindustrian di pantai barat dan beberapa kawasan pantai timur Semenanjung Malaysia. Nilai anggaran pekali Hurst diperoleh adalah di antara 0.54 dan 0.99. Penerapan kaedah-kaedah ini, juga memperlihatkan perkaitan ciri-ciri penskalaan dan sifat swa-serupa dengan sifat panjang-ingatan. Ianya dirujuk kepada keserupaan dalam perlakuan turun-naik aras kepekatan ozon maksimum harian dalam selang masa kecil dan selang masa besar. Hasil kajian yang diperoleh adalah serupa dengan hasil kajian dari negara-negara lain.

DETECTION OF LONG-MEMORY TIME SERIES DATA WITH QUANTILE APPROACH

ABSTRACT

The presence of autocorrelation is a major issue in time series analysis. Typically, autocorrelation occurs when the lag between data decay exponentially to zero. When decaying occurs hyperbolically to zero, the observation would seem to increase to a further considerable period of time before falling to a new level. Decay also occur for quite long periods before it rise again. Such a phenomenon is known as long-memory. The importance of long-memory properties have been discussed by many researchers from various disciplines. In addition to its ability to take into account the correlation in the data, the knowledge of long-memory is use full for assessing the effectiveness of measures of control strategies. The nature of long-memory is characterized by scaling which can lead to the provision of effective monitoring report with a simple architecture though it is comprehensive and easily accessible, thus helping to minimize the monitoring cost. The presence of long-memory is measured based on Hurst coefficient which can be obtained by various methods. This study was to modify some of the existing methods through quantile approach. The Scaled Dispersion Ratio (NSB) and the Moving Detrended (NTB) are the two modified methods that been proposed and their ability are investigated through simulation. Also, these methods are applied to the ozone data recorded for a period of 9 to 12 years (1998-2009) for 34 air quality monitoring stations in Peninsular Malaysia. Hurst coefficient estimates that obtained using the modified methods give comparable result and much better than the estimates results of the existing methods. This means that, the method can be offered as an alternative to the methods currently available. Next, using the existing method and modified method, the estimated value of the Hurst coefficient for stations under study is done. Adjusted Hurst Index (IHT) constructed to characterize long-memory phenomenon spatially, were then mapped to the whole of Peninsular Malaysia. Mapping is done using the Krige. The study found that the presence of long-memory patterns occur in many urban and industrial areas on the west coast and some areas of the east coast of Peninsular Malaysia. The estimated Hurst coefficient value obtained is in between 0.54 and 0.99. Application of these methods also showed the relevance of scaling characteristics and the nature of self-similar long-memory properties. It is referred to the similarity in the fluctuation behavior for daily maximum ozone concentration level in small time intervals and large time intervals. The results found in this study are similar to findings from other countries.

KANDUNGAN

	Halaman
PENGAKUAN	ii
PENGHARGAAN	iii
ABSTRAK	iv
ABSTRACT	v
KANDUNGAN	vi
SENARAI JADUAL	xiii
SENARAI ILUSTRASI	xv
SENARAI SIMBOL	xix
SENARAI SINGKATAN	xxi
BAB I	PENDAHULUAN
1.1	Motivasi Kajian 1
1.2	Pernyataan Masalah 9
1.3	Objektif Kajian 10
1.4	Inovasi dan Sumbangan Kajian 11
1.5	Pengorganisasian Tesis 13
BAB II	PENJEJAKKAN SIRI MASA PANJANG-INGATAN
2.1	Pengenalan 15
2.2	Siri Masa 15
2.2.1	Prinsip Asas Siri Masa: Domain Masa 16
2.2.1.1	Proses Kepegunan dan Fungsi Autokorelasi 17
2.2.1.2	Model Box-Jenkins (ARIMA) 19
2.2.2	Prinsip Asas Siri Masa: Domain Frekuensi 20
2.2.3	Siri Masa Pendek-Ingatan lawan Siri Masa Panjang-ingatan 21
2.3	Ciri-Ciri Siri Masa Panjang-Ingatan 25
2.3.1	Proses Swa-Serupa Dan Ciri-ciri Penskalaan 26

2.4	Permodelan Siri Masa Panjang-Ingatan	29
2.4.1	Model Pergerakan Brownian Pecahan, fBm dan Hingar Gaussian Pecahan, fGn.	30
2.4.2	Penunjuk Panjang-ingatan: Pekali <i>HURST</i> , H	32
2.5	Kaedah Penjejakan Kehadiran Panjang-Ingatan Suatu Siri Masa Dengan Penganggaran Pekali <i>Hurst</i> , H	34
2.5.1	Kaedah Julat Penskalaan Semula, R/S	35
2.5.2	Kaedah Varians Penskalaan Semula, V/S	36
2.5.3	Kaedah Varians Tertingkap Berskala, SWV	37
2.5.4	Kaedah Analisis Nyah-trend Turun-naik, DFA	38
2.5.5	Kaedah Nyah-trend Purata Bergerak, DMA	40
2.6	Ilustrasi Kaedah-Kaedah Penganggaran Pekali <i>Hurst</i>	41
2.6.1	Ilustrasi kaedah R/S , V/S dan SWV	43
2.6.2	Ilustrasi kaedah DFA dan DMA	49
2.7	Panjang-Ingatan Dalam Pelbagai Disiplin Ilmu	55
2.8	Rumusan	60
2.9	Kerangka Teori	61
BAB III	KAEDAH NISBAH SERAKAN BERSKALA (NSB) SUATU KAEDAH ALTERNATIF KEPADA PENGANGGARAN PEKALI HURST	
3.1	Pengenalan	63
3.2	Sekilas Tinjauan Kajian Kepelbagaian Kaedah Penambahbaikan Penganggaran Pekali <i>Hurst</i> , H	65
3.3	Penginovasian Kaedah: Kaedah Nisbah Serakan Berskala, NSB	67
3.3.1	Ukuran Serakan: Persentil dan Kuantil	67
3.3.2	Julat Antara Kuartil, IQR dan Median Sisihan Mutlak, MED	68
3.3.3	Kaedah Nisbah Serakan Berskala, NSB	69
3.4	Pengujian Pengesahan Kaedah Ubahsuai (NSB) dan kaedah Klasik	72
3.4.1	Penjanaan Siri Masa Panjang-ingatan: Simulasi Davis & Harte	72

3.4.2	Ujian Pencapaian	75
3.4.3	Prosedur uji kaji	77
3.5	Hasil Analisis Ujikasi Secara Simulasi	78
3.5.1	Nilai Anggaran Pekali Hurst, $\hat{H}$	79
3.5.2	Nilai Pincang, Bs_H dan Min Ralat Kuasa Dua, $MRKD_H$	83
3.6	Kesan Panjang Data Ke Atas Kaedah Penganggaran Pekali Hurst, H	86
3.6.1	Analisis Kategori Pertama: Ukuran Serakan Bagi Kedua-dua Pengangka dan Penyebut Dengan Ciri-Ciri Sukatan yang Sama.	87
3.6.2	Analisis Kategori Kedua: Ukuran Serakan Bagi Kedua-dua Pengangka dan Penyebut Dengan Ciri-Ciri Sukatan Yang Berbeza.	89
3.7	Ujian Pencapaian Berdasarkan Pekali Korelasi Konkordan Lin (LCC)	92
3.8	Pengujian Kaedah Ke Atas Data Dengan Kewujudan Gangguan	94
3.8.1	Kesan Terhadap Nilai Anggaran Pekali Hurst, $\hat{H}$	94
3.8.2	Kesan Terhadap Kebolehubahan (Variasi)	96
3.9	Rumusan	98
BAB IV	KAEDAH NYAH-TREND BERGERAK (NTB) SUATU KAEDAH ALTERNATIF PENGANGGARAN PEKALI HURST	
4.1	Pengenalan	99
4.2	Penjejakan Kaedah Nyah-trend: DFA Dan DMA	100
4.2.1	Nyah-Trend Turun-naik, DFA	100
4.2.2	Nyah-tend Purata Bergerak, DMA	103
4.3	Peginovasian Kaedah: Kaedah Nyah-trend Bergerak, (NTB)	105
4.3.1	Kaedah Nyah-Trend Turun-naik Terhubung (DFA-bd)	107
4.3.2	Kaedah Nyah-Trend Median Bergerak (DMM).	108
4.3.3	Ukuran Serakan Berasaskan Kuantil	109

4.4	Ujian Pengesahan Dan Kebolehpercayaan Kaedah-kaedah Penganggaran Pekali Hurst	110
4.4.1	Penjanaan Data Simulasi, fBm: Data Stabil	110
4.4.2	Penjanaan Data Simulasi, fBm: Data dengan Kehadiran Trend	111
4.4.3	Ujian Pencapaian	112
4.5	Hasil Analisis Uji kaji Simulasi Siri Data Stabil	112
4.5.1	Nilai Anggaran Pekali Hurst, $\hat{H}$	112
4.5.2	Nilai Pincang, B_{S_H} Dan Min Ralat Kuasa Dua, $MRKD_H$	115
4.5.3	Statistik Ujian Ukuran Serakan Berasaskan Persentil atau Kuantil: IQR Dan MED	118
4.6	Hasil Analisis Simulasi Siri Data Dengan Kehadiran Trend Linear	122
4.7	Kesan Panjang Data Terhadap Kaedah-Kaedah Penganggaran	125
4.7.1	Nilai Anggaran Pekali Hurst, $\hat{H}$	126
4.7.2	Nilai Pincang, B_{S_H}	128
4.7.3	Nilai Min Ralat Kuasa Dua, $MRKD_H$	129
4.8	Rumusan	132
BAB V	SUMBER DAN PERIHAL DATA: PROFIL TABURAN ARAS KEPEKATAN OZON MAKSIMUM HARIAN DI SEMENANJUNG MALAYSIA	
5.1	Pengenalan	134
5.2	Ozon	134
5.2.1	Ozon di Permukaan (Troposfera)	135
5.3	Sumber Dan Data Kajian	139
5.4.1	Data Kajian	140
5.4.2	Rawatan Data Lenyap	142
5.4	Unit Analisis Dan Definisi Ozon Maksimum Harian	144
5.5	Corak Taburan Data Di Stesen-stesen Kajian	147

5.5.1	Pola Taburan Harian Purata Maksimum Harian Aras Kepekatan Ozon	148
5.5.2	Pola Taburan Bulanan Purata Maksimum Harian Aras Kepekatan Ozon	151
5.5.3	Taburan Aras Kepekatan Ozon Maksimum Bagi Tahun 2000 Dan 2005	153
5.6	Profil Aras Kepekatan Ozon Maksimum Harian Di Semenanjung Malaysia.	158
5.6.1	Pemetaan Aras Kepekatan Ozon Maksimum Harian Menerusi Kaedah Krige	158
5.6.2	Statistik Ikhtisar Aras Kepekatan Ozon Maksimum Harian di Semenanjung Malaysia	159
5.6.3	Purata Maksimum Bulanan Aras Kepekatan Ozon Maksimum Harian di Semenanjung Malaysia	161
5.7	Rumusan	163
BAB VI	PENJEJAKKAN PANJANG-INGATAN SIRI MASA ARAS KEPEKATAN OZON MAKSIMUM HARIAN DI SEMENANJUNG MALAYSIA	
6.1	Pengenalan	165
6.2	Tanda Awal Kehadiran Panjang-Ingatan	166
6.3	Keputusan Analisis Penganggaran Pekali Hurst Menggunakan Kaedah Ubahsuai Dan Kaedah Klasik	167
6.3.1	Anggaran Pekali Hurst, $\hat{H}$	169
6.3.2	Purata Berwajaran Anggaran Pekali Hurst, $\bar{\hat{H}}_w$.	176
6.3.3	Pola Taburan Anggaran Pekali Hurst, $\hat{H}$ Di Stesen Kajian.	181
6.4	Indeks Baharu Bagi Mencirikan Kawasan Panjang-Ingatan Di Semenanjung Malaysia	182
6.4.1	Indeks Hurst Terlaras, IHT	183
6.4.2	Pola Taburan Panjang-Ingatan Di Semenanjung Malaysia	185
6.5	Ciri-ciri Penskalaan Dan Sifat Keberterusan Aras Kepekatan Ozon Maksimum Harian	188

6.5.1	Ciri-ciri Penskalaan Dan Keberterusan Di Kawasan Bandar	189
6.5.2	Ciri-ciri Penskalaan Dan Keberterusan Di Kawasan Perindustrian	191
6.5.3	Ciri-ciri Penskalaan Dan Keberterusan Di Kawasan Pinggir Bandar	193
6.6	Rumusan	195
BAB VII	RUMUSAN DAN KESIMPULAN	
7.1	Rumusan Hasil Kajian	197
7.2	Kesimpulan Dan Penutup	200
	RUJUKAN	201
	LAMPIRAN	
A	Hasil Ujian Simulasi Ke Atas 100 Siri Data Siri Masa Menggunakan Kaedah Nisbah Serakan Berskala (NSB).	212
B	Hasil Ujian Simulasi Ke Atas 100 Siri Data Siri Masa Menggunakan Kaedah Nyah-Trend Bergerak (NTB).	236
C	Nilai Purata Anggaran, $\bar{\hat{H}}$, Had Atas Dan Had Bawah Bagi 95% Selang Keyakinan, Varians, Nilai Pincang Dan Min Ralat Kuasa Dua Bagi Kaedah Nisbah Serakan Berskala (NSB).	248
D	Nilai Purata Anggaran, $\bar{\hat{H}}$, Had Atas Dan Had Bawah Bagi 95% Selang Keyakinan, Varians, Nilai Pincang Dan Min Ralat Kuasa Dua Bagi Kaedah Nyah-Trend Bergerak (NTB).	252
E	Hasil Ujian Simulasi Ke Atas Data Yang Dicemar.	258
F	Hasil Analisis Profil Data Siri Masa Aras Kepekatan Ozon Maksimum Harian.	267
G	Hasil Analisis Panjang-Ingatan Ke Atas Data Siri Masa Aras Kepekatan Ozon Maksimum Harian.	268
H	Atur Cara R Bagi Menjana Siri Masa Panjang-Ingatan.	273
I	Atur Cara R Bagi Kaedah Klasik Dan Kaedah Ubah Suai : Kaedah Nisbah Serakan Berskala (NSB).	275

J	Atur Cara R Bagi Kaedah Klasik Dan Kaedah Ubah Suai : Kaedah Nyah-Trend Bergerak (NTB).	277
K	Atur Cara R Penjanaan Siri Masa fGn Dengan Gangguan.	284
L	Atur Cara R Penjanaan Siri Masa fBm Dengan Kehadiran Trend.	285
M	Atur Cara R Bagi Mendapatkan Plot Kotak Nilai Pekali Hurst (H) Menggunakan Kaedah NSB.	286
N	Atur Cara R Bagi Mendapatkan Plot Nilai Pincang (Bs) Menggunakan Kaedah NSB.	288
O	Atur Cara R Bagi Mendapatkan Plot Nilai Min Ralat Kuasa Dua (MRKD) Menggunakan Kaedah NSB.	290
P	Atur Cara R Untuk Plot Min Dan 95% Selang Keyakinan Bagi Nilai Pekali Hurst (H) Menggunakan Kaedah NTB.	292
Q	Atur Cara R Untuk Plot Nilai Pincang (Bs) Menggunakan Kaedah NTB.	295
R	Atur Cara R Untuk Plot Min Ralat Kuasa Dua (MRKD) Menggunakan Kaedah NTB	297
S	Atur Cara R Bagi Profil Data Aras Kepekatan Ozon Maksimum Harian.	299
T	Atur Cara R Analisis Panjang-Ingatan Data Aras Kepekatan Ozon Maksimum Harian.	302
U	Senarai Penerbitan.	304

SENARAI JADUAL

No. Jadual		Halaman
2.1	Pengklasifikasian proses fraktal dan pekali Hurst, H .	33
2.2	Sampel data untuk ilustrasi pengiraan kaedah penganggaran pekali H .	42
2.3	Data dan statistik bagi kaedah R/S dan V/S.	46
2.4	Data dan statistik bagi kaedah SWVld dan SWVbd.	48
2.5	Nilai log bagi kaedah R/S, V/S dan SWV dan anggaran nilai pekali Hurst $\hat{H}$.	49
2.6	Data dan statistik bagi kaedah DFA	54
2.7	Data dan statistik bagi kaedah DMA	55
3.1	Skala pencapaian pekali Konkordans Lin, ρ_c .	77
3.2	Peratusan kepincangan dan peratusan min ralat kuasa bagi $H = 0.5$ dan $H = 0.7$.	85
3.3	Peratusan kepincangan bagi siri masa dengan panjang data yang berbeza bagi $H = 0.5$ dan $H = 0.7$..	91
3.4	Pencapaian Kaedah penganggaran pekali Hurst berdasarkan ukuran pekali korelasi Konkordans Lin (LCC).	93
4.1	Keputusan analisis anggaran nilai H menggunakan ukuran serakan SD, MED dan IQR bagi $H = 0.4, 0.5$ dan 0.7 .	120
4.2	Analisis bagi data dengan kehadiran Trend Linear.	123
4.3	Kesan kehadiran trend terhadap penggunaan kaedah penganggaran pekali Hurst.	125

No. Jadual

Halaman

5.1	Maklumat data kajian bagi 34 buah stesen pemantauan kualiti udara di Semenanjung Malaysia.	143
5.2	Contoh data aras kepekatan ozon maksimum harian di stesen Ipoh, Perak (N3) untuk 20 hari yang pertama bagi tahun 1998.	146
6.1	Keputusan bagi kaedah Klasik R/S, V/S dan kaedah ubah suai Qc/S, Rc/Q, Rc/M dan Qc/Q (Kaedah Nisbah Serakan Berskala, NSB)	164
6.2	Jadual 6.2 Keputusan bagi kaedah klasik DFA, DMA dan kaedah ubah suai DMAw, DMMs dan DMMw. (Kaedah Nyah-trend Bergerak, NTB)	166
6.3	Contoh pengiraan purata berwajaran bagi 8 buah stesen di bahagian tengah Semenanjung Malaysia.	168
6.4	Purata berwajaran pekali Hurst dengan kaedah Nisbah Serakan Berskala (NSB).	170
6.5	Purata berwajaran pekali Hurst dengan kaedah Nyah-Trend Bergerak (NTB).	171
6.6	Indeks Hurst Terlaras bagi 34 buah stesen kajian.	174

SENARAI ILUSTRASI

No. Rajah		Halaman
2.1	Plot siri masa dan plot fungsi autokorelasi, ACF .	24
2.2	(a) Siri masa swa-serupa dengan skala masa n_1 dan n_2 . (b) Siri masa swa-serupa bagi skala masa n_1 dengan dua skala pembesaran M_y dan M_x .	26
2.3	Siri masa (a) fGn dan (b) fBm bagi tiga nilai $H = 0.2, 0.5$ dan 0.7 dengan $T=2048$ cerapan.	31
2.4	Ketumpatan spektral Siri masa (a) fGn dan (b) fBm bagi tiga nilai H adalah $0.2, 0.5$ dan 0.7 dengan $T=2048$ cerapan.	32
2.5	Pemetakan keseluruhan data, $N=256$ kepada M kemungkinan sub selang bersaiz n .	43
2.6	Log-log plot bagi kaedah R/S, V/R, SWVld dan SWVbd	49
2.7	Log-log plot kaedah DFA bagi saiz selang $n = 8, 16, 64$ dan 128 .	52
2.8	Log-log plot bagi kaedah DMA mudah, MAs dan berwajaran,MAw.	53
2.9	Tentatif Perjalanan Kajian	62
3.1	Gambaran pengukuran Pekali Konkordans Lin (LCC)	76
3.2	(a) Plot siri masa panjang-ingatan fGn dengan pekali $H = 0.7$ yang stabil dan (b), (c), (d) adalah data yang telah dicemar dengan peratus gangguan $10\%, 50\%, 100\%$ dan 350% .	78
3.3	Plot kotak bagi 100 nilai anggaran, $\hat{H}$ bagi kaedah klasik (a) dan (b), dan kaedah ubahsuai (c) hingga (l).	81
3.4	Nilai purata anggaran H bagi 100 siri masa	82
3.5	Nilai pincang, Bs_H dan min ralat kuasa dua, $MRKD_H$ bagi kaedah (a) Rc/S, Qc/S, Mc/S dan Vc/S, (b) Rc/M, Sc/M, Qc/M, dan Mc/M, dan (c) Rc/Q, Sc/Q, Qc/Q, dan Mc/Q.	84

3.6	(a) Nilai purata anggaran, $\bar{\hat{H}}$, (b) Nilai pincang, Bs_H (c) Nilai min ralat kuasa dua, $MRKD_H$ bagi siri masa dengan empat panjang yang berbeza bagi kaedah Mc/M, Rc/Q dan Qc/Q.	88
3.7	a) Nilai purata anggaran, $\bar{\hat{H}}$, (b) Nilai pincang, Bs_H (c) Nilai min ralat kuasa dua, $MRKD_H$ bagi siri masa dengan empat panjang yang berbeza diukur dengan kaedah Mc/S, Sc/M, Qc/S, Sc/Q, Qc/M, Mc/Q dan Rc/M.	90
3.8	Plot nilai purata anggaran, $\bar{\hat{H}}$ dengan peratusan gangguan bagi kaedah klasik (a) dan (b), dan kaedah ubah suai (c),(d),(e),(f),(g) dan (h). Paksi-x merujuk kepada nilai peratus dan paksi-y adalah nilai purata anggaran.	95
3.9	Plot min ralat kuasa dua, $MRKD_H$ dengan peratusan gangguan bagi kaedah klasik (a) dan (b), dan kaedah ubah suai (c),(d),(e),(f),(g) dan (h). Paksi-x merujuk kepada nilai peratus dan paksi-y adalah nilai min ralat kuasa dua.	97
4.1	Plot siri masa fBm data stabil dan data dengan kehadiran trend .	106
4.2	Nilai purata anggaran, $\bar{\hat{H}}$ dan 95% selang keyakinan bagi 100 nilai anggaran, $\hat{H}$. yang diukur dengan kaedah klasik (a) dan (c), dan kaedah ubah suai (b), (d), (e) dan (f).	111
4.3	Nilai Pincang, Bs_H dan min ralat kuasa dua, $MRKD_H$. bagi perbandingan antara kaedah klasik dan kaedah ubah suai.	114
4.4	Nilai Pincang, Bs_H dan Min ralat kuasa dua, $MRKD_H$ bagi data stabil yang diperoleh menggunakan kaedah (a) DFA-l dan DFA-bd, (b) DMAs dan DMMs, dan (c) DMAw dan DMMw dengan tiga ukuran serakan, SD, MED dan IQR.	121

No. Rajah

Halaman

4.5	Nilai Pincang, Bs_H dan min ralat kuasa dua, $MRKD_H$ bagi data dengan kehadiran trend yang diperolehi menggunakan kaedah (a) DFA-l dan DFA-bd, (b)DMAs dan DMMs, dan (c) DMAw dan DMMw dengan tiga ukuran serakan, SD, MED dan IQR.	124
4.6	Nilai purata anggaran, $\bar{H}$ bagi kaedah (a) DFA-l, (b) DMAw dan (c) DMMw yang diuji ke atas data stabil dan data dengan kehadiran trend untuk empat siri data S1, S2, S3 dan S4 dengan panjang yang berbeza.	127
4.7	Nilai pincang, Bs_H bagi siri masa S1, S2, S3 dan S4 stabil dan dengan kehadiran trend untuk kaedah (a) DFA-l, (b) DMAw dan (c) DMMw dengan ukuran SD, MED dan IQR.	130
4.8	Nilai min ralat kuasa dua, $MRKD_H$ bagi siri masa S1, S2, S3 dan S4 stabil dan dengan kehadiran trend untuk kaedah (a) DFA-l, (b) DMAw dan (c) DMMw dengan ukuran SD, MED dan IQR.	131
5.1	Pembentukan ozon di udara.	136
5.2	Peta Malaysia.	139
5.3	Lokasi stesen pemantauan kualiti udara di Semenanjung Malaysia.	141
5.4	Plot-kotak bagi taburan aras kepekatan ozon maksimum harian bagi stesen-stesen kajian.	147
5.5	Pola taburan harian di kawasan bandar.	149
5.6	Pola taburan harian di kawasan perindustrian.	150
5.7	Pola taburan harian di kawasan pinggir bandar.	150
5.8	Pola taburan bulanan di kawasan bandar.	151
5.9	Pola taburan bulanan di kawasan perindustrian.	152
5.10	Pola taburan bulanan di kawasan pinggir bandar.	153
5.11	Pola taburan harian aras kepekatan ozon maksimum bagi tahun 2000 dan 2005.	156

No. Rajah

Halaman

5.12	Pola taburan bulanan aras kepekatan ozon maksimum bagi tahun 2000 dan 2005.	157
5.13	Pemetaan statistik ikhtisar taburan aras kepekatan ozon maksimum harian.	160
5.14	Pola taburan aras kepekatan maksimum bulanan bagi bulan Januari hingga Disember.	162
6.1	Plot siri masa aras kepekatan ozon maksimum harian di 12 buah stesen kajian.	159
6.2	Plot Fungsi Autokorelasi, ACF aras kepekatan ozon maksimum harian di 12 buah stesen kajian.	160
6.3	Plot log-log $F(n)$ lawan n .	161
6.4	Taburan anggaran pekali Hurst menggunakan pendekatan serakan nisbah berskala (NSB).	172
6.5	Taburan anggaran pekali Hurst menggunakan pendekatan nyah-trend bergerak (NTB).	173
6.6	Pemetaan taburan panjang-ingatan berdasarkan Indeks Hurst Terlaras bagi kaedah nisbah serakan berskala (NSB).	176
6.7	Pemetaan taburan panjang-ingatan berdasarkan Indeks Hurst Terlaras bagi kaedah nyah-trend bergerak (NTB).	177
6.8	Plot log-log bagi punca kuasa dua aras kepekatan ozon maksimum harian $F(n)$ melawan selang masa, n (hari) bagi stesen-stesen di kawasan bandar.	180
6.9	Plot log-log bagi punca kuasa dua aras kepekatan ozon maksimum harian, $F(n)$ melawan selang masa, n (hari) bagi stesen-stesen di kawasan perindustrian.	181
6.10	Plot log-log bagi punca kuasa dua aras kepekatan ozon maksimum harian, $F(n)$ melawan selang masa, n (hari) bagi stesen-stesen di kawasan pinggir bandar.	183

SENARAI SIMBOL

$S(f)$	fungsi ketumpatan spektrum
f	frekuensi
Y_t	siri masa stokastik
γ_k	autokovarian dengan k susulan
ρ_k	autokorelasi dengan k susulan
B_H	model pergerakan Brownian pecahan
G_H	model hingar Gaussian pecahan
$=_d$	kebersamaan dari segi taburan
N	jumlah cerapan
n	bilangan cerapan
x_j	sampel siri masa dengan cerapan ke- j
$x_{n(i-1)+j}$	cerapan ke- j dalam selang ke- i bersaiz n
y_k	siri masa terintegrasi dengan k cerapan
$y_{n(i-1)+k}$	cerapan ke- k dalam selang ke- i bersaiz n
$\hat{y}_{n(i-1)+k, T}$	anggaran cerapan ke- k dalam selang ke- i bersaiz n yang diperoleh menggunakan teknik T
$F(n)$	statistik serakan
$F_Q(n)$	statistik serakan menggunakan ukuran julat pertengahan kuartil
$F_M(n)$	statistik serakan menggunakan ukuran median sisihan mutlak
Q_1	kuartil pertama
Q_3	kuartil ketiga
H	pekali Hurst/ nilai sebenar
$\hat{H}$	nilai anggaran pekali Hurst
$\bar{\hat{H}}$	purata nilai anggaran pekali Hurst
$\bar{\hat{H}}_w$	purata nilai anggaran pekali Hurst berwajaran

χ_{H_j}	indeks Hurst terlaras bagi setiap stesen j
V_{O_3}	pekali variasi aras kepekatan ozon maksimum harian
BS_H	nilai pincang bagi nilai anggaran pekali Hurst
S_H^2	varians bagi nilai anggaran pekali Hurst
$MRKD_H$	nilai min ralat kuasa dua bagi anggaran pekali Hurst
D_c	pengangka bagi siri masa kumulatif
D	penyebut bagi siri masa asal
R_c	julat bagi siri masa kumulatif
V_c	varians bagi siri masa kumulatif
S_c	sisihan piawai bagi siri masa kumulatif
Q_c	julat pertengahan kuartal bagi siri masa kumulatif
M_c	median sisihan mutlak bagi siri masa kumulatif
ρ_c	pekali konkordans Lin
n_{c_1}	skala masa kritikal pertama
n_{c_2}	skala masa kritikal kedua
B	kawasan bandar
I	kawasan perindustrian
PB	kawasan pinggir bandar

SENARAI SINGKATAN

ASMA	Alam Sekitar Sendirian Berhad
R/S	Julat Penskalaan Semula
V/S	Varians Penskalaan Semula
SWV	Varians Tertingkap Berskala
DFA	Nyah-trend Turun-naik
DMA	Nyah-trend Purata Bergerak
DMM	Nyah-trend Median Bergerak
SD	Sisihan Piawai
MED	Median Sisihan Mutlak
IQR	Julat Antara Kuartil
NN	Kejiranan Terdekat
SKNN	K -Kejiranan Terdekat Berjujukan
NSB	Nisbah Serakan Berskala
NTB	Nyah-trend Bergerak
IHT	Indeks Hurst Terlaras

BAB I

PENDAHULUAN

1.1 MOTIVASI KAJIAN

Siri masa panjang-ingatan adalah satu siri masa yang berbeza dengan siri masa pendek-ingatan. Pada dasarnya, perbezaan di antara kedua-dua siri masa ini dilihat kepada sifat autokorelasi masing-masing. Siri masa pendek-ingatan adalah suatu siri tak bersandar atau bersandar untuk suatu susulan masa terdekat, manakala siri masa panjang-ingatan adalah suatu siri masa yang saling bersandar untuk suatu susulan masa yang panjang. Sifat ketakbersandaran dalam siri masa ini lebih dikenali sebagai autokorelasi. Autokorelasi merujuk kepada hubungan di antara cerapan lepas dengan cerapan masa depan mengikut susulan masa atau lat. Pada kebiasaannya, nilai autokorelasi menyusut secara eksponen ke sifar dan ini hanya berlaku jika suatu siri masa itu pendek-ingatan. Namun dalam amalan sebenar, perkara sedemikian sukar diperoleh dan lazimnya, autokorelasi ini menjadi satu permasalahan utama dalam analisis siri masa. Autokorelasi ini boleh berlaku sama ada autokorelasi positif atau negatif. Autokorelasi positif boleh dianggap satu keadaan tertentu yang dikenali sebagai "keberterusan", yakni kecenderungan sesuatu sistem itu kekal berada pada keadaan yang sama sebelum ia berubah ke keadaan yang lain. Nilai cerapan kelihatan terus meningkat untuk suatu tempoh yang lama sebelum menyusut ke paras yang baharu. Penyusutan juga berlaku untuk suatu tempoh yang agak lama sebelum meningkat semula. Manakala, autokorelasi negatif pula dikenali sebagai "anti-keberterusan" yang memperlihatkan keadaan dengan peningkatan nilai cerapan yang

berlaku sebelumnya akan diikuti penyusutan selepasnya, begitu juga sebaliknya. Keadaan sebegini di kenali sebagai fenomena panjang-ingatan.

Menurut Zhu dan Liu (2003), suatu siri masa itu dikatakan suatu siri masa panjang-ingatan sekiranya nilai-nilai bersebelahannya berkorelasi sama ada berkorelasi positif atau negatif. Ingatan di sini merujuk kepada ingatan dalaman yang saling berkait antara cerapan-cerapan sebelum dan cerapan-cerapan selepasnya dan ia berlaku dalam suatu tempoh yang lama. Secara intuitif, kehadiran panjang-ingatan di dalam data siri masa mengimplikasikan bahawa terdapat korelasi yang sangat kuat antara setiap titik data berikutan suatu peristiwa dipengaruhi oleh peristiwa-peristiwa sebelumnya. Menurut Beran (1994), ciri-ciri statistik suatu siri masa panjang-ingatan dengan siri masa pendek-ingatan adalah berbeza. Jika siri masa pendek-ingatan dengan varians terhingga, sampel minnya adalah asimptotik normal dan varians sampelnya berkadar dengan $1/n$, dengan n adalah saiz sampel. Jumlah fungsi autokorelasinya juga adalah terhingga dan bersamaan dengan sifar. Maka, keputusan pentaabiran statistik seperti selang keyakinan dan ujian hipotesis bagi min populasi boleh diperoleh dengan mudah. Namun, sesetengah cerapan siri masa walaupun kelihatan pegun, didapati hukum sampel besar atau teorem had memusat gagal dipenuhi apabila varians sampelnya menghampiri sifar dengan sangat perlahan menjauhi $1/n$. Yakni, jumlah fungsi autokorelasinya adalah tak terhingga. Maka, jika pentaabiran statistik dibuat berdasarkan keputusan ini, ianya akan memberikan keputusan yang tidak bererti ke atas min populasi. Demikian juga, kaedah formal analisis siri masa yang terkenal iaitu kaedah Box dan Jenkins (ARIMA), tidak lagi sesuai digunakan bagi siri masa panjang-ingatan (Beran 1994; Zhu & Liu 2003). Kaedah Box dan Jenkins ini popular mungkin disebabkan ia suatu model yang fleksibel yang boleh mengendalikan apa-apa siri, pegun atau tidak, bermusim atau tidak bermusim dan mudah diperoleh dalam pelbagai perisian statistik. Maklumat lanjut berkenaan model Box dan Jenkins boleh diperoleh dalam Bowerman dan O'Connell (2005), Cowpertwait dan Metcalfe (2008), Box et al.(1994) dan Brockwell dan Davis (1991).

Kehadiran panjang-ingatan di data siri masa memberikan model ARIMA(p,d,q) bukanlah satu model yang terbaik. Ini kerana, apabila varians sampel menyusut dengan sangat perlahan menjauhi $1/n$, bilangan parameter yang banyak di dalam model ARIMA(p,d,q) diperlukan dan parameter ini akan menjadi tak terhingga bila mana berlakunya pertambahan saiz sampel. Penggunaan terlalu banyak parameter dalam sesuatu kajian adalah tidak perlu terutamanya apabila ianya meningkatkan ketidakpastian pentaabiran statistik, dan kesukaran untuk mentafsir parameter. Oleh itu, sekiranya terdapat tanda-tanda penyusutan varians secara perlahan, maka permodelan dilakukan secara eksplisit. Pendekatan yang paling mudah yang boleh digunakan ialah penyusutan varians sampel berkadaran terhadap $n^{-\alpha}$ untuk suatu $\alpha \in (0,1)$ yang dikenali sebagai model panjang-ingatan (Beran 1994). Antara rujukan yang boleh dilihat berkenaan model panjang-ingatan adalah Brockwell (2004), Robinson (2003), Palma (2007) dan Peter (1994).

Selain daripada sifat autokorelasi, kehadiran panjang-ingatan bagi suatu siri masa juga dilihat berdasarkan domain frekuensinya. Kehadiran panjang-ingatan memberikan fungsi ketumpatan spektrum, $S(f)$ menjadi tidak terbatas apabila frekuensi, f menghampiri sifar. Keadaan ini berbeza bagi siri masa pendek-ingatan dengan memperlihatkan keadaan fungsi ketumpatan spektrum, $S(f)$ adalah bersamaan sifar apabila frekuensi, f menghampiri sifar. Walau bagaimanapun, perbincangan panjang-ingatan suatu siri masa dalam tesis ini hanya tertumpu pada domain masa iaitu berdasarkan sifat autokorelasinya.

Siri masa panjang-ingatan ini, kerap dijumpai pada data siri masa seperti limpahan air sungai, harga pasaran saham dan rangkaian internet. Ianya telah lama mendapat perhatian para penyelidik apabila sekitar tahun 50an, Hurst seorang ahli hidrologis, dalam kajiannya terhadap limpahan air sungai Nil membuktikan secara empirik bahawa paras air sungai setiap tahun menunjukkan ciri-ciri panjang-ingatan dengan paras air sungai meningkat dalam tempoh yang lama sebelum ianya kembali surut. Juga, beliau dapati, paras air sungai Nil tidak dapat dimodelkan dengan model klasik ARMA. Kajian beliau telah membuahkan hasil sumbangan yang besar dalam usaha merangka satu strategi kawalan banjir. Beliau mempersembahkan satu rekaan

empangan yang mampu menampung air semaksimum mungkin, juga dengan jaminan kekeringan air dalam empangan tidak akan berlaku (Hurst et al. 1965). Selain itu, sumbangan besar beliau juga adalah dengan memperkenalkan pekali Hurst, H yang terbukti menjadi penunjuk kepada kehadiran panjang-ingatan bagi suatu data siri masa. Bagi para pelabur saham pula, ciri-ciri panjang-ingatan yang dilihat dalam siri masa harga pasaran saham telah membawa kepada penyediaan langkah perancangan strategik dalam sesuatu pelaburan. Pemantauan tempoh pulangan modal yang mengimplikasikan penolakan hipotesis teori pasaran saham telah membawa kepada pengaplikasian model panjang-ingatan berikutan telah terbukti harga pasaran saham gagal dimodelkan secara model klasik ARMA (Peter 1994; Skjeltorp 2000; Couillard & Davison 2005; Erfani & Samimi 2009; Serinaldi 2010). Rangkaian lalu lintas internet yang mempamerkan ciri-ciri panjang-ingatan dikatakan telah menjejaskan banyak aspek dari segi reka bentuk rangkaian dan reka bentuk komponen rangkaian yang biasanya bertindak sebagai tulang belakang dalam pembentukan keseluruhan rangkaian. Justeru, memahami sifat rangkaian lalu lintas internet berkelajuan tinggi serta sistem komunikasi jalur lebar adalah penting untuk reka bentuk, operasi dan penilaian prestasi rangkaiannya. Implikasi panjang-ingatan dalam kejuruteraan internet adalah dilihat pada asasnya sebagai satu model sumber yang memberi maklumat bila mana individu pengguna internet dijangka menunjukkan kepelbagaian yang melampau dalam tempoh antara ketibaan data paket. Manakala, pekali Hurst, H dilihat sebagai satu tanda aras yang bermakna untuk melihat data tersekat-sekat berbanding ukuran indeks penyebaran, nisbah min puncak dan pekali variasi bagi masa antara ketibaan data paket. Juga, sifat kesesakan yang dihasilkan oleh model lalu lintas rangkaian internet panjang-ingatan dilihat secara drastiknya berbeza daripada apa yang diramalkan oleh model biasa (Shah et al. 2000; Gao & Rubin 2001; Park et al. 2005; Kherani & Kumar 2005; Clegg et al. 2010).

Kebanyakan kajian yang dijalankan mengaplikasikan kaedah yang dicadangkan oleh Hurst et al. (1965) iaitu kaedah Julat penskalaan Semula, R/S yang mencirikan sifat panjang-ingatan dengan pekali Hurst, H . Terdapat juga pendekatan lain yang digunakan, namun pendekatan yang paling mendapat perhatian adalah melalui analisis penskalaan seperti yang dilakukan oleh Hurst. Antara kaedah-kaedah

analisis penskalaan selain R/S yang mendapat perhatian para penyelidik adalah kaedah Varians Penskalaan Semula, V/S yang diperkenalkan oleh Cajueiro dan Tabak (2009), kaedah Nyah-trend Turun-naik, DFA diperkenalkan oleh Peng et al. (1995) dan kaedah Nyah-trend Purata bergerak, DMA yang diperkenalkan oleh Carbone et al. (2004a). Kaedah-kaedah ini memperlihatkan dua teknik berbeza, iaitu kaedah R/S dan V/S mempertimbangkan ujian statistik secara pemiawaian antara dua ukuran serakan tanpa mengambil kira kehadiran trend, yakni trend dinyahkan terlebih dahulu sebelum analisis penskalaan dilakukan. Manakala, kaedah DFA dan DMA pula mempertimbangkan trend yang ada dibuang seiring dengan perlaksanaan analisis penskalaan dan ujian statistik yang digunakan adalah ukuran serakan sisihan piawai sebagai ujian penentuan kehadiran panjang-ingatan. Namun, kesemua kaedah ini berkongsi matlamat yang sama iaitu mengesan ketaktentuan atau turun-naik data pada skala masa yang berbeza dengan masing-masing mempunyai kelebihan dan kelemahan tersendiri.

Selain itu, kajian berkenaan panjang-ingatan ini juga mendapat perhatian penyelidik dalam usaha memulihara kelestarian alam sekitar. Kaedah-kaedah popular ini, juga beberapa pendekatan lain telah digunakan pada pemboleh ubah alam sekitar seperti siri pemendapan berasid, suhu, kelajuan angin dan agen pencemaran udara iaitu aras kepekatan ozon (O_3), nitrogen dioksida (NO_2), sulfur dioksida (SO_2) dan zarah terampai bersaiz diameter kurang daripada 10 mikrometer (PM_{10}). Kai et al. (2008) menjalankan kajian terhadap indeks pencemaran udara (IPU) harian dan tiga indeks pencemaran iaitu sulfur dioksida (SO_2), nitrogen dioksida (NO_2) dan zarah terampai (PM_{10}) di Shanghai, China. Hasil pengaplikasian kaedah R/S, DFA dan spektrum telah menunjukkan bahawa siri masa bagi bahan pencemar yang berbeza mempamerkan sifat dinamik yang berbeza untuk penskalaan jangka masa panjang. Weng et al. (2008) pula telah menjalankan kajian di dua kawasan bandar di Kaohsiung, Taiwan untuk melihat struktur yang mendasari ozon di permukaan dan sekali gus meningkatkan kapasiti peramalan di kawasan bandar. Hasil analisis menggunakan kaedah R/S menunjukkan bahawa, siri masa kepekatan ozon adalah berterusan dengan nilai pekali Hurst, H yang diperoleh menghampiri 0.75. Keputusan yang diperoleh ini serupa dengan hasil kajian yang dilakukan oleh Windsor