

*MEMBINA KAITAN TAKLINEAR ANTARA KEPEKATAN
ZARAHAN TERAMPAI DENGAN VARIAT-VARIAT
METEOROLOGI*

*Sabariah Baharun
Jabatan Matematik
Mohd. Rashid Mohd. Yusof
Jabatan Kejuruteraan Kimia
Fakulti Kej. Kimia & Kej. Sumber Asli
Universiti Teknologi Malaysia
54100 KUALA LUMPUR*

*Yunus Jaafar
Unit Statistik
MARDI, Serdang,
50774 KUALA LUMPUR*

Dibentangkan di:-

MARDI

14hb. Julai, 1990

MEMBINA KAITAN TAKLINEAR ANTARA KEPEKATAN ZARAHAN TERAMPAI DENGAN VARIAT-VARIAT METEOROLOGI

Sabariah Baharun,
Jabatan Matematik,
Mohd. Rashid Mohd. Yusof,
Jabatan Kejuruteraan Kimia,
Universiti Teknologi Malaysia
54100 KUALA LUMPUR.

Yunus Jaafar
Unit Statistik
MARDI, Serdang,
50774 KUALA LUMPUR.

ABSTRAK

Beberapa kaedah permodelan dengan menggunakan pakej SAS telah dikaji untuk menghasilkan bentuk model yang paling sesuai bagi menerangkan kaitan antara kepekatan zarah terampai (pembolehubah ramalan) TSP (Total Suspended Particulate) dengan beberapa faktor meteorologi sebagai pembolehubah penerang. Kaedah permodelan yang telah digunakan termasuklah permodelan regresi berganda, kaedah regresi mundur langkah demi langkah dan model taklinear. Keputusan menunjukkan bahawa kaitan taklinear merupakan bentuk permodelan yang paling sesuai menerangkan gejala pencemaran ini.

PENGENALAN

Memang sudah menjadi lumrah bagi penyelidik untuk membentuk atau menghasilkan model yang sesuai bagi bahan yang dikajinya. Tidak ada suatu model pun yang boleh dikatakan memberi penerangan yang jitu terhadap sesuatu perlakuan. Oleh itu setiap model diperolehi perlu diperbaiki dari masa ke semasa bergantung kepada data-data yang dipungut. Hampir semua bentuk fungsi secara teorinya menunjukkan perkataan yang taklinear keadaannya.(1) Kerumitan dalam pembentukan model taklinear tersebut telah menyebabkan kebanyakan kajian sekarang ini tertumpu kepada bentuk linear. Dengan perkembangan pesat alat menghitug seperti komputer telah dapat mempermudah permodelan taklinear ini diramalkan.

Sehubungan dengan ini, satu kajian untuk meninjau kaedah permodelan paling sesuai bagi menerangkan hubungan antara kepekatan zarah terampai di udara Kuala Lumpur dengan faktor-faktor meteorologi (halaju angin, jangkamasa hujan, jumlah hujan, suhu, cahaya matahari, pancaran suria dan kelembapan) telah dilakukan dan dibincangkan dalam kertas kerja ini.

METODOLOGI

Pengumpulan data.

- Lokasi kajian.

Data zarah terampai, TSP diperolehi dari Stesen Pengawasan Kualiti Udara Universiti Teknologi Malaysia, Kuala Lumpur. Program pengawasan tersebut bermula pada bulan Januari 1986 hingga ke hari ini. Lokasi stesen serta kedudukan kawasan perindustrian diberikan dalam Rajah 1. Ianya berada 2 km. ke timur laut Bandaraya Kuala Lumpur dan terletak lebih kurang 10 km daripada kawasan perindustrian Petaling Jaya pada arah yang sama.

Penyampelan

Pengumpulan data harian (24 jam) TSP telah dilakukan dengan menggunakan alat High Volume Air Sampler yang mampu mengumpul sebarang saiz partikel sehingga sebesar 50 mikron. Alat tersebut ditempatkan di atas bumbung sebuah bangunan setinggi empat tingkat. Perihal detil tentang peralatan dan prosidur penyampelan telah dinyatakan di tempat lain. Faktor meteorologi yang dilaporkan dalam kajian ini telah diambil dari stesen jabatan kajicuaca yang paling hampir dengan stesen pengawasan kualiti udara, UTM.

Analisis Data

Pakej SAS telah digunakan dalam menganalisis data dalam kajian ini. Langkah pertama dalam analisis ini ialah pengamatan terhadap matriks korelasi bagi keseluruhan pembolehubah untuk meninjau pembolehubah penerang yang mempunyai koefisien korelasi yang bererti terhadap pembolehubah respon (TSP). Regresi berganda dilakukan untuk melihat kaitan linear keseluruhan pembolehubah penerang dan sumbangannya terhadap TSP.

Seterusnya prosidur regresi undur langkah demi langkah digunakan untuk menentukan pembolehubah yang memberikan sumbangan bererti terhadap TSP.(3)

Penumpuan seterusnya adalah terhadap permodelan taklinear.

Model Taklinear

Secara matematik, kita boleh nyatakan sesuatu perkara penting dalam sains dengan penerangan bagi fenomena yang ditinjau dalam persamaan berbentuk;

$$y = f(x, \theta) + \epsilon$$

yang y adalah satu atau lebih nilai ukuran bagi respon (ii. pembolehubah bersandar atau hasil bagi sesuatu sistem yang dikaji), ϵ adalah ralat kajian dan $f(x, \theta)$ pula adalah ungkapan matematik mengandungi p parameter $\theta_1, \theta_2, \dots, \theta_p$ (dinyatakan θ) dan k pembolehubah (x_1, x_2, \dots, x_k (dinyatakan x)).

Sebagai contoh jika y suatu respon tunggal, fungsi f boleh mengambil bentuk;

$$f(x, \theta) = \theta_1 (1 - \exp(-\theta_2 x))$$

yang x boleh merupakan sebarang pembolehubah yang mempunyai kaitan dengan y .

Model-model taklinear yang sudah diterokai dalam statistik adalah seperti berikut:

1. Model Michaelis - Menten

$$f(x, \theta) = \frac{\theta_1 x}{\theta_2 + x}$$

2. $f(x, \theta) = \theta_1 (1 - e^{-\theta_2 x})$

3. $f(x, \theta) = \theta_1 x^{\theta_2}$

4. Model Regresi Asimtotik

$$f(x, \theta) = \theta_1 + \theta_2 e^{\theta_2 x}$$

5. Model Logistik (model taklinear bagi pertumbuhan tanaman dan organisme)

$$f(x, \theta) = \frac{\theta_1}{1 + \theta_2 e^{\theta_3 x}}$$

6. Model pertumbuhan Gompertz.

$$f(x, \theta) = \theta_1 e^{-e^{-(\theta_2 - \theta_3 x)}}$$

$$7. \quad f(x, \theta) = \theta_1 (1 - \theta_2 e^{-\theta_3 x})$$

8. Model pertumbuhan log logistik

$$f(x, \theta) = \theta_1 - \ln(1 + \theta_2 e^{-\theta_3 x})$$

$$9. \quad f(x, \theta) = \theta_1 + \frac{\theta_2}{x^{\theta_3}}$$

10. Model pertumbuhan Morgan, Mercer, Florin

$$f(x, \theta) = \frac{\theta_2 \theta_3 + \theta_1 x^{\theta_4}}{\theta_3 + x^{\theta_4}}$$

11. Model pertumbuhan Richards

$$f(x, \theta) = \frac{\theta_1}{(1 + \theta_2 e^{-\theta_3 x})^{1/\theta_4}}$$

12. Model Duncan

$$f(x, \theta) = \frac{\theta_1}{\theta_1 - \theta_2 (e^{-\theta_2 x} - e^{-\theta_1 x})}$$

$$13. \quad f(x, \theta) = e^{\theta_1 + \theta_2 x}$$

Model 3 telah digunakan dalam kajian ini.

TAKRIF KETAKLINEARITIAN

Perbezaan penting adalah antara model yang linear dalam parameter dan model yang tidak. Model linear boleh ditulis sebagai;

$$f(x, \theta) = \sum_{i=1}^p \theta_i g_i(x)$$

bagi fungsi g_i yang hanya bergantung kepada nilai x dan tidak pada nilai θ . Sebagai contoh model linear adalah:

$$f(x, \theta) = \theta_1 + \theta_2 x$$

$$f(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2$$

Selain model yang boleh ditulis seperti dalam bentuk di atas kebanyakan model matematik adalah yang menyajikan gejala tabii (natural models) dengan cocok taklinear dalam parameter atau dinyatakan model taklinear.

Cara yang mudah bagi menentukan sesuatu model itu linear atau taklinear adalah dengan memeriksa pembeza (terbitan) bagi f terhadap setiap parameter θ_i . Jika $\delta f / \delta \theta_i$ tidak bergantung kepada sebarang θ , model tersebut adalah linear dalam θ_i dan jika f linear dalam semua p parameter ($\theta_1, \theta_2, \dots, \theta_p$) maka model dinyatakan sebagai model linear.

Sebahagian model tak linear boleh ditukarkan ke bentuk linear dengan mengambil Logaritma atau songsangannya atau dengan sebarang transformasi.(5) Tidak semua model dengan mudah boleh ditukar ke bentuk linear.

Model tak linear boleh dianalisis dengan menggunakan regresi taklinear. Seperti regresi linear, anggaran parameter adalah dengan mengambil nilai θ , yang meminimumkan hasil tambah sisihan kuasa dua.

Berbanding dengan regresi linear, yang mudah dianggarkan parameternya, regresi taklinear memerlukan kaedah terlelar (iterative) yang mudah dihasilkan dengan memakai program pakej SAS yang menerapkan komputer kerangkautama. Kaedah-kaedah yang boleh digunakan dalam menganggarkan parameter adalah;

1. Gauss-Newton
2. Marquardt
3. Kecerunan
4. DUD (kaedah yang tidak menggunakan pembeza)

Dalam kajian ini, kaedah menganggar parameter dengan anggaran kuasa dua terkecil menggunakan kaedah Marquardt telah digunakan. Untuk mengetahui lebih lanjut tentang kaedah tersebut di atas boleh merujuk di tempat lain. (4,5,7,8).

Langkah-langkah untuk mencapai pembentukan model tidak linear adalah seperti berikut:-

- 1) memplot data mentah
- 2) menganggar nilai permulaan
- 3) menganggar secara kasar model yang hampir dengan arah aliran plot data mentah
- 4) menggunakan pakej SAS untuk melelar nilai hampiran parameter dalam model
- 5) mengadakan analisis sisihan untuk menentukan sama ada terdapat varians pemalar ralat
- 6) menguji sumbangan parameter dalam model dengan menerapkan ujian t terStudent (studentized t test)
- 7) menerapkan matrik korelasi parameter untuk menguji model itu supaya tidak berlebihan parameter dalam model/peraga

KEPUTUSAN DAN PERBINCANGAN

Taburan purata bulanan bagi TSP dalam masa empat tahun ditunjukkan dalam Rajah 2. Daripada graf didapati ada beberapa bulan yang tiada mempunyai nilai TSP yang dipungut. Secara keseluruhan, didapati ada variasi berlaku dalam kepekatan TSP dari masa ke semasa dalam mendalar (domain) vektor percubaan.

Kaitan antara variat dalam data percubaan (Jadual I) menunjukkan bahawa jangkamasa hujan dan suhu mempunyai korelasi yang bererti dengan TSP, masing-masing pada aras bererti rantau keberangkalan 5% dan 1%.

Dalam model lengkap keseluruhan ini (Jadual 2) variat-variati X_2 , X_4 dan X_5 menyumbang bererti model dengan nilai-nilai $t_3 = -2.189$ ($Pr > T = 0.03$), $t_4 = +3.384$ ($Pr > T = 0.0009$) dan $t_5 = -2.84$ ($Pr > T = 0.005$). Keputusan ini disokong oleh kajian seterusnya dengan menggunakan regresi undur langkah demi langkah.

Versi undur dalam regresi langkah demi langkah bermula dengan model lengkap persamaan regresi dan membuang variat yang tidak mustahak satu demi satu, pada setiap langkah. Daripada kaedah ini diperolehi tiga variat penerang iaitu jangkamasa hujan, suhu dan pancaran suria menyumbang bererti kepada respon. Pemilihan ini dikuatkan dengan nilai statistik Mallows (Jadual 3) yang terkecil. Kemungkinan ada anggapan bahawa dengan menggunakan regresi undur langkah demi langkah ini, satu model paling cocok bagi data telah dihasilkan. Namun begitu variat-variat ini bukan merupakan variat penerang yang nisbah variasinya bagus. Ini boleh dilihat dari nilai R^2 yang cuma 13.7% yang bermakna gabungan variat ini hanya menerangkan 13.7% sahaja kepada model ini. Dengan ini satu alternatif yang diambil dalam kajian ini ialah menerokai model taklinear. Secara rambang variat suhu telah digunakan dalam menghasilkan model taklinear. Taburan TSP terhadap suhu digambarkan dalam Rajah 3. Hasilnya, model berikut:

$$y = 0.6153 x^{1.43}$$

telah diperolehi (Jadual 4).

Daripada analisis sisihan (Rajah 4) bagi menguji model tersebut didapati bahawa varians ralatnya sekata dan ini menunjukkan model tersebut adalah sesuai bagi menerangkan kaitan antara TSP dan suhu.

Dengan ujian t-terStudent (Studentized t-test) didapati bahawa nilai anggaran parameter adalah sangat bererti dalam sumbangannya terhadap model. Namun begitu dengan menganalisis matriks korelasi kedua anggaran parameter tersebut menunjukkan bahawa keduanya mempunyai koefisien korelasi yang sungguh bererti pada aras bererti rantau keberangkalian 1%. Ini bermakna bahawa model ini mempunyai lebih parameter dan seterusnya model ini mungkin boleh diperbaiki lagi dengan hanya mengambil satu sahaja parameter.

RUMUSAN

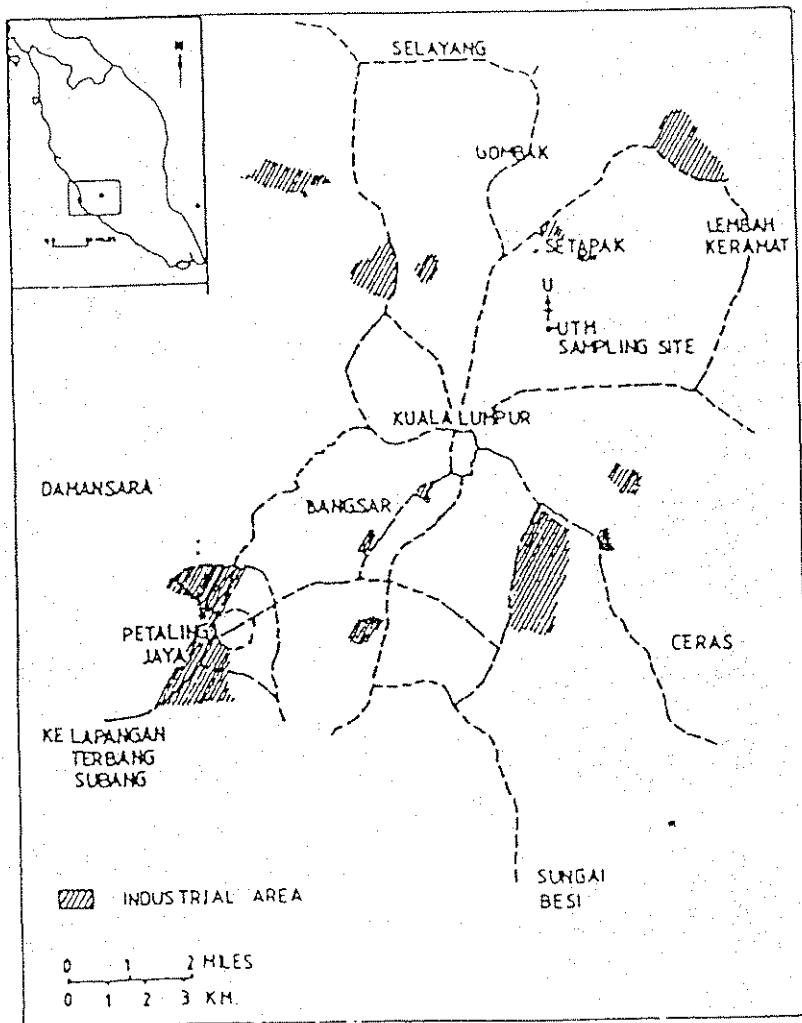
Pelbagai model regresi telah dibandingkan. Pemilihan model yang sesuai untuk mewakili data bergantung kepada objektif dan keadaan data itu sendiri. Pada hakikatnya dalam kajian ini, pengutipan data TSP dan data variat meteorologi telah dibuat dalam dua lokasi yang berasingan. Ini mengakibatkan kesan variat penerang terhadap variat respon (TSP) tidak boleh diungkapkan dalam model yang lengkap kebagusan suainya). Walaupun terdapat 13 model yang telah diterokai namun hanya satu model sahaja dapat diuji. Walau bagaimanapun usaha ke arah permodelan tak linear yang lebih mantap akan diteruskan.

PENGHARGAAN

Setinggi-tinggi terima kasih kepada Pengarah Bahagian Kajian Tekno-Ekonomi dan Sosial, Tuan Haji Sameon Abdullah dan Ketua Unit Metodologi Statistik, Dr. Lee Chong Soon, yang telah memberikan kerjasama dan galakan dan seterusnya kepada pihak MARDI umumnya yang telah membenarkan pengarang dalam menggunakan khidmat dan kepakaran yang ada dalam menjayakan penulisan ini.

RUJUKAN

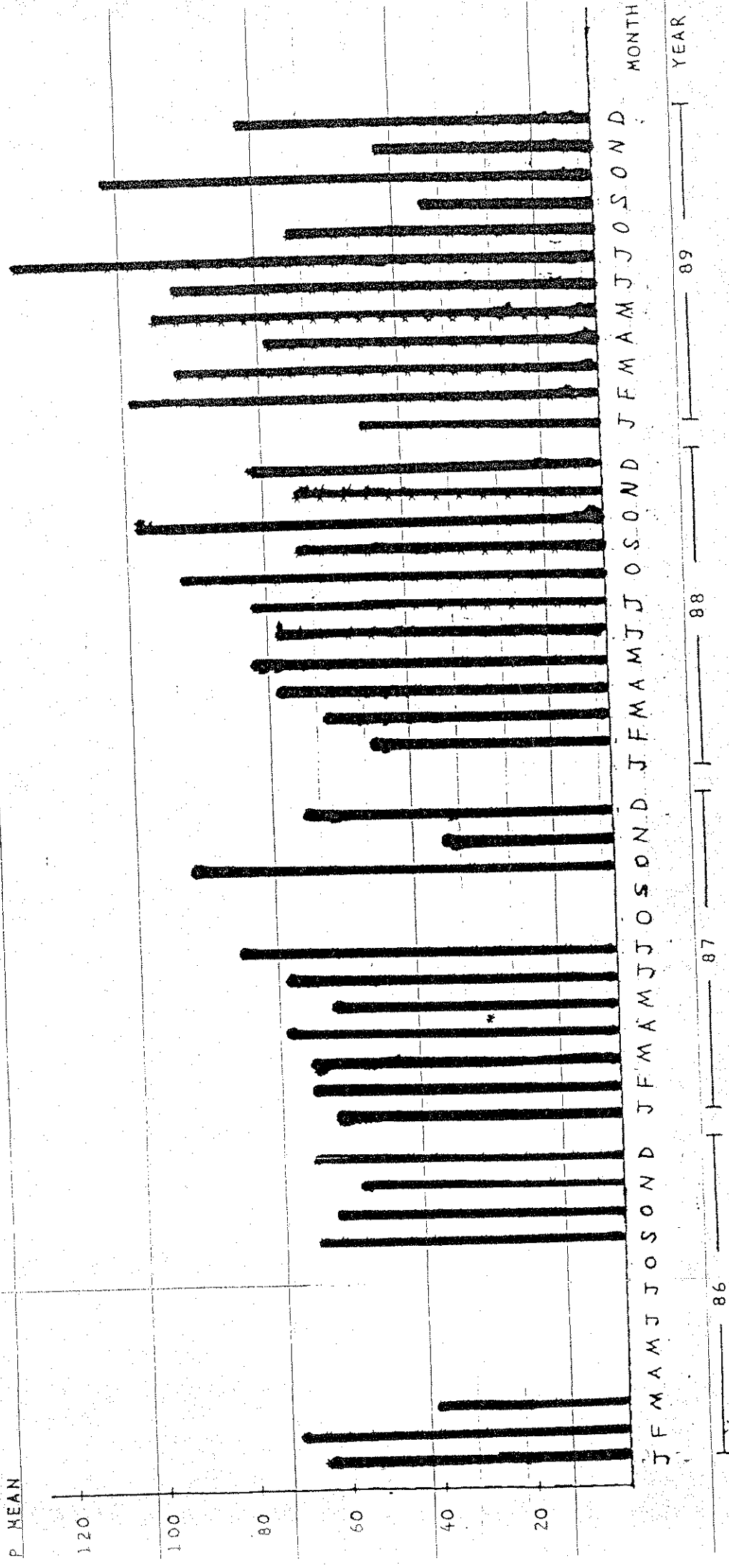
1. G.E. Box dan W.G. Hunter (1962) 'A Useful Method for Model-Building'. *Technometrics*, Vol.4 (301-318).
2. Mohd. Rashid M.Y. (1987) 'An Investigation of PM10 Concentrations at One Site of Kuala Lumpur' Kertas kerja yang dibentangkan pada Simposium Ketiga Jurutera Kimia Malaysia, Kuala Lumpur.
3. J.W. Gorman dan R.J. Toman (1966) 'Selection of Variables for Fitting Equations to Data'. *Technometrics*. Vol.8 (27-51).
4. N. Draper dan H. Smith (1981) 'Applied Regression Analysis', 2nd. Edition John Wiley & Sons, Inc.
5. Snadecor, G.W. dan Cochran, W.G. (1990) 'Statistical Methods,' Seventh Edition, Ames. Iowa : The Iowa State University Press.
6. Kotz, S, Johnson, N dan Read, C. (1985) 'Encyclopaedia of Statistical Sciences' Vo.6, John Wiley.
7. Weisberg, S (1985). 'Applied Linear Regression', 2nd. Ed. New York : John Wiley & Sons.
8. SAS User's Guide : Statistics. Version 5 Edition, SAS Institute Inc. Cary North Carolina.
9. Chatfield, C. and Collins, A.J. (1980) Introduction to Multivariate Analysis, London : Chapman and Hall



Rajah 1: Lokasi Stesen Kualiti Udara (UTM) dan Kedudukan Industri

REGISTRATION

BAR CHART OF MEANS



Rajah 2

JADUAL 1: Matrik Korelasi

	TSP	WSPEED	AVSPEED	DURRAIN	AMTRAIN	TEMP	SOLRAD	SUNSHINE	RELHUMID
TSP	1.00000	-0.08429	-0.11995	-0.23038	-0.13427	0.17699	-0.03790	-0.00296	-0.06493
	0.0000	0.2463	0.1215	0.0013	0.0773	0.0143	0.6115	0.9676	0.3722
	191	168	191	174	182	191	191	191	191
WSPEED	-0.08429	1.00000	0.71387	-0.02017	-0.10124	-0.00249	0.02110	0.00639	-0.07389
	0.2463	0.0000	0.0001	0.7569	0.1371	0.9695	0.7513	0.9219	0.2562
	191	238	201	238	217	238	228	238	238
AVSPEED	-0.11995	0.71387	1.00000	-0.08376	-0.26036	0.20183	0.21540	0.13232	-0.30007
	0.1215	0.0001	0.0000	0.2371	0.0003	0.0041	0.0028	0.0611	0.0001
	168	201	201	201	185	201	191	201	201
DURRAIN	-0.23038	-0.02017	-0.08376	1.00000	0.48480	-0.53992	-0.41369	-0.29007	0.52871
	0.0013	0.7569	0.2371	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001
	191	238	201	238	217	238	228	238	238
AMTRAIN	-0.13427	-0.10124	-0.26036	0.48480	1.00000	-0.36276	-0.22529	-0.14873	0.43661
	0.0773	0.1371	0.0003	0.0001	0.0000	0.0001	0.0011	0.0295	0.0001
	174	217	185	217	217	217	208	217	217
TEMP	0.17699	-0.00249	0.20183	-0.53992	-0.36276	1.00000	0.69328	0.45754	-0.71387
	0.0143	0.9695	0.0041	0.0001	0.0001	0.0000	0.0001	0.0001	0.0001
	191	238	201	238	217	238	228	238	238
SOLRAD	-0.03790	0.02110	0.21540	-0.41369	-0.22529	0.69328	1.00000	0.63110	-0.70285
	0.6115	0.7513	0.0028	0.0001	0.0011	0.0001	0.0000	0.0001	0.0001
	182	228	191	228	208	228	228	228	228
SUNSHINE	-0.00296	0.00639	0.13232	-0.29067	-0.14873	0.45764	0.63110	1.00000	-0.46202
	0.9676	0.9219	0.0611	0.0001	0.0285	0.0001	0.0001	0.0000	0.0001
	191	238	201	238	217	238	228	238	238
RELHUMID	-0.06493	-0.07389	-0.30007	0.52871	0.43661	-0.71387	-0.70285	-0.46202	1.00000
	0.3722	0.2562	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000
	191	238	201	238	217	238	228	238	238

Jadual 2: Model Linear Berganda

$$Y = -149.7 - 2.06X_1 - 0.04 X_2 \\ -0.08 X_3 + 8.24 X_4 - 0.07 X_5 \\ + 0.16 X_6 + 0.43 X_7$$

$$R^2 = 0.153$$

$$F_{5,158} = 4.067$$

$$= (\text{Pr} < 0.004)$$

Legenda:

- X_1 = halaju angin
- X_2 = jangkamasa hujan
- X_3 = jumlah hujan
- X_4 = suhu
- X_5 = pancaran suria
- X_6 = cahaya matahari
- X_7 = kelembapan

Jadual 3 : Ringkasan Nilai Statistik Dalam Permodelan Undur Langkah Demi Langkah

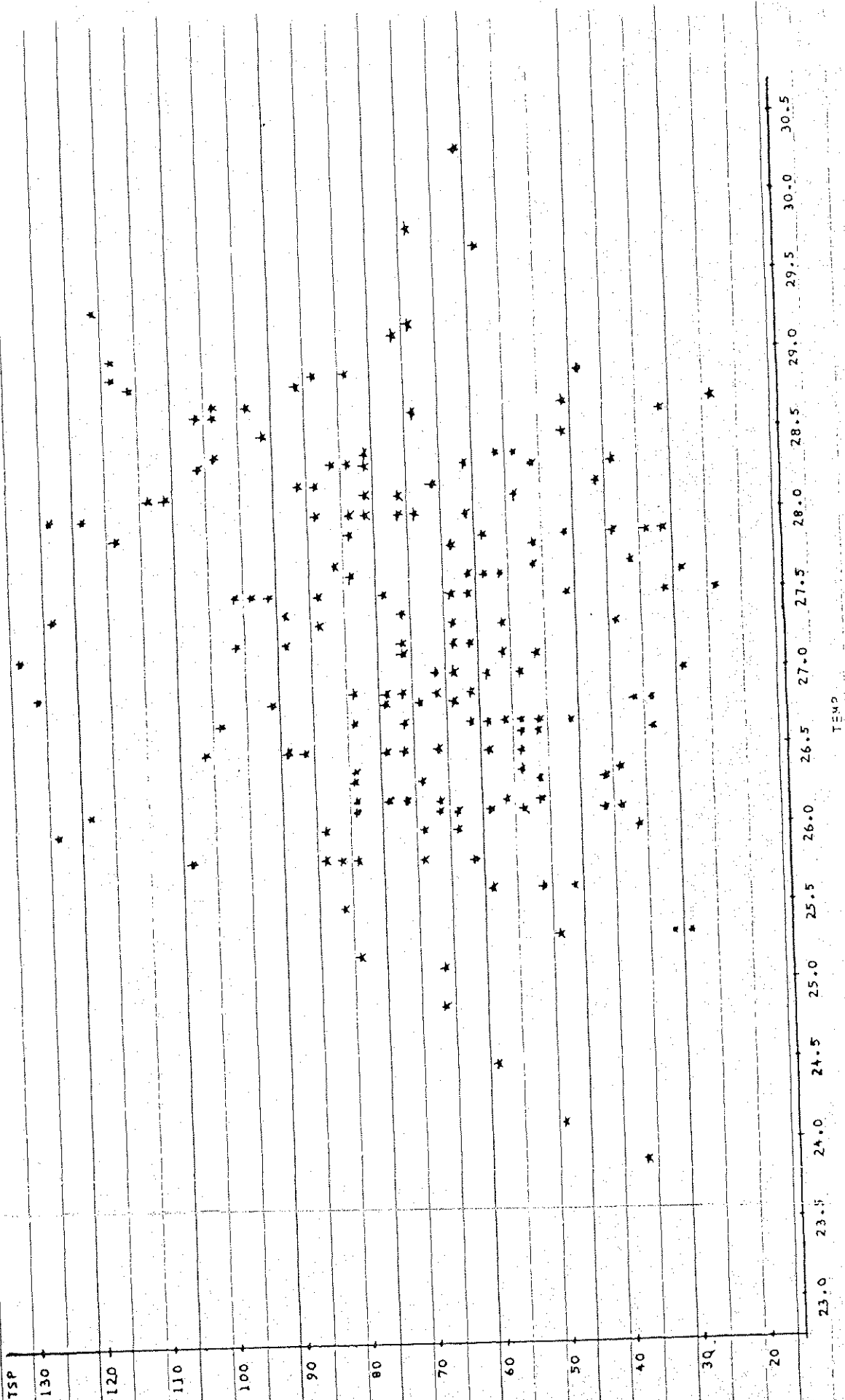
Variat	P	C _p	R ²
Keseluruhan variat	8	8	0.153
X ₁ ,X ₂ ,X ₃ ,X ₄ ,X ₅ ,X ₇	7	6.07	0.152
X ₁ ,X ₂ ,X ₄ ,X ₅ ,X ₇	6	4.38	0.151
X ₁ ,X ₂ ,X ₄ ,X ₅	5	3.22	0.146
X ₂ ,X ₄ ,X ₅	4	2.92	0.137

Legenda:

- P = bilangan sebutan dalam persamaan regresi linear yang dikaji
- C_p = Statistik Mallows
- R² = Koeffisien penentuan berganda

KEPEKATAN PURATA BULANAN KESELURUHAN ZARAHAN TERAMPAI

PLOT OF TSP/TEMP SYMBOL USED IS *



NOTE: 48 OBS HAD MISSING VALUES 15 OBS HIDDEN

DEPENDENT VARIABLE: TSP METHOD: MARQUARDT

ITERATION	B0	B1	RESIDUAL SS
0	0.620000000	1.420000000	94277.472617198
1	0.627048709	1.424299806	93704.934839831
2	0.622001205	1.426659883	93704.849051288
3	0.616059121	1.429558622	93704.834524257
4	0.615317440	1.429936572	93704.832382806
5	0.615311880	1.429939523	93704.832382313

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS DEPENDENT VARIABLE TSP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	2	919000.4568	459500.2284
RESIDUAL	189	93704.8324	495.7928
UNCORRECTED TOTAL	191	1012705.2892	
(CORRECTED TOTAL)	190	96729.3065	

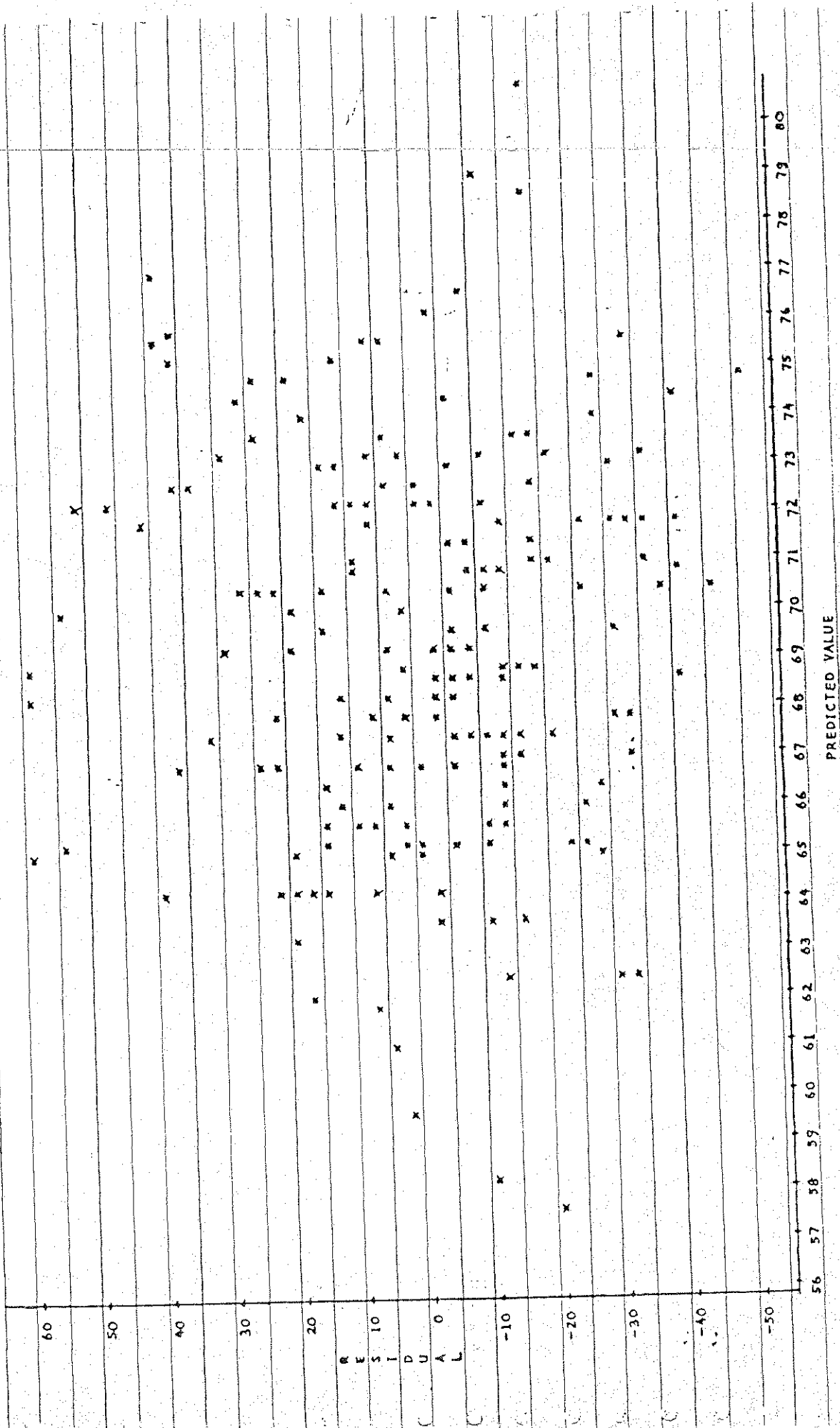
PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL LOWER	UPPER
B0	0.615311800	1.1976355083	-1.7274384076	2.9580621667
B1	1.429939523	0.5036755751	0.2785710170	2.5813080298

ASYMPTOTIC CORRELATION MATRIX OF THE PARAMETERS

	B0	B1
CORR		
B0	1.0000	-0.9997
B1	-0.9999	1.0000

ANALYSIS RESIDUAL BAGI MODEL $80 \times (TEMP \times B1)$

PLOT OF TSPRESKTSPHAT SYMBOL USED IS *



NOTE: 48 OBS HAD MISSING VALUES 12 OBS HIDDEN