《報文》 HWAHAK KONGHAK Vol. 17, No. 4, August. 1979, pp. 293—304.
(Journal of the Korean Institute of Chemical Engineers)

대기 오염 사건의 지속 시간에 관한 통계학적 모델

朴 玉 鉉
동아대학교 공과대학 위생공학과
(접수 1979. 3. 28)

The Statistical Models of Duration of Atmospheric Pollution Events

Park Ok Hyun

Department of Sanitary Engineering, College of Engineering,
Dong-A University, Busan 600-01, Korea
(Received March 28, 1979)

요 약

1976년 Australia의 New South Wales주의 1개 측정지점으로 부터 매시간 평균 SO_2 농도 자료, 그리고 16개 측정지점들로 부터 매일 평균 SO_2 농도 자료가 수집되었다. 어떤 농도를 초과한 대기오염사건의 지속시간에 관한 간단하고 정확한 통계학적 모델을 만들므로서 일반적이고도 새로운 진리를 발견하기 위하여 그 자료들을 처리하였다.

Abstract

Daily average and hourly average SO_2 concentrations data in 1976 were collected from 16 stations and 1 station respectively in New South Wales, Australia. The data have been processed with regard to the duration of pollution events for which the concentration exceeded some levels in order to find out general new things through formulating simple and accurate models.

---

1. Introduction

Usually the quality standards are formulated as average values that should never be exceeded or that could be exceeded with given frequencies. However, air pollution effects except eye irritation, taste and odour have been found to be propprtional to exceeded dosage expressed as the product of pollutants concentration exceeded threshold limit and exposure duration[1]

$$\text{Dosage} = (C - C_t)\ t \tag{1}$$

where C=pollutants concentration

C_t=threshold concentration

t=continued exposure time

Therefore it is of relevant interest to know the distribution of concentration and also a description of the concentration events in terms of their number and the distribution of durations.

2. Literature review

According to Kalpasanov, Y. and G.Kurchatove[2] (1976) the statistical disstributions of chemical pollutants are not normal and not all the results reported for different pollutants in various cities fit an exponential distribution even for 5 minute averages[7].

Application of logarithmic transformation brings the distributions closer to the log normal, but does not completely normalize them [2,10,11], though Larsen et al[3] (1976) indicate that such air pollution data show log normal distribution in all cities regardless averaging time.

The near log-normality of pollutant concentrations can be explained on the basis of the near log-normality of wind speed distributions alhough this explanation does nott establish that wind speed distributions are solely responsible for observed concentration distributions[11].

Since air pollution data typically exihibit a distribution which can be approximated by a log normal distribution, the summary statistics usually estimated are the geometric mean M_g and the standard geometric deviation S_g[4].

Larsen, R.I.[3] (1976) established equation (2) and Darby, W.P. and C.J. Gregory[4] (1976) made equation(3) and equation(4) on the basis of log normal distribution.

$$S_{g(c)} = 0.67\ M_{g(c)}^{-0.33} \tag{2}$$

$$M_{g(c)} = \exp\ (\mu) \tag{3}$$

$$S_{g(c)} = [\exp(\sigma^2)]^{0.5} \tag{4}$$

where $S_{g(c)}$=standard geometric deviation of pollutant concentration.

$M_{g(c)}$=geometric mean of pollutants concentration.

$\mu = \ln C/n$, the average of the logarithms of the original pollutant concentrations.

$\sigma = [\Sigma(\ln C - \mu)^2/n]^{0.5}$, the standard deviation of the logarithms of the original pollutant concentrations.

Duration of events D can be described by means of the arithmetic mean $M_{a(d)}$, the geomatric mean $M_{g(d)}$, the standard deviation $S_{a(d)}$, the standard geometric deviation $S_{g(d)}$ and the maximum observed duration D_m which are the statistics of the duration D[1].

Durfuca, G. and M. Giugliano[1] (1977) established equation (5).

$$M_{(d)}, S_{(d)} = \frac{1}{C^{k1}} \exp(K_2 + K_3\ \bar{C}_1) \tag{5}$$

where $M_{(d)}$=arithmetic or geometric mean of the duration of SO_2 concentrations in hours

$S_{(d)}$=standard deviation or standard geometric deviation of the duration of SO_2 concentrations in hours.

K_1, K_2 and K_3=numerical constants.

$\bar{C}_1$=long term average SO_2 concentration.

It was indicated by the above authors that equation(5) fails for the values of C less than 0.10 ppm.

According to Durfuca, G. and M. Giugliano [1] (1977), if the duration D is log normally distributed, then $M_{a(d)}$ and $M_{g(d)}$ must be related by equation(6).

$$\frac{M_{a(d)}}{M_{g(d)}} = \exp[0.5(\ln S_{g(d)})] \qquad (6)$$

Again the authors expressed total exposure time as

$$T = A \exp[-K(C+B)] \qquad (7)$$

where $K = 17.4 \exp(-4.86\bar{C})$

T=the total time for which C is exceeded in hours.

A, B=constants.

This paper concentrates on finding new things through the test of validity of equation(5) and equation(7) for New South Wales, Australia.

3. Data

SO_2 was determined by scrubbing the sample air with a dilute solution of H_2O_2 and Sulphuric acid (British Standard method No. 1747 part 3). Any SO_2 present in the sample air was converted to Sulphuric acid, and the resulting increase in acidity is determined by titration (British Standard Method) or conductivity (Continuous Method).

The data were obtained from New South Wales, State Pollution Control Commission's statewide air monitoring net work as shown in *Fig.* 1.

Hourly average data were obtained from Sydney city monitoring station and daily average data are from 16 stations (4 in Western N.S.W., 8 in Central Sydney and 4 in Southern N.S.W.) in 1976.

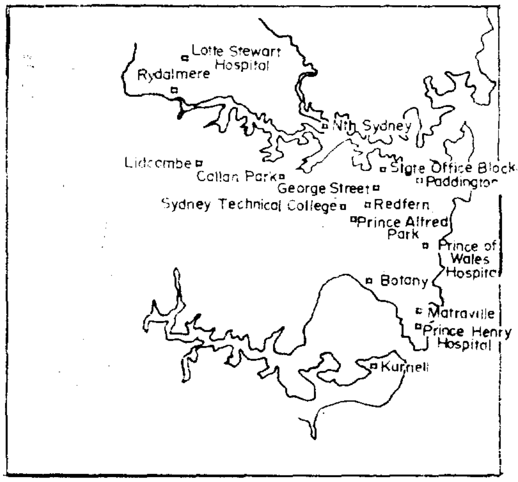

Fig. 1. Location of SO_2 monitoring sites in N.S.W.

4. Distribution

4.1. Distribution of pollutants concentration

The usual first step in analyzing air pollution data is to examine the frequency distribution[9].

4.1.1. Normality

Normality of pollutants concentration was checked under the consideration that it could be site dependent.

Normality tests based on estimats of skewness r_1, kurtosis r_2 and kolmogorov coefficient λ which can be expressed as equation (8), equation(9) and equation(10) respectively are more convenient than the chi-squared test[2,6].

$$r_1 = \frac{\mu_3}{\mu_2^{1.5}} \qquad (8)$$

where $\mu_2 = \frac{\Sigma(x_i - \bar{x})^2 \cdot f_i}{\Sigma f_i}$

$$\mu_3 = \frac{\Sigma(x_i - \bar{x})^3 \cdot f_i}{\Sigma f_i}$$

$$r_2=\frac{\mu_4}{\mu_2^2}-3 \quad (9)$$

where $\mu_4=\frac{\Sigma(x_i-\bar{x})^4\cdot f_i}{\Sigma f_i}$

$$\lambda=D\sqrt{n} \quad (10)$$

where D=the maximal dfference between empiric and theoretical distributions.

n=number of samples.

The illustration of computation of Kolmogorov's test for normality is as *Table* 1.

Table 1. Illustration of computation of kolmogorov's test.

1	2	3	4	5	6	7	8	9	10
concentration in $\mu g/m^3$	x_i	f_i	$f_i\cdot x_i$	Σf_i	$\Sigma f_i/n$	$t=\frac{x_i-\bar{x}}{S_x}$	$F(t)$	$1/2+F(t)$	9−6
2.5…… 7.5	5	9	45	9	0.027	−2.60	−0.495	0.005	0.022
7.6……12.5	10	6	60	15	0.045	−2.01	−0.478	0.022	0.023
12.6……17.5	15	23	345	38	0.113	−1.43	−0.424	0.076	0.037
17.6……22.5	20	40	800	78	0.232	−1.85	−0.302	0.198	0.034
22.6……27.5	25	93	2325	171	0.509	−0.26	−0.103	0.397	0.112
27.6……32.5	30	94	2820	236	0.789	0.32		0.626	0.163
32.6……37.5	35	40	1400	305	0.908	0.91		0.819	0.089
37.6……42.5	40	17	680	322	0.958	1.49		0.932	0.026
42.6……47.5	45	8	360	330	0.982	2.08		0.981	0.001
47.6……52.5	50	4	200	334	0.994	2.66		0.996	0.002
52.6……57.5	55	1	55	335	0.997	3.25		0.999	0.002
57.6……62.5	60	1	60	336	1.000	3.83		1.000	0.009

$$\bar{x}=\frac{\Sigma x_i\cdot f_i}{\Sigma f_i}=\frac{9150}{336}=27.232$$

$$S_x=\sqrt{\frac{\Sigma(x_i-\bar{x})^2\cdot f_i}{\Sigma f_i}}=\sqrt{\frac{24575.89}{336}}=8.552$$

$\lambda=D\sqrt{n}=0.163\quad\sqrt{336}=2.99$

* explanation

column

1; concentration in interval

2; center of interval

3; frequency in the respective interval

4; variable times frequency

5; cumulative frequencies

6; cumulative frequencies of the empirical distribution

7; secondary variable $t=\frac{x_i-\bar{x}}{S_x}$

8; $F_{(t)}=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}\cdot dt$

9; cumulative frequencies of the normal distribution

10; differences between empirical and normal distribution

Statiatical data concerning normality are summarized in *Table* 2. *Fig.* 2-1, *Fig.* 2-2 and *Fig.* 2-3 show the normality of SO_2 data for various stations.

Table 2. Statistical data of normality of SO_2 concentrations.

	site	no. of values counted	max.	mean	S.D.	r_1	standard error of r_1 $\pm 1.96\sqrt{\frac{24}{n}}$	r_2	standard error of r_2 $\pm 1.96\sqrt{\frac{6}{n}}$	λ
1	Dundas	336	60.6	27.2	8.4	0.06	−0.262	1.25	−0.524	2.99
2	Kensington	3363	77.0	42.2	11.5	0.34	−0.252	0.47	−0.504	2.61
3	Kurnel	362	129.0	61.7	20.3	0.66	−0.252	−0.02	−0.505	5.37
4	Lanecove	327	56.0	25.3	27.6	−0.04	−0.265	1.12	−0.631	2.10
5	Little bay	322	62.0	33.2	8.5	−0.07	−0.268	0.05	−0.535	1.99
6	Matraville	354	68.0	32.8	10.0	0.11	−0.255	0.38	−0.510	1.99
7	Rozzela	347	70.0	35.9	12.0	0.21	−0.258	0.06	−0.515	1.38
8	Rydalmere Hospital	354	12.0	30.4	6.5	0.19	−0.255	0.68	−0.510	3.25
9	Macquarie Univ.	340	67.0	29.3	9.8	0.44	−0.260	1.43	−0.251	2.51
10	Stratfield	354	112.0	47.0	17.0	1.02	−0.255	0.96	−0.510	3.07
11	Sydney George-Marcket street	341	152.0	70.2	19.7	0.40	−0.260	1.85	−0.520	1.96
12	Sydney George st. north	355	100.0	43.0	11.3	1.02	−0.255	3.32	−0.510	3.03
13	Sydney Queen-Victoria building	301	92.0	40.0	15.8	0.57	−0.277	0.48	−0.553	2.05
14	Sydney State-office block	362	83.0	33.4	12.3	0.67	−0.252	1.48	−0.505	2.15
15	Sydney Tech. College	350	134.0	46.1	17.9	1.26	−0.257	2.70	−0.513	2.43
16	Woolooware	343	44.0	25.1	7.1	−0.06	−0.259	0.12	−0.518	2.54

For large samples, n greater than 200 say, rough tests of normality may be obtained by comparing the sample estimates of skewness r_1 and the excess of kurtosis r_2 with the approximate values of their standard errors which are $(6/n)^{0.5}$, $(24/n)^{0.5}$ respectively[6].

For example, for $n=1000$, the approximate standard error$(24/n)^{0.5}$ of the estimate of kurtosis is 0.155, and assuming a symetrical distribution, 95% of values should be between −0.30 and +0.30(±1.96 times 0.155).

At $\lambda > 0.50$, the hypothesis for normality (log-normalty) of the distribution is rejected at a level of significance $p > 0.05$.

The test of skewness and kurtosis indicates that 50% and 44% of places around New South Wales respectively have normality, while kolmogorov test shows that the hypothesis for normality is rejected for all places.

It is clear that the normality criteria of the test of skewness and kurtosis is significantly apart from that of kolmogorov test. For 13 places out of 16 have positive kurtosis meaning more peaked(narrow) distribution and 12 places have positive skewness indicating right tailed, clustering to the left.

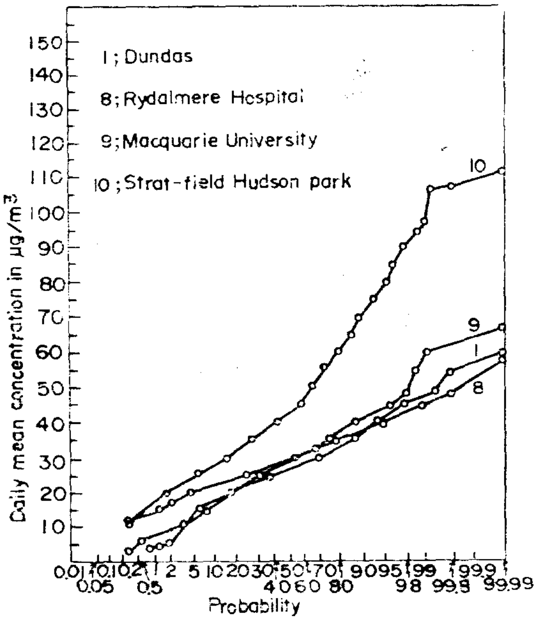

Fig. 2-1. Normality plot of SO_2 data for Western N.S.W.

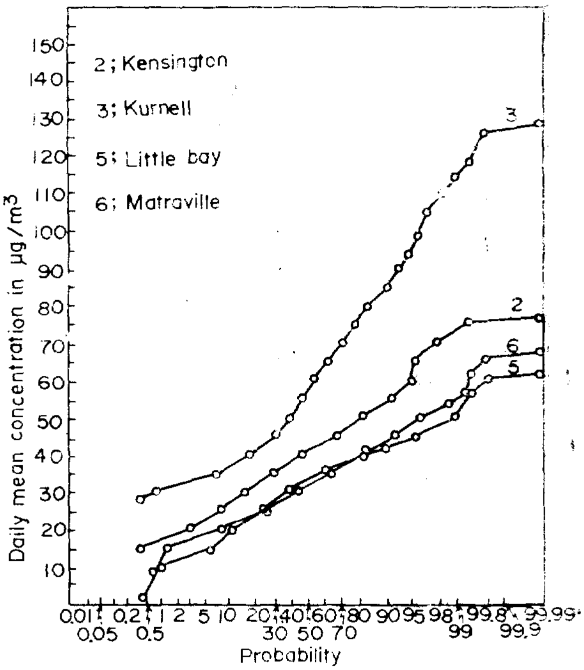

Fig. 2-3. Normality plot of SO_2 data for Southern N.S.W.

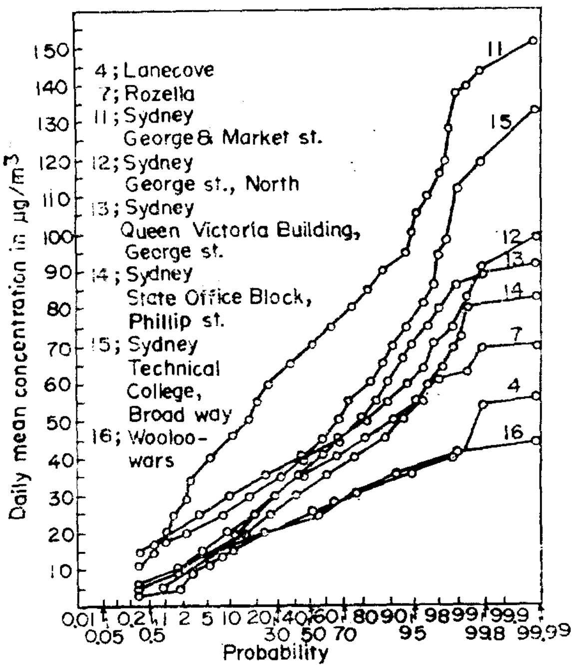

Fig. 2-2. Normality plot of SO_2 data for Central Sydney.

4.1.2. Log-normality

Geometric mean and standard geometric deviation were calculated by using equation (3) and equation (4).

Calculation of kolmogorov test for log-normality was similarly done as *Table* 1 on the basis of following modification of log-normal distribution funtion.

$$F(C)=\frac{1}{\sqrt{2\pi}\ \log\sigma_g}\int_{-\infty}^{\log C}\exp-\frac{1}{2}\left(\frac{\log C-\log C_{gm}}{\log\sigma_g}\right)^2 d\log C \qquad (11)$$

Let $t=\dfrac{\log C-\log C_{gm}}{\log\sigma_g}$,

$d\log C=\log\sigma_g\, dt$

Thus $F(t)=\dfrac{1}{\sqrt{2\pi}\ \log\sigma_g}\exp\left(-\dfrac{t^2}{2}\right)\log\sigma_g\, dt$

$$=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)dt \qquad (12)$$

Statistical data concerning log-normality are summarized in *Table* 3 and *Fig.* 3-1~3 show the log-normality of SO_2 data for various stations.

From *Table* 3 and *Fig.* 3-1~3, It is confirmed that logarithmic transformation brings the distributions closer to the log-normal, but it does not completely normalize them in all places except one. Again the log-normality of SO_2 concentrations distributions appears to be site dependent.

Table 3. Statistical data of log-normality of SO_2 concentratios

station number	geometric mean C_{gm} ($\mu g/m^3$)	standard geometric deviation σ_g ($\mu g/m^3$)	kolmogorov coefficient λ
1	24.9	1.265	1.03
2	39.8	1.244	1.68
3	56.4	1.418	0.93
4	23.2	1.293	1.84
5	31.1	1.254	2.15
6	30.3	1.327	1.43
7	34.0	1.303	2.25
8	27.5	1.236	0.49
9	26.1	1.379	1.00
10	39.8	1.558	1.32
11	66.5	1.265	1.27
12	39.7	1.254	0.73
13	[illegible]	1.417	1.34
14	[illegible]	1.343	2.36
15	40.7	1.440	0.64
16	22.5	1.333	0.91

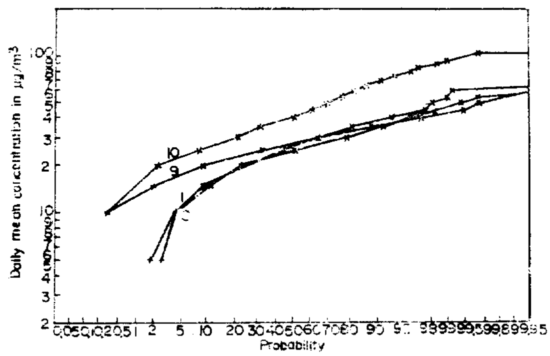

Fig. 3-1. Log-normality plot of SO_2 data for western N.S.W.

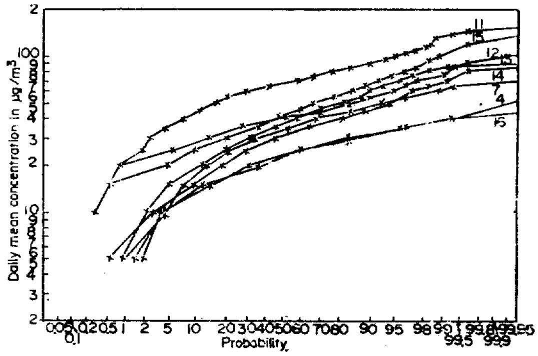

Fig. 3-2. Log-normality plot of SO_2 data for central Sydney

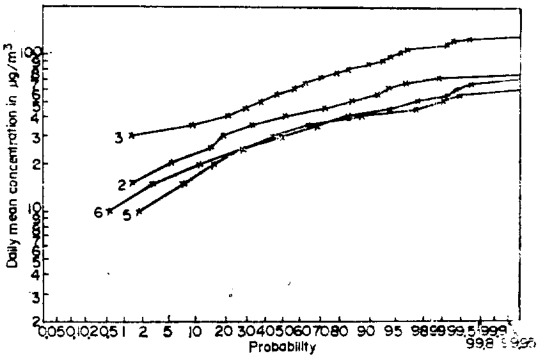

Fig. 3-3. Log-normality plot of SO_2 data for southern N.S.W.

4.2. Distribution of time of SO_2 pollution events.

4.2.1. Total time for which concentration is exceeded.

Fig. 4.1. and *Fig.* 4-2 show the distribution of $\log_{10}$(Duration) with daily average SO_2 concentration in $\mu g/m^3$ for various stations. The data of *Fig.* 4-1 and *Fig.* 4-2 can be expressed by;

$$\text{Log } T = -KC + B \tag{13}$$

where T is the total time for which C is exceeded in % of days in a year.

B is constant(intercept).

C is concentrations in $\mu g/m^3$

K can be expressed as equation(14) or equation(15)

$$K=-0.00062\bar{C}_1+0.61 \qquad (14)$$

$$K=\exp(0.1485\bar{C}_1-2.813) \qquad (15)$$

where $\bar{C}_1$ is annual average of daily mean concentrations in $\mu g/m^3$.

Therefore if mean concentration (*Table* 2) and B(*Table* 4) values for each site are given, we can estimate the total time for which C is exceeded by using equation(13) and equation(14) or equation(15). Considering equation(7) it is clear that K is site dependent.

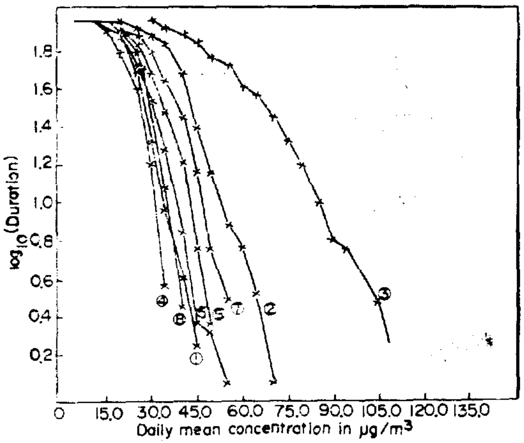

Fig. 4-1. Relationship between total duration and concentration exceeded

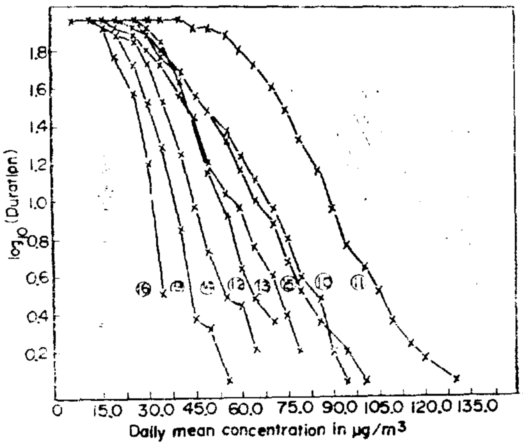

Fig. 4-2. Relationship between total durnation and concentrtaion exceeded

Table 4. Statistical data of log function of total time for which concentration is exceeded

station number	$\bar{C}_1$	K	B
1	27.232	0.0401	2.534
2	42.135	0.0401	1.784
3	61.796	0.0240	1.838
4	25.382	0.0412	1.438
5	33.199	0.0367	2.553
6	32.881	0.0368	1.686
7	35.851	0.0353	1.705
8	30.353	0.0383	1.363
9	29.176	0.0389	1.800
10	46.921	0.0299	1.924
11	70.161	0.0212	2.176
12	43.000	0.0317	1.760
13	39.950	0.0332	1.888
14	33.415	0.0365	2.462
15	46.157	0.0302	1.943
16	25.044	0.0414	1.442

4.2.2. Duration for which concentration is exceeded.

The values of $M_{a(d)}$ for various hourly SO_2 concentrations exceeded, C(pphm) from the Sydney monitoring station were computed and are shown in *Fig.* 5.

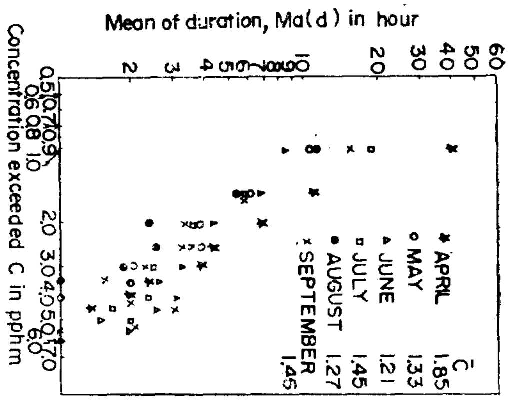

Fig. 5. Relationship between $M_{a(d)}$ and C exceeded

The basic equation expressing *Fig.* 5 may be established as

$$\ln M_{a(d)} = -K \ln C + \ln D \quad (16)$$

$$\therefore\ M_{a(d)} = \frac{D}{C^k} = \frac{1}{C^k}\ \exp.\ B'$$

where B' = intercept

K = slops of exponential function.

Thus another graph *Fig.* 6. showing the relationship between B' and $\bar{C}_2$(monthly average of hourly mean concentration) was drawn and equation(17) was eventually produced.

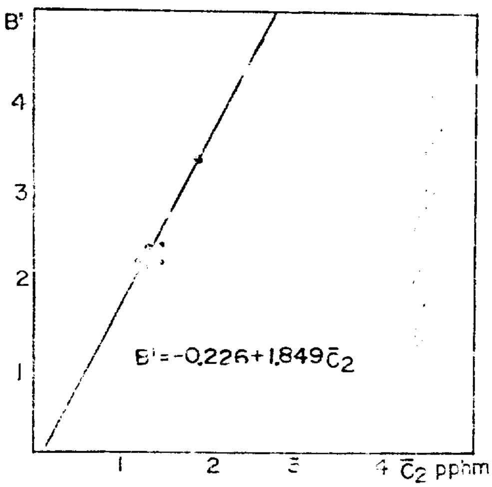

Fig. 6. Relationship between B' and $\bar{C}_2$

$$M_{a(d)} = \frac{1}{C^{1.53}} \exp(-0.226 + 1.849\bar{C}_2) \quad (17)$$

Equation(17) behaves quite well for values of C greater than 1.5 pphm as shown in *Fig.* 7.

In the same manner as above, $S_{a(d)}$, $M_{g(d)}$ and $S_{g(d)}$ can be formulated as equation(18), equation(19) and equation(20).

$$S_{a(d)} = \frac{1}{C^{1.748}} \exp.(1.378 + 0.935\ \bar{C}_2) \quad (18)$$

$$M_{g(d)} = \frac{1}{C^{0.80}} \exp.(-0.008 + 1.139\bar{C}_2) \quad (19)$$

$$S_{g(d)} = \frac{1}{C^{1.27}} \exp.(-0.071 + 0.176\ \bar{C}_2) \quad (20)$$

Comparison between calculated values and observed values of $S_{a(d)}$, $M_{g(d)}$ and $S_{g(d)}$ are shown in *Fig.* 8, *Fig.* 9 and *Fig.* 10.

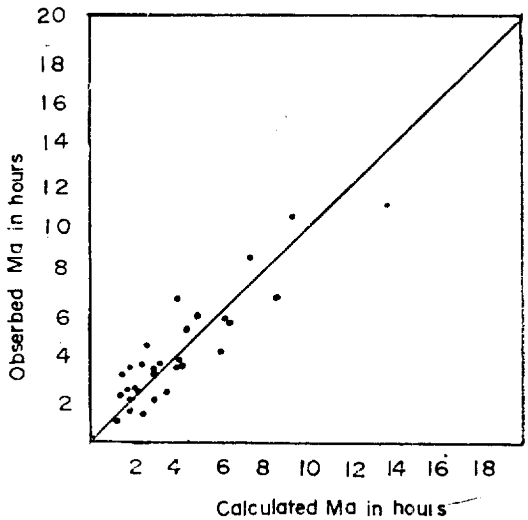

Fig. 7. Relationship between calculated M_a and observed M_a

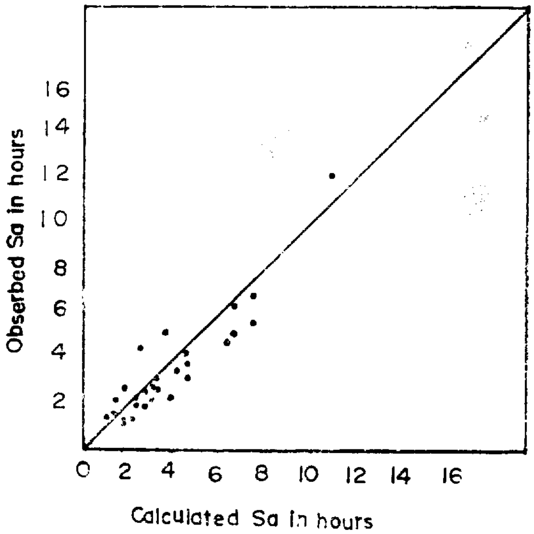

Fig. 8. Relationship between calculated S_a and observed S_a

5. Conclusion

1). It was found that normality criteria

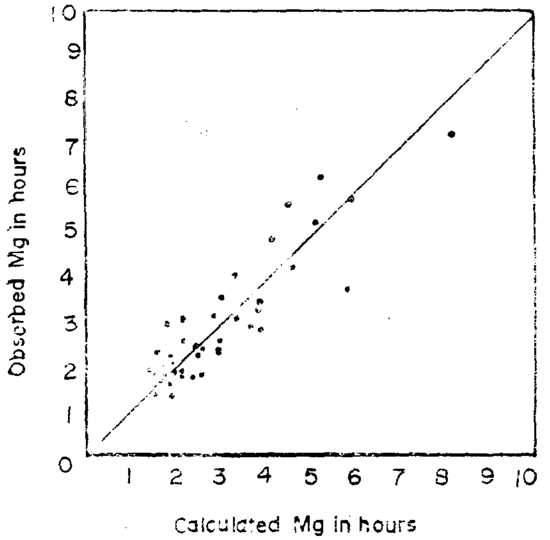

Fig. 9. Relationship between calculated M_g and observed M_g

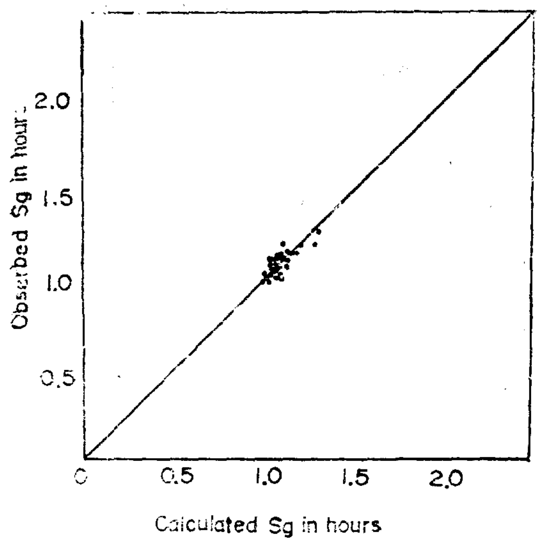

Fig. 10. Relationship between calculated S_g and observed S_g

of the test of skewness and kurtosis is significantly apart from that of kolmogorov test. Normal distribution according to skewness and Kurtosis is not normal one according to kolmogorov coefficients.

The distributions of SO_2 data from New South Wales are not normal and the normality of SO_2 data appears not to be site dependent. However, it was confirmed that logarithmic transformation brings the distributions closer to the log-normal and the log-normality of SO_2 concentrations distributions appears to be site dependent.

2). For Sydney monitoring station, M, $S=\frac{1}{C^{k1}}\exp(K_2+K_3\cdot\bar{C})$ may apply for SO_2 data greater than 1.5 pphm.

The values of K_1, K_2, and $K_3$are site dependent and vary with time(i.e. month). The low limit(i.e. 1.5 pphm) is also site dependent.

3) The total time for which concentration is exceeded can be expressed as

$$\text{Log } T=-KC+B.$$

K and B are again site dependent and K is a linear or exponential function of $\bar{C}_1$, annual average of daily mean concentrations.

Acknowledgement

Sincere gratitude is expressed to Mr. Ross Hayes, scientific officer of S.P.C.C. of N.S. W., Australia, for friendly serving air pollution data. Author wishes to express thanks to associate professor E.T. Linacre for his encouragement and some assistance in collecting data. Appreciation is also expressed to my students, S.O. Beak, K.S.Jeung and Y.W. Ha for their help with some calculations and drawing.

References

1. G. Durfuca, and M. Giugliano, Atm. Env., **11** (1977), 729.
2. Y. Kalpasanov and G. Kurchatov, J.A.P. C.A., **26** (1976), 961.
3. R.I. Larsen, C.E.Zimmer, D.A.Lynn and

K.G.Blemel, J.A.P.C.A., **17** (1976), 85.
4. W.P. Darby, and C.J.Greqory, J.A.P.C.A., **26** (1976), 233.
5. New South Wales, State Pollution Control Commission, Qnarterly Report of Air Quality Measurements in new South Wales (1976).
6. J.V. Maher and J.N. McRAE, "Upper wind Statistics," Commonwealth of Bureau of Meteorology, Australia. (1964).
7. P.J. Barry, Stochastic properties of atmospheric diffusivity. Preprint of a contribution to the "Effects of Sulphur in Canada" AECL-5012 (1975).
8. R.I. Larsen, J.A.P. C.A., **20** (1970), 214.
9. D.T. Mage, proceedings of the 68th annual meeting of the air polltion control association, (1975), 75-51. 4.
10. L.T. Brasser, S.A. Clarenburg, H.de. Graaf and J.R. pruys, Atm. Env., **5** (1971), 505.
11. K.E. Bencala and J.H. Seinfeld, Atm. Env., **10** (1976) 941.
12. T.C. Curran and N.H. Frank, proceedings of the 68th Annual Meeting of the Air Pollution Control Association, Boston, Massachusetts, (1975), 75-51. 3.