© *Proc. Indian natn. Sci. Acad.*, **48**, A, No. 2, 1982, pp. 144–160.
Printed in India.

Atmospheric Physico-Chemical Analysis (Air Pollution)

A STATISTICAL ANALYSIS OF AIR POLLUTION DATA FROM FIVE SAMPLING STATIONS IN SOUTHERN BENGAL

A. K. SAHA, F.N.A., A. SAHA *and* N. DAS

Department of Geology, Presidency College, Calcutta-700 073, India

and

A. MUKHOPADHYAY

Department of Mechanical Engineering, Jadavpur University, Calcutta-700 032, India

(*Received 17 June 1981; after revision 16 November 1981*)

Computer processing of the systematic data on seven air pollution and four meteorologic parameters (max. and min. temperatures, relative humidity and wind speed) over a period of nearly two years at 5 stations (Durgapur, Debagram, Jhalda, Jhargram and Digha) in the southern part of West Bengal have brought out the following particularly interesting results :– (*a*) the individual pollution parameters at particular stations may be estimated in terms of the meteorological variables at moderate to high level of confidence; (*b*) effective discrimination of the seasonwise pollution levels between pairs of sampling stations is possible through use of linear discriminant functions; (*c*) the pollution levels of Durgapur, Debagram—Jhalda and Digha—Jhargram form three distinct clusters in progressively decreasing order; (*d*) Varimax-rotated *R* mode factor analysis suggests influence of at least five distinct processes (or causes) in determining the distribution and migration of the pollutants; and (*e*) clear discrimination between the seasonal data for individual stations is possible through use of factor-score diagrams.

Keywords : Air Pollution; Multivariate Statistical Analysis; Southern Bengal

INTRODUCTION

As part of an inter-disciplinary and inter-institutional Research Project to study some problems of air pollution in southern West Bengal, systematic sampling for particulate matter and some gaseous pollutants along with collection of meteorologic data was carried out over a period of a little over two years between December 1977 and January 1980 at five sampling stations (Jhalda, Durgapur, Debagram, Digha, Jhargram), 4-5m above the ground level. The pollution parameters sampled are —

(*i*) Suspended matter in air, sampled for 24-hour periods on glass fibre filters, usually thrice every week.
(*ii*) Dustfall, sampled for one month periods.
(*iii*) Sulphur dioxide.
(*iv*) Nitrogen oxide.
(*v*) Total oxidants.

(*iii*)–(*v*): } Sampled for a couple of 12-hour periods (8 a.m.-8 p.m.; 8 p.m.-8 a.m.), usually thrice every week.

(*vi*) Hydrogen sulphide.
(*vii*) Carbon monoxide.
} Sampled occasionally for 24-hour periods.

(*viii*) Periodical sampling of pollens.

(*ix*) Occasional sampling for mycological and bacteriological studies.

The equipment for sampling the particulate and gaseous matter was locally fabricated and standardised. The different analyses were carried out in accordance with the specifications laid down in the Indian Standards Institution, standard nos. IS 5182, parts I, II, IV, V, IX and XII. The details of the procedure are given elsewhere (Saha *et al.*, 1980).

The five sites (Fig. 1), where systematic sampling for air pollution was carried out are :—

(1) *Durgapur*—Highly industrialised (including a steel plant), partly rocky undulating plain area; sampling site was on roof of a pump house in the compound of Durgapur Govt. College which is situated on the northern outskirts of the Durgapur City complex;

(2) *Jhalda*—Semi-urban, rocky undulating plain area with some minor lac-processing factories and a railway junction close-by; sampling site was on roof of the old A. M. College building;

(3) *Debagram*—Purely agricultural rural, alluvial flat area with no industry; sampling site was on roof of a farm house surrounded by agricultural land;

(4) *Jhargram*—A semi-urban, forested area on a lateritic plain area; sampling site was on roof of the pump house in the compound of the Jhargram Raj College;

(5) *Digha*—Sea-coast resort with coastal sand dunes and with no industries; sampling site was on roof of the 'Nirala' canteen building.

A large volume of systematic diurnal data were accumulated for the following pollution and meteorologic variables for the five stations, which were stored on computer punch-cards in BASIC Format, upto 6 columns being used for each variable :—

1. Maximum temperature (MAXT)
2. Minimum temperature (MINT)
3. Relative humidity (RLHUM)
4. Windspeed (WNDSP)
5. Suspended matter (SUSMT)
6. Sulphur dioxide, day (SO_2DY)
7. Sulphur dioxide, night (SO_2NT)
8. Nitrogen oxides, day (NO_2DY)
9. Nitrogen oxides, night (NO_2NT)
10. Total oxidants, day (OXIDY)
11. Total oxidants, night (OXINT)

The data for each station were grouped seasonwise as follows :—

November to February—Winter; March to June—Summer; July to October—Rainy season; the mean seasonwise values for the four pollution variables at the five stations are given in Table I.

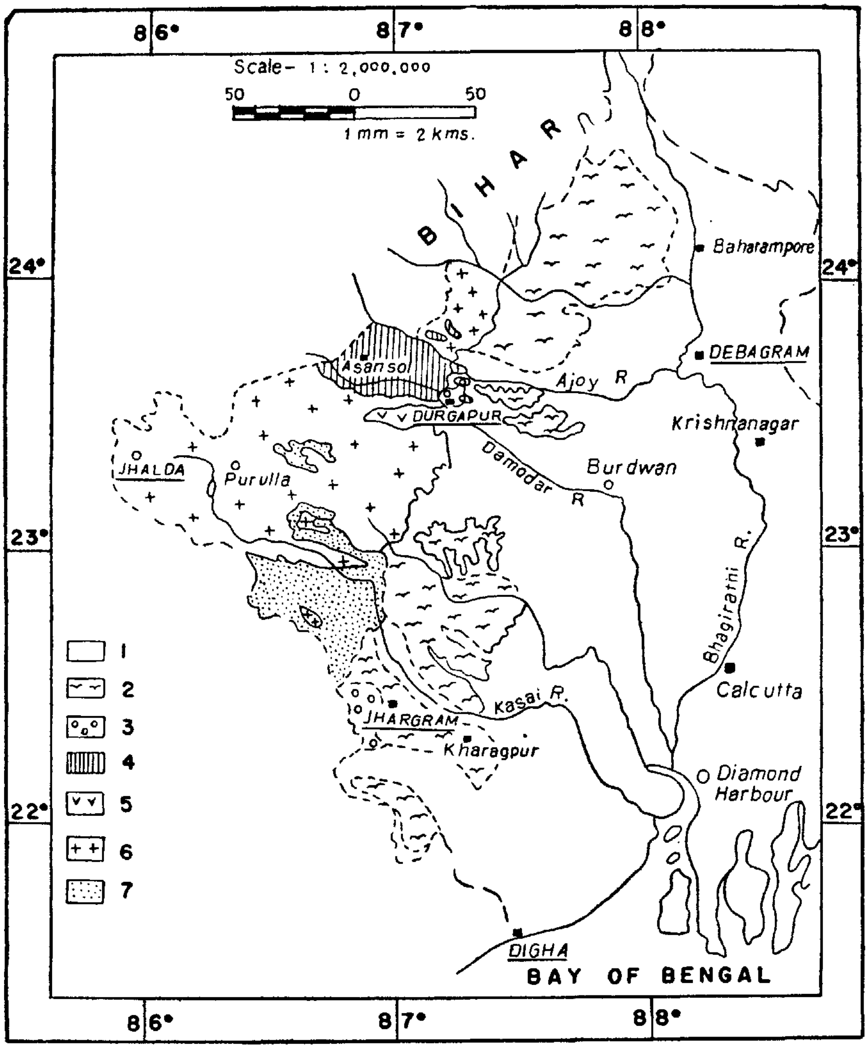

FIG. 1. Map of southern West Bengal showing locations of the air-sampling stations in relation to surface geology (sampling stations underlined). 1-Recent Alluvinm, 2-Older Alluvium & Laterite, 3-Tertiaries, 4-Gondwanas & Rajmahal Trap, 5-Anorthosite, 6-Granite & Gneiss, 7-Metamorphites.

For most of the statistical computations, the day and night values for the gaseous pollutants were combined as 24 hour values.

Using the systematic pollution and meteorologic data as mentioned above, four sets of statistical analysis were undertaken with the following objectives :—

(*a*) To study the inter-correlations among the individual pollution and meteorological variables;

(*b*) To find out if the values of individual pollution parameters at particular stations could be estimated successfully on the basis of one or more of the meteorologic parameters;

TABLE I

Seasonal means of suspended matter and gaseous pollutants

(All measurements in $\mu g/m^3$)

Stations	Period	Suspended Matter	SO_2	N-Oxides	Total Oxidants
Debagram	I : Dec. 77–Feb. 78	104.67	5.01	64.23	13.06
	II : Mar. 78–Jun. 78	177.78	10.53	25.71	28.44
	III : Jul. 78–Oct. 78	40.38	15.08	10.70	14.83
	IV : Nov. 78–Feb. 79	149.49	9.22	22.91	19.86
	V : Mar. 79–Jun. 79	202.38	10.78	14.56	12.59
	VI : Jul. 79–Oct. 79	47.35	10.09	26.30	25.10
	VII : Nov. 79–Dec. 79	100.40	4.90	39.95	3.80
Durgapur	I : Dec. 77–Feb. 78	195.68	11.81	48.74	11.33
	II : Mar. 78–Jun. 78	189.52	16.94	21.76	20.39
	III : Jul. 78–Oct. 78	57.66	13.01	7.02	16.72
	IV : Nov. 78–Feb. 79	202.33	21.44	42.75	8.16
	V : Mar. 79–Jun. 79	273.45	20.48	34.55	14.65
	VI : Jul. 79–Oct. 79	82.56	23.20	26.09	24.30
	VII : Nov. 79–Jun. 80	193.99	26.37	73.16	6.35
Jhalda	I : —	—	—	—	—
	II : —	—	—	—	—
	III : Jul. 78–Oct. 78	71.49	15.81	8.38	18.71
	IV : Nov. 78–Feb. 79	168.77	15.74	28.00	19.07
	V : Mar. 79–Jun. 79	190.47	7.39	14.23	19.86
	VI : Jul. 79–Oct. 79	117.92	10.68	34.29	15.85
	VII : Nov. 79–Dec. 79	233.41	33.42	48.17	6.05
Digha	I : Jan. 78–Feb. 78	92.75	28.09	24.14	43.08
	II : Mar. 78–Jun. 78	74.88	9.54	8.29	19.17
	III : Jul. 78–Oct. 78	43.53	10.81	12.31	16.68
	IV : Nov. 78–Feb. 79	110.54	14.71	36.04	11.23
	V : Mar. 79–Jun. 79	101.59	10.53	10.19	4.18
	VI : Jul. 79–Oct. 79	80.82	21.14	26.16	3.91
	VII : Nov. 79–Dec. 79	124.86	22.68	43.39	5.25
Jhargram	I : —	—	—	—	—
	II : Mar. 78–Jun. 78	87.30	7.79	16.18	17.98
	III : Jul. 78–Oct. 78	50.27	10.52	9.22	11.42
	IV : Nov. 78–Feb. 79	110.86	9.27	15.81	5.32
	V : Mar. 79–Jun. 79	103.84	8.39	7.41	8.64
	VI : Jul. 79–Oct. 79	36.67	8.84	8.40	1.85
	VII : Nov. 79–Dec. 79	95.59	5.57	29.93	5.06

(*c*) To compare the incidence of pollution at the five sampling stations computed season-wise;

(*d*) To try to identify the significant natural processes controlling the pollution levels at the sampling stations, and to assess their relative importance.

METHODS OF COMPUTATIONS

Prior to the computations, the data sets for the different stations were grouped and labelled as follows :—

	Debagram	Durgapur	Jhalda	Jhargram	Digha
Nov. 77 - Feb. 78	DBG 1	DRG 1	—	—	—
Mar. 78 - June 78	DBG 2	DRG 2	—	JGM 2	DGA 2
Jul. 78 - Oct. 78	DBG 3	DRG 3	JLD 3	JGM 3	DGA 3
Nov. 78 - Feb. 79	DBG 4	DRG 4	JLD 4	JGM 4	DGA 4
Mar. 79 - Jun. 79	DBG 5	DRG 5	JLD 5	JGM 5	DGA 5
Jul. 79 - Oct. 79	DBG 6	DRG 6	JLD 6	JGM 6	DGA 6
Nov 79 - Jan. 80	DBG 7	DRG 7	JLD 7	JGM 7	DGA 7

Starting with the pollution and meteorological data for the five stations stored on computer punch-cards in BASIC Format, four sets of statistical analysis were done as follows, using the Burroughs 6700 computer at the Regional Computer Centre, Calcutta.

AN 1 : Computation of the correlation coefficient matrix for four meteorological and pollution variables at each sampling station, the day and night values of each gaseous pollutant being combined into single 24-hour values. The BASIC program package CORREL was used for this purpose.

AN 2 : Stepwise multiple regression to predict the four pollution variables in terms of one or more than one combinations of the meteorological parameters, MAXT, MINT, RLHUM and WNDSP for the data sets DRG 2–4, JLD 4–5, JGM 3–4, DGA 5 (10 sets of computations). The BASIC program package STEPR was used for these computations.

AN 3 : Computation of multivariate analysis, Mahalanobis D^2 and LDF using the data for individual seasons (i. e., summer, rainy season and winter) comprising of the seven pollution parameters, SUSMT, SO2DY, SO2NT, NO2DY, NO2NT, OXIDY and OXINT, in order to assess the season-wise multivariate differences among the 5 sampling stations. The Mahalanobis D^2 values indicate the degree of dissimilarity between the pairs of different stations considering the seven pollution parameters together, while the Linear Discriminant Function (LDF) represents linear equation for the hyperplane which most effectively separates the plots for the individual observations belonging to two stations in the multi-dimensional space represented by the seven pollution parameters. A FORTRAN program modified from Wolleben *et al.* (1968) was used for this purpose. The program consists of computation of the multivariate means for the data sets for the five stations, Mahalanobis D^2, F values for the D^2 values and LDF between all possible pairs of the data sets; LDFs of the individual diurnal data sets were also computed as a measure of the efficiency of the LDF.

AN 4 : R-mode Factor analysis (using Varimax rotation) including computation of rotated Factor Score matrix, the variables being SUSMT, SO2, NO2, OXID (i.e. four pollution variables) and MAXT, MINT, RLHUM and WNDSP (i.e., four meteorological parameters) was carried out separately with data for 5 stations. As is well known, Factor analysis helps interpret the structure of similarity coefficient matrix (such as, linear correlation coefficients) between several variables (R-mode) or between the samples (Q-mode) in a given data set (cf. Harman, 1967). In *R*-mode factor analysis, the similarity coefficients between the variables are considered as vectors in a multi-dimensional space. A number of orthogonal axes are then successively fitted to the cluster of these vectors such that the maximum possible degree of association of the variables is taken into consideration. These orthogonal axes (or factors) are expressed as a linear combination of coefficients (or loadings) of the individual variables. Usually a relatively small number of factors account for the major part of the total variations in the data. Each factor is supposed to represent an independent source of the variations and may give some clue to the genetic processes. For more meaningful interpretation, the orthogonal factors are commonly rotated rigidly so that the variance of the factor loadings on each factor is a maximum (varimax rotation). Each particular observation (sample) may be computed into a *score* by projecting it on to one factor axis; thus for each diurnal observation of pollution and meteorological variables, a factor score was computed for each varimax-rotated factor.

The FORTRAN IV Program after Davis (1973) with suitable adaptations was used for the computations, the data sets for the eight variables being first standardised.

In carrying out the above-noted analysis, those diurnal data sets in which more than one pollution variable is below detection limit were omitted.

RESULTS AND DISCUSSIONS

A. *Inter-Correlations of the Pollution and the Meteorologic Variables*

The following are the significant results obtained:

(1) A general tendency for the gaseous pollutants to decrease with windspeed is evident for all the sampling stations (Table II).

(2) The significantly high correlations between the variables are by and large positive.

(3) A general positive correlation among the gaseous pollutants is evident.

(4) SUSMT has significant negative correlation with RLHUM except at Debagram and Digha; also SUSMT has moderate positive correlation with NO_x, especially at Durgapur.

(5) As expected, a strong positive correlation between MXTMP and MNTMP occurs at all the stations.

(6) Except at Debagram, NO_x is moderately correlated negatively with MAXT and MINT.

B. *Stepwise Multiple Regression Analysis*

It is interesting to note that in spite of non-inclusion of a large number of variables (such as fluctuations in factory emissions, wind direction, rainfall etc.) which must be affecting the diurnal values of the pollution variables, it is possible in most cases to estimate for the individual seasons SUSMT, SO_2 and NO_x at the

TABLE II

Correlation matrices for stationwise pollution and meteorologic data

A. Digha (113 data sets)

	SO_2	NO_x	OXID	SUSMT	MAXT	MINT	RLHUM	WNDSP
SO_2	1.000							
NO_x	0.531*	1.000						
OXID	0.104	0.201	1.000					
SUSMT	0.173	0.251	0.166	1.000				
MAXT	0.017	−0.340*	−0.369*	−0.217	1.000			
MINT	−0.121	−0.465*	−0.262	−0.265	0.763*	1.000		
RLHUM	−0.069	−0.034	−0.235	−0.194	0.176	0.3000	1.000	
WNDSP	−9.071	−0.065	−0.026	−0.044	−0.051	0.133	0.110	1.000

B. Durgapur (94 data sets)

	SO_2	NO_x	OXID	SUSMT	MAXT	MINT	RLHUM	WNDSP
SO_2	1.000							
NO_x	0.240	1.000						
OXID	0.002	−0.25	1.000					
SUSMT	0.236	0.376*	0.136	1.000				
MAXT	0.032	−0.341*	0.333*	0.325*	1.000			
MINT	−0.240	−0.467*	0.257	0.083	0.745*	1.000		
RLHUM	−0.125	−0.581*	0.145	−0.518*	0.04	0.185	1.000	
WNDSP	−0.106	−0.233	0.101	0.097	0.332*	0.379*	0.102	1.000

C. Jhargram (100 data sets)

	SO_2	NO_x	OXID	SUSMT	MAXT	MINT	RLHUM	WNDSP
SO_2	1.000							
NO_x	−0.004	1.000						
OXID	0.220	0.027	1.000					
SUSMT	0.13	0.134	0.016	1.000				
MAXT	−0.006	−0.309*	0.040	−0.151	1.000			
MINT	0.082	0.239	−0.034	−0.187	0.761*	1.000		
RLHUM	−0.022	−0.038	−0.080	−0.337*	−0.129	−0.236	1.000	
WNDSP	−0.070	−0.079	0.209	−0.043	0.199	0.283	0.004	1.000

D. Jhalda (90 data sets)

	SO_2	NO_x	OXID	SUSMT	MAXT	MINT	RLHUM	WNDSP
SO_2	1.000							
NO_x	0 405*	1.000						
OXID	−0.159	−0.172	1.000					
SUSMT	0 268	0.223	−0.227	1.000				
MAXT	−0.217	−0.261	−0.150	0.286	1.000			
MINT	−0.284	−0.393*	−0.140	−0.114	0.736*	1.000		
RLHUM	0.001	0.036	0.178	−0.397*	−0.503*	−0.146	1.000	
WNDSP	0.087	−0.121	0.112	0.002	0.127	0.086	−0.138	1.000

E. Debagram (54 data sets)

	SO_2	NO_x	OXID	SUSMT	MAXT	MINT	RLHUM	WNDSP
SO_2	1.000							
NO_x	−0.083	1.000						
OXID	−0.024	0.086	1.000					
SUSMT	−0.054	0.173	−0.133	1.000				
MAXT	0.116	0.125	−0.095	0.014	1.000			
MINT	0.124	−0.030	−0.104	−0.338*	0.703*	1.000		
RLHUM	−0.62	0.057	0.077	0.286	−0.337*	−0.325*	1.000	
WNDSP	−0.213	0.035	−0.276	0.010	−0.276	−0.367	−0.020	1.000

Asterisk (*) indicates statistically significant ($\alpha = .05$) correlation.

individual stations with moderate to high level of confidence using the data for only four meteorological variables, viz. MAXT, MINT, WNDSP and RLHUM. In general, predictability is highest for SUSMT, followed successively by NO_x, SO_2 and OXID. Seasonwise predictability appears to be best in winter when the meteorologic conditions are relatively stable, and is poorest in the Rainy season. Out of the four stations the data for which were analysed, those from Jhalda were found to

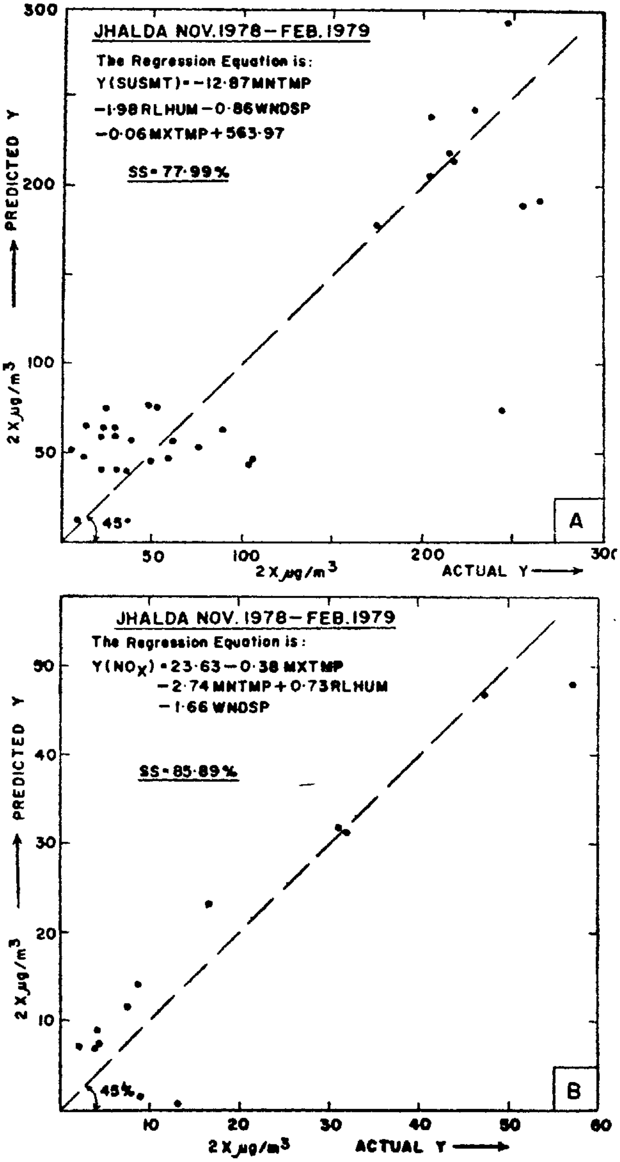

FIG. 2. Diagrams to show degree of fit of multiple regression equations to predict pollution parameters at Jhalda in terms of the meteorologic parameters — *A*. SUSMT. *B*. NO_x.

yield highest predictabilities (Fig. 2) followed successively by Durgapur, Digha and Jhargram. The predictability of the pollution parameters would certainly have been higher if the wind direction and rainfall data could be incorporated in the regression equations.

C. *Results of Multivariate Discrimination Studies*

The inter-station Mahalanobis-D^2 values for the data on 3 seasons (Summer, Rainy season and Winter) are shown in Table III. It is evident that the D^2 values

TABLE III

Mahalanobis-D^2 values and percentages of fit of the corresponding LDF's (in parenthesis) for pairs of sampling stations on the basis of Seasonwise data for 7 pollution parameters

	Jhalda	Debagram	Durgapur	Digha	Jhargram
A. *Summer*	28	13	46	42	33
Nos. of measurements					
1. Jhalda	—				
2. Debagram	1.148 NS	—			
3. Durgapur	1.581**	2.511**	—		
	(71.6%)	(76.3%)			
4. Digha	2.488**	2 039*	5.037**	—	
	(85.7%)	(89.1%)	(86.4%)		
5. Jhargram	2.548**	2.255*	5.241**	0.316 NS	
	(80.3%)	(84.8%)	(87.3%)		
B. *Rainy Season*					
No. of measurements	33	28	32	34	35
1. Jhalda	—				
2. Debagram	2.433**	—			
	(78.7%)				
3. Durgapur	1.459**	2.345**	—		
	(70.3%)	(83.3%)			
4. Digha	2.752**	2.388**	0.938 NS	—	
	(85.7%)	(75.8%)			
5. Jhargram	5.662**	1.109*	1.730**	1.612**	
	(80.9%)	(74.6%)	(73.2%)	(75.4%)	
C. *Winter*					
No. of. measurements	29	13	16	37	31
1. Jhalda	—				
2 Debagram	1,119 NS	—	—		
3. Durgapur	2.431**	3.071*			
	(84.4%)	(89.7%)			
4. Digha	1.692**	2.269**	4.637**	—	
	(80.3%)	(70.9%)	(92.4%)		
5. Jhargram	2.389**	1.028 NS	5.437**	1.033*	
	(83.3%)		(91.5%)	(69.1%)	

*Statistically significant at 95 per cent confidence level
**Statistically highly significant at 99 per cent confidence level
NS Statistically non-significant at 95 per cent confidence level.

for most of the pairs of the stations indicate statistically significant differences among themselves. For the summer data, the D^2 values can be grouped into two clusters, cluster I consisting of the data for Jhargram and Digha and cluster II consisting of the data for Durgapur-Jhalda-Debagram; Digha and Jhargram Summer data are particularly similar. For the Rainy season data, there are again two clusters, but this time Jhargram and Debagram form one cluster while Durgapur and Jhalda cluster with Digha. For the winter data, Durgapur is markedly different from the other four stations, which form two sub-clusters—Digha-Jhargram on the one hand and Jhalda-Debagram on the other.

Of the three seasons, meteorological conditions are most stable in winter and the result of multivariate study for winter data bring out the distinctly worse pollution picture for the industrial Durgapur with respect to the other stations; out of the other four stations, Digha and Jhargram have generally lower pollution levels than Jhalda and Debagram. The effects of this fall-out from the factory emissions from Durgapur-Asansol belt and the extensive coal burning in Raniganj coalfield appear to be responsible for the relatively higher pollution levels at Jhalda and Debagram than at the southern stations of Digha and Jhargram. In summer, the effect of relatively strong winds tends to disperse the pollutants from Durgapur-Asansol region, so that Durgapur pollution level comes closer to those at Jhalda and Debagram and these three stations form a cluster which is distinct from the Jhargram-Digha cluster.

In the rainy season, the pollution level at all the stations is appreciably lower (Table I), obviously due to the scrubbing effect of the rainfall. In particular, the polution level at Durgapur is brought down most, while this lowering is least in the case of Digha, so much so that Durgapur and Digha data for this season show statistically non-significant difference. On the other hand, the relatively heavy rainfall at Debagram lowers the pollution level there more than at Jhalda and Jhargram, so that Debagram and Jhargram come closer while Jhalda pollution level becomes more or less similar to those at Durgapur and Digha.

Effective discrimination between the pollution levels for all possible pairs of five stations is possible through use of the linear discriminant function, except of course in those cases where the multivariate distance between the two stations is statistically non-significant. As will be evident from Table IV, the percentage of misfits of the respective LDF's varies from 7.6 per cent to 30.9 per cent.

D. *Factor Analysis*

Varimax-rotated R-mode Factor Analysis based on 4 pollution variables (SUSMT, SO_2, NO_x, OXID) and 4 meteorologic variables (MAXT, MINT, RLHUM and WNDSP) indicates for all the stations, 5 Factors (each with high loadings on one or more variables) which account for 82–89 per cent of the total variance (Table IV). As already mentioned, each of these factors may be representing an independent source of variation in the data matrix and these may give some clue to genetic processes (cf. Harman, 1967).

TABLE IV

Varimax-rotated factor loadings matrices

Durgapur (94 data sets)

Factors Variables	I	II	III	IV	V
SO_2	-0.0631	0.1208	0.9785*	-0.0026	0.0468
NO_x	-0.4610	0.7275*	0.1268	0.1081	0.0976
OXID	0.1748	-0.0492	0.0024	-0.9762*	-0.0282
SUSMT	0.2735	0.7950*	0.2120	-0.1769	-0.1000
MAXT	0.9130*	0.0723	0.1095	-0.1771	-0.1317
MINT	0.8854*	-0.1239	-0.1993	-0.0737	-0.1722
RLHUM	0.0656	-0.8803*	0.0394	-0.1239	-0.0733
WNDSP	0.2194	-0.0392	-0.0482	-0.0280	-0.9717*
Variance (λ_i)	1.9918	1.9753	1.0741	1.0502	1.0191
					$\Sigma\lambda_i = 7.1105$

Jhalda (90 data sets)

Factors Variables	I	II	III	IV	V
SO_2	-0.0197	-0.8692*	0.1840	0.0968	-0.0655
NO_x	-0.0833	-0.7032*	-0.2122	0.2926	0.1134
OXID	0.1243	0.1381	0.0717	0.0984	0.9695*
SUSMT	-0.7815*	-0.3744	-0.1008	-0.0306	-0.0234
MAXT	-0.4481	0.1409	0.0433	-0.8374*	-0.0311
MINT	0.0761	0.2166	0.0445	-0.9391*	-0.1045
RLHUM	0.8454*	0.1499	0.1600	0.1334	0.1390
WNDSP	-0.0585	0.0208	0.9672*	-0.0567	0.0670
Variance (λ_i)	1.5582	1.4989	1.0591	1.7098	0.9933
					$\Sigma\lambda_i = 6.8193$

Debagram (54 data sets)

Factors Variables	I	II	III	IV	V
SO_2	-0.1293	-0.0239	-0.1018	0.9221*	0.0257
NO_x	-0.0885	0.1671	0.8598*	-0.1194	-0.0339
OXID	0.0924	0.9063*	0.1005	-0.0304	-0.0988
SUSMT	0.1982	-0.4611	0.5783*	0.0978	0.4057
MAXT	-0.8250*	-0.0734	0.2467	0.1587	-0.2085
MINT	-0.8970*	0.0565	-0.1442	0.0566	-0.1999
RLHUM	0.1973	0.0844	0.0286	-0.0211	0.9234
WNDSP	-0.6103*	-0.3173	-0.0860	-0.4965	0.2410
Variance(λ_i)	1.9690	1.1789	1.1840	1.1504	1.1703
					$\Sigma\lambda_i = 6.6526$

Jhargram (100 data sets)

Factors Variables	I	II	III	IV	V
SO_2	−0.0565	−0.0072	0.9171*	−0.0278	0.0098
NO_x	0.1942	0.0629	0.0059	−0.0055	0.9707*
OXID	0.1086	0.0789	0.4764	0.7209*	−0.0138
SUSMT	0.2676	0.8263*	−0.0431	0 0189	−0.0189
MAXT	−0.8734*	−0.0121	−0.0008	0.0797	−0.1912
MINT	−0.9415*	0.0166	0.0322	0.0721	−0.0566
RLHUM	0.2999	−0.8047*	−0.0595	0.0167	−0.1014
WNDSP	−0.2507	−0.0545	−0.2593	0.8110*	−0.0003
Variance (λ_i)	1.9264	1.3442	1.1417	1.1904	1.9930
					$\Sigma\lambda_i = 6.5957$

Digha (113 data sets)

Factors Variables	I	II	III	IV	V
SO_2	−0.1137	−0.9039	−0.0517	−0.0966	0.1031
NO_x	0.4296	−0.7862*	−0.02448	0.0235	0.0863
OXID	0.2697	−0.1829	−0.1308	−0.7928*	−0.1441
SUSMT	0.1646	−0.1645	0.0398	−0.0585	0.9010*
MAXT	−0.9216*	0.0049	−0.1023	0.1748	−0.0564
MINT	−0.8815*	0.1432	−0.1599	0·1386	−0.1734
RLHUM	−0.0545	−0.1061	0.2299	0.7148*	−0.3911
WNDSP	−0.0296	0.0611	0.9667*	0.0279	0.0222
Variance(λ_i)	1.9276	1.5312	1.0454	1.2033	1.0373
					$\Sigma\lambda_i = 6.7448$

The most important and consistent Factor at all the stations is the "Temperature Factor," in which MAXT and MINT have high loadings with the same sign, which means that a process in which MAXT and MINT play dominant roles is represented by this Factor, At Debagram this Factor includes also WNDSP.

SUSMT (with oppositely signed high loading on RLHUM) constitutes the second-ranked Factor at three stations (Durgapur, Jhalda and Jhargram). At Durgapur, this Factor has high loading also on NO_x and at Digha, SUSMT by itself constitutes an individual Factor. At Debagram, SUSMT—NO_x and RLHUM figure in separate Factors. The Factor in which SUSMT figures with high loading appears to represent a process of derivation of materials blown along the ground surface or close to the ground surfece; proportion of such materials is negatively influenced by relative humidity of the atmosphere. Association of NO_x with this Factor at Debagram and Durgapur may be interpreted to mean that at least a part of the NO_x there is derived by blowing out of the decomposed fertilisers in the heavily fertilised fields.

SO_2 constitutes the dominant constituent of another Factor either singly or jointly with NO_x (as at Jhalda and Digha). This Factor is best interpreted as the relatively long-distance migration of gaseous matter at medium to high levels in the atmosphere.

Total oxidants (OXID) have high loadings in separate Factors at Durgapur. Jhalda and Debagram suggesting that the process of generation and migration of the oxidants is distinct from those of the other pollutants. At the two southern stations, Jhargram and Digha, however, OXID has high loadings in the same Factor along with WNDSP and RLHUM respectively. The processes of formation and distribution of oxidant in the atmosphere thus appear to be different from those of other pollutants.

Wind speed (WNDSP) has high loadings in individual Factors at Durgapur, Jhalda and Digha, while at Debagram it is associated with the TEMPR Factor and at Jhargram it is associated with the OXID Factor. Wind speed thus appears to represent an independent cause affecting migration and distribution of the pollutants.

Summarising, at least five distinct processes (or causes) appear to have controlled the distribution and migration of the pollutants. These are, (*a*) the temperature factor, (*b*) a process of blowing of ground-derived materials, (*c*) a process of long distance air-borne migration of gaseous matter at medium to high levels in the atmosphere, (*d*) wind speed, and (*e*) an unknown process which has controlled the formation and migration of the oxidants.

The Varimax-rotated Factor Scores

A study of the Varimax-rotated Factor scores for the individual diurnal data for the five stations has revealed some interesting results.

At *Durgapur*, the two most important Varimax-rotated Factors are the temperature Factor (F-1) and another Factor (F-2) which has high positive loadings on SUSMT and NO_x and negative loading on RLHUM. These two Factors together constitute over 49.5 per cent of the total variance. On plotting the Factor scores for F-1 against those for F-2 (Fig. 3A) it is found that the plots for the three seasons lie in distinct fields, which reflect distinctly lower values for F-2 in rainy season than in summer and winter. Another interesting point is that there are distinct linear trends in the Factor scores for winter and rainy seasons which indicate negative correlation between the scores for the two Factors. In the F-3 (which has high loading on SO_2 alone) *vs.* F-2 Factor score diagram (Fig. 3B) a moderate tendency of the SO_2 Factor to increase with decrease of the other Factor is evident for both winter and rainy season data. This suggests that in these two seasons the process controlling SO_2 concentration and that controlling SUSMT, NO_x and RLHUM are negatively correlated.

At *Jhalda*, the scores for the TEMPR Factor show clear demarcation between the seasons. This is best seen in the F-1 (SUSMT, RLHUM) vs F-4 (TEMPR) Factor score diagram (Fig. 3C). Some of the other Factor score diagrams not

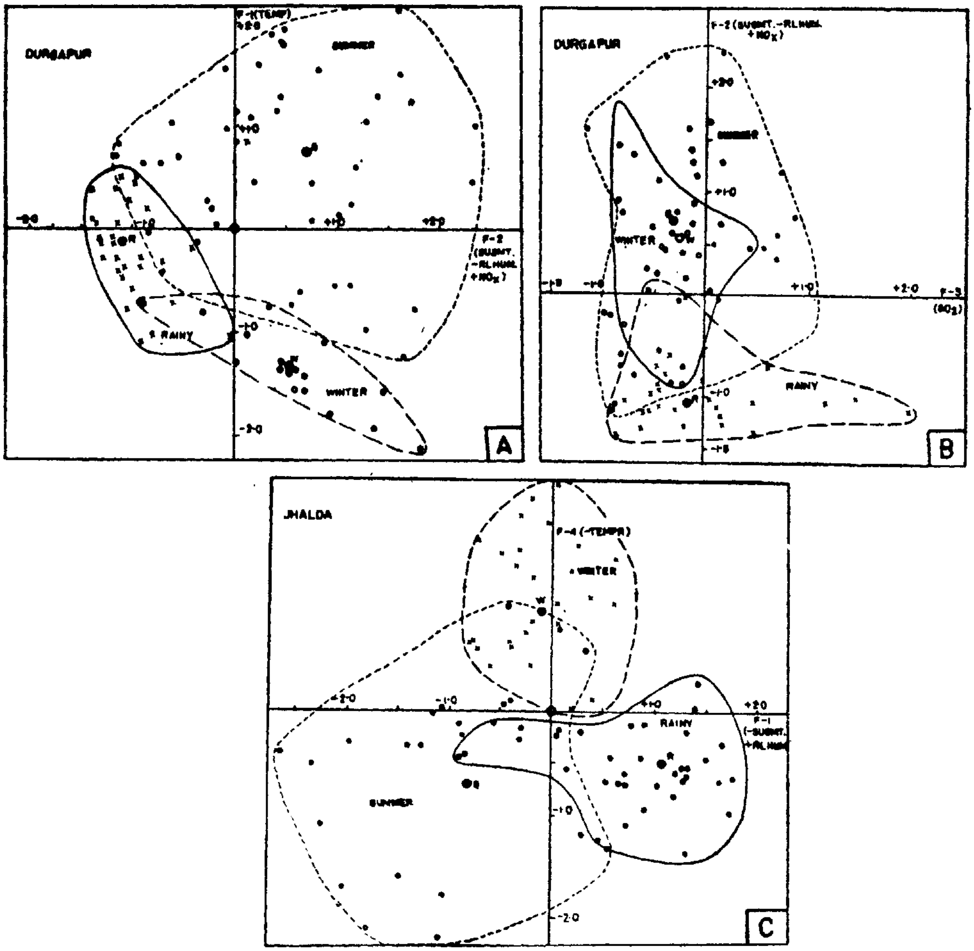

FIG. 3. Factor score diagrams based on Varimax-rotated Factor analysis of the 24-hour data on pollution and meteorologic parameters. A. Durgapur, F-1 *vs.* F-2; B. Durgapur, F-2 *vs.* F-3; C. Jhalda, F-1 *vs.* F-4.

involving TEMPR also show good discrimination between the seasons. For example, the F-1 (SUSMT, RLHUM) *vs.* F-2 (SO_2, NO_x) diagram shows very good discrimination between the plots for rainy season and those for the other seasons, with some overlap between the winter and the summer plots.

The structures of the Varimax-rotated Factors for *Debagram* are appreciably different from those of the other stations in that here NO_x and SUSMT come in one Factor, WNDSP joins with TEMPR Factor while RLHUM constitutes a single Factor. The clustering of NO_x and SUSMT in one Factor is possibly indicative of the derivation of much of the NO_x from the grounds, i.e., the heavily fertilised fields. The F-1 (TEMPR, WNDSP) *vs.* F-3 (NO_x, SUSMT) Factor score diagram (Fig. 4A) shows clear discriminations between the seasons, a negative linear trend for the summer plots and a positive linear trend for the rainy season plots. This suggests that in summer the main pollutants (NO_x, SUSMT) decrease with increasing temperature while in rainy season the pollution level increases with temperature.

At *Jhargram* the two most important Factors are F–1 (MAXT, MINT) and F–2 (SUSMT, RLHM). The corresponding Factor score diagram (Fig. 4B) shows very good discrimination between winter and summer: the rainy season data overlap largely on those for the summer. However, the appreciably lower SUSMT and higher RLHUM in rainy season than in summer are reflected in the distinctive clustering of the plots for the two seasons.

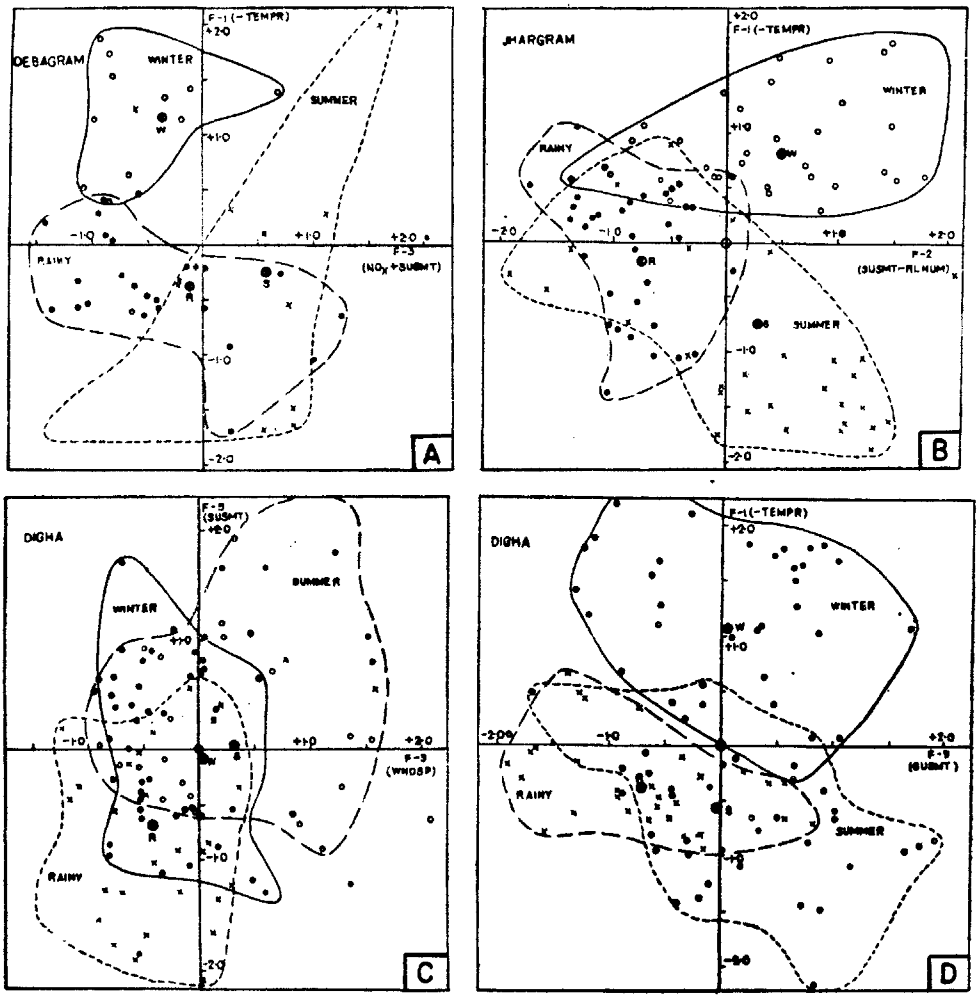

FIG. 4. Factor score diagrams based on Varimax-rotated Factor analysis of the 24-hour data on pollution and meteorologic parameters. A. Debagram, F-1 *vs.* F-3; B. Jhargram, F-1 *vs.* F-2; C. Digha, F-3 *vs.* F-5; D. Digha, F-1 *vs.* F-5.

At *Digha* the effect of WNDSP on SUSMT is clearly indicated in the F-3 (WNDSP) *vs.* F-5 (SUSMT) Factor score diagram (Fig. 4C). Corresponding to the higher windspeed in summer there is distinctly higher SUSMT while in rainy season both windspeed and SUSMT are low. However, because of interaction with other Factors, the variations of SUSMT with WNDSP are neither linear nor regular. As in the case of the other stations, there is good discrimination between summer and winter on F-1 (TEMP) *vs.* F-5 (SUSMT) Factor score diagram (Fig. 4D), while the rainy season plots overlap partly on to the summer plots. An interesting point is that on this diagram the plots for the individual seasons show elongated trends which suggest increasing SUSMT with increasing temperature.

CONCLUSIONS

Computer processing of air pollution data for 5 sampling stations in southern Bengal along with the associated meteorological variables brings out several interesting results, most significant of which are as follows :—

1. Suspended particulate matter tends to decrease with increasing relative humidity, while the gaseous pollutants decrease with increasing wind speed.

2. In most cases it is possible to set up for individual seasons multiple regression equations to estimate at moderate to high level of confidence, the individual pollution parameters in terms of the four meteorological variables, viz., maximum and minimum temperatures, relative humidity and wind speed, predictability being the highest for SUSMT, followed successively by NO_x, SO_2 and OXID.

3. The multivariate distances based on seven pollution parameters for most of the pairs of sampling stations are statisticaly significant and effective discrimination between such pairs of stations is possible through use of linear discriminant functions. Cluster analysis of the D^2 values indicates that in Summer and Winter, Durgapur has distinctly higher pollution level than the rest, Debagram and Jhalda cluster together while Jhargram and Digha form the third cluster with distinctly lower pollution level. The scrubbing effect of the rains during the rainy season brings down appreciably the pollution levels at all the stations.

4. Varimax-rotated *R*-mode Factor analysis with pollution and meteorological variables suggests that at least five distinct processes (or causes) have affected the distribution and migration of the pollutants, viz., (*a*) temperature, (*b*) wind speed, (*c*) a process of low-level *migration* of ground-derived materials. (*d*) a process of long-distance *migration* of airborne gaseous matter at medium-high levels of the atmosphere, and (*e*) an unknown factor controlling *migration* of the total oxidants. Clear discrimination between the seasonal data for individual stations is possible using some of the Factor score diagrams.

ACKNOWLEDGEMENTS

Sincere thanks are due to Dr S. N. Banerjee of Regional Computer Centre, Calcutta and Shri Basudev Mukherjee of Presidency College for assistance in the computations.

REFERENCES

Davis, J. C. (1973) *Statistics and Data Analysis in Geology.* John Wiley,New York.

Harman, H. H. (1967) *Modern Factor Analysis* (2nd edn.). Univ. Chicago Press, Chicago.

Indian Standards Institution (1969–75) Indian standard methods for measurement of air pollution. *IS* 51 82., Parts I, II, IV, V, IX and XII.

Saha, A. K., Dasgupta, S. P., Mukhopadhyay, A., Biswas, A. B., Chatterjee, H., Sen T., Thamayya, A., Chatterjee, B. D., and Chanda, S. (1980) On some problems of atmospheric pollution in southern Bengal. Unpubl. Rep., Dept. of Science & Technology, New Delhi.

Wolleben, J. A., Pauken, R. J., and Dearien, J. A. (1968) FORTRAN IV Program for multivariate paleontologic analysis using on IBM System/360 Model 40 Computer *Kansas Geol Surv. Computer Contr.*, **20**. 1–12.