return to contents page

Receptor Modelling of Air Quality Data Using Neural Networks

By John P. Grubert, [1]

Introduction

According to Hopke (1991) the management of ambient air quality is a difficult but important problem. In general, it involves the identification of the sources of materials emitted into the air, the estimation of the emission rates of the pollutants, the understanding of the transport-diffusion processes involved and the physical and chemical changes that can take place during the transport. All of these elements can be expressed in a mathematical model in order to compute the observable changes that take place in the airborne concentrations if various actions are taken. These actions could include new sources due to new industries or emission controls on existing facilities in order to reduce pollution.

However, the atmosphere is very complex and although the mathematical dispersion models used are complicated and require very fast and large storage capacity computers, these models are still insufficient to permit the development of useful air quality management systems. Thus it is necessary to have alternative methods available to assist in the identification of sources. Such methods are called receptor models since they are focused on the behaviour of the ambient environment at the point of impact as opposed to the source-oriented dispersion models.

Receptor models are generally of two types:
(i) regression analyses known as chemical mass balance,
(ii) multivariate techniques such as factor analysis, and cluster analysis.
Chemical mass balance as described by Kowalczyk et al. (1978), needs a prior knowledge of the number of sources. For example, in a study of a Chicago district, aerosols were split into six components: automobile emissions, cement dust, soil dust, emissions from coal-fired plants, emissions from oil-fired plants and steel industry emissions.

According to the literature multivariate techniques are able to determine the number, nature and mass concentrations of the particle source in a region without prior knowledge of the sources. According to Hopke (1981) of the multivariate statistical techniques that have been used as source-receptor models, factor analysis is the most widely employed. The basic objectives of factor analysis is to allow the variation within a set of data to determine the number of independent factors, i.e. groups of particles.

---

1. Department of Civil Engineering, University of East London, Barking Campus, Longbridge Road, Degenham, Essex, RM8 2AS.

This paper describes an original and alternative statistical approach using backpropagation neural networks which has the same objectives as factor analysis but utilises a completely different technique. The method is therefore valuable as an independent solution for comparison purposes and unlike factor analysis requires no knowledge of statistics.

Receptor modelling using neural networks was applied to air quality data collected from the ambient aerosol in a district of Birmingham in the West Midlands. Sixty-nine fine fraction (dia. < 2.5 μm) and seventy coarse fraction (2.5 μm < dia. < 11 μm) aerosol samples were used in the study, the object of which was to identify and quantify sources of pollution.

For this study the data set consisted of the concentrations (ng/m^3) of a large number of metals and anionic species, plus meteorological data such as relative humidity (%), temperature (^{o}C), wind speed (km/hr), solar radiation (W/m^2) and daily rainfall (mm). This data was taken over 70 sampling days during the winter of 1990 by Lukana and Harrison (1991), for the fine and coarse data respectively.

The results of the neural network analysis were compared with the factor analysis results obtained by Wormald (1994) using the same data sets. Wormald's dissertation verified the results of an earlier factor analysis study by Birmingham University. However, the meteorological data was used only in the neural network analysis and not in the factor analysis study. Also, unlike the neural analysis, it was not possible in the factor analysis method used by Wormald to find the mass concentrations of any of the elements only their matrix loading factors.

Neural Network Methodology

The data consists of twenty-one elements for the coarse fraction samples, twenty-three elements for the fine fraction samples plus in each case five relevant meteorological measurements. Concentrations of each element, expressed in ng/m^3 was measured each day for 70 days. The data showed that the order of magnitude and range of concentrations for each element varied widely. For example, silicon (Si) ranged from 0 to 26.1 ng/m^3 whereas sulphur (S) ranged from 395 to 3818 ng/m^3. This difference in the value and range of the concentrations causes difficulties since in order to compare and find relationships between the elements it is necessary that each element's concentration varies over approximately the same range. In factor analysis the data can be standardised in a variety of ways, one method of which is to subtract the variable from its mean and divide by its standard deviation in order to create a transformed value for the variable. In neural networks this transformation is done automatically since every variable can only be represented from 0 to 1. However, unlike most factor analysis transformations which are linear, modern backpropagation networks usually use a sigmoidal function to transform the data.

Because neural networks have not been used before for this type of application, it was not initially clear how to apply them. For example, with this type of database there is no obvious output or answer or dependent variable resulting from a set of inputs or independent variables, as is the case in most backpropagation applications. For this type of problem we need to identify groups of elements which are associated with each other because they represent specific sources of pollution. It should be realised that in a complicated aerosol such as Birmingham one specific group of elements can represent several different sources of pollution. We do not know the sources so we cannot start with the answers and then train the network to tell us which sources are polluting the atmosphere on a particular day. Once we know the sources this could well be another useful application. Logically, the problem suggests that we use an unsupervised associative type of neural network which is presented with the data and then self organises or categorises it into groups or sources. This type of complicated neural architecture does exist, Kohonen is an example, but backpropagation can also be used in an associative way. Also, since over 95% of all neural networks used world-wide are of the backpropagation type it is more useful to demonstrate how they can be used to solve this type of application than to experiment with network architectures that are highly specialised and unfamiliar to most engineers and scientists.

Backpropagation Procedure

The elements were analysed starting with the fine fraction data plus meteorological conditions, obtaining their groups, then repeating the whole process without the meteorological data. The concentrations from the two sets of results were compared and averaged for each pollutant group found and any significant meteorological conditions identified. This calculation was executed twice, firstly in order to see how much the meteorological conditions affected the results and secondly to have the results from two independent neural calculations. Neural networks never produce the same results twice even with identical data and network parameters, and since we have no answers to check the results with it is important to repeat the procedure. The same method was used using the coarse fraction data, giving four different calculations in total.

Because a supervised type of network is being used, the only sensible option open to us is to use one variable at a time for the input and observe its effects on the remaining variables in the output section. Also, it will be assumed that if an increase in input A results in an increase in output B, then there is only a significant group relationship between them if likewise an increase in input B produces an increase in output A.

When using the BrainMaker neural network the steps involved in the calculation are:
(1) Load the data from the database into the neural shell. Since we have no answers for checking purposes use all the data available for training. Shuffle the data twice.
(2) Identify one variable as the INPUT, the remainder, with the exception of the collection day which is not used, become the PATTERN or output.
(3) Set the screen display to THERMOMETERS and the training tolerance parameter to 0.40. Use the default values for the other parameters i.e. learning rate constant equal to 1.0, with one hidden layer containing the same number of neurons as in the output layer.
(4) Train the network and save it when as many of the 69 or 70 input/output data pairs are matched or recorded as GOOD. Since the data set is not ideal, for some input elements, this may be as low as 30 and will probably train within 200 iterations or runs. With this type of data nothing is gained by letting the network run for more than 200 runs, in fact the training could become worse.

(5) Edit the network input of the trained network by setting this single element to its maximum value, i.e. 8/8 units of concentration. (The thermometers, which replace the actual numerical values of concentration, are split into eight parts). Note the resulting units of concentration for all the elements in the output section plus the meteorological values. Now set the input concentration to zero, i.e. 0/8 units, and record the output values. Calculate and record the difference between the two values for each element and meteorological variable.
(6) Repeat sections (1) to (5) above for all the remaining elements and meteorological variables. Also, when recording the data assume that the single input variable used is also an output.

When using the fine data set with meteorological variables, this procedure will result in training 28 different neural networks, 23 for the atmospheric elements and 5 for the meteorological variables. Finally, a neural network can be built using published data on composition profiles of a number of sources in order to identify the possible origins of the groups of elements found in this study.

Neural Network Results for the Fine Fraction Data

In order to have two independent solutions, the procedures for finding groups of elements from the fine data was executed twice. Firstly, the data on element concentrations was trained together with its meteorological data, then secondly trained without the meteorological data. As mentioned previously, since we have no method of checking whether the neural networks are training the data correctly, it is important to train at least two sets of networks. Only if the results are similar can we be sure that the network training is not producing random results. Also, the meteorological data should not significantly influence the composition of the groups of elements found.

Fine Fraction Data with Meteorological Variables

Table 1 shows the values obtained after steps 5 and 6 of the Backpropagation Procedure are executed, the meteorological data are not shown here. For example when Cl set to a thermometer value of 8 (the maximum value), the resulting thermometer outputs are: P = 7, S = 7, K = 8, Ca = 3, Si = 8, etc. When the input for Cl is set to 0, the resulting thermometer outputs are: P = 3, S = 4, K = 3, Ca = 3, Si = 0, etc. The differences between the values are thus: input Cl = (8-0) = 8, and outputs P = (7-3) = 4, S = (7-4) = 3, K = (8-3) = 5, Ca = (3-3) = 0, Si = (8-0) = 8, etc. These differences are shown in Table 1 on the third row for input Cl.

Table 2 shows the values in Table 1 after neglecting thermometer values for the elements less than two. The accuracy of these procedures is not very high and it does not help to retain thermometer differences which are not significant. Together with these values are their reciprocal element values shown in brackets. For example, when nickel (Ni) is used as input, the resulting phosphorus (P) output difference is 4.

OUT IN	**P**	**S**	**Cl**	**K**	**Ca**	**Si**	**Ti**	**V**	**Cr**	**Mn**	**Fe**	**Ni**	**Cu**	**Zn**	**As**	**Se**	**Br**	**Pb**	**Na**	**NH_4**	**NO_3**	**SO_4**	**C**
P	8	1	3	0	0	2	2	1	0	1	4	5	-1	-1	-3	2	-2	1	0	-2	-1	-3	0
S	3	8	1	0	0	0	-1	4	1	-1	3	2	-1	-2	1	0	-2	0	1	5	1	7	3
Cl	4	3	8	5	0	8	6	4	3	4	3	7	1	0	1	4	-2	4	5	-4	-3	0	2
K	3	-2	3	8	3	2	5	5	1	4	5	3	2	0	2	2	3	5	3	-1	-3	-1	1
Ca	1	0	0	2	8	1	3	3	1	5	5	-4	1	0	1	-1	2	5	1	-2	-2	0	6
Si	2	-2	8	4	0	8	6	2	3	3	2	5	1	3	1	2	-2	3	3	-3	-3	-1	1
Ti	4	0	7	5	2	7	8	4	2	6	6	2	3	1	0	2	1	5	6	-3	-3	0	1
V	2	1	2	4	2	1	5	8	2	5	6	4	-1	-1	4	2	1	5	1	0	-1	3	4
Cr	-1	0	-1	1	-2	0	-1	1	8	-3	-1	-2	-2	0	1	-1	-1	-1	0	1	-1	1	-3
Mn	-1	-3	4	5	4	4	6	5	2	8	7	-2	-3	-3	5	2	5	7	5	-4	-3	2	5
Fe	-2	-2	0	6	5	0	5	7	2	8	8	-3	1	-3	2	-1	6	7	-1	-3	-2	-1	7
Ni	4	4	5	5	-2	3	0	1	-3	-3	-2	8	5	5	-3	2	0	-1	3	-3	-3	-3	-3
Cu	2	-3	4	4	1	3	-1	-1	3	-1	-2	-2	8	1	4	2	-2	-2	5	-4	-3	-2	-3
Zn	2	-4	3	-2	-3	4	-2	-2	-2	-3	-3	2	-1	8	-1	5	-2	-2	-2	-4	-3	-2	-2
As	-2	1	0	3	1	0	1	3	0	3	3	0	0	-1	8	0	-1	4	2	0	-2	1	4
Se	3	1	3	4	-2	3	-2	1	2	-2	-2	3	0	4	-2	8	1	-2	0	-3	-3	-1	-2
Br	-1	1	0	3	2	0	6	4	1	3	6	-2	-1	-1	0	-1	8	7	-2	-1	-3	-1	2
Pb	0	-3	0	5	4	0	3	4	1	7	6	-1	0	-2	4	0	7	8	5	-3	-2	-3	6
Na	2	-2	5	3	1	5	1	0	1	2	1	3	4	1	2	2	0	3	8	-2	-2	-3	0
NH_4	0	2	-4	1	-3	-3	-3	2	0	-1	-2	-2	-4	-3	1	0	-3	-3	0	8	5	4	1
NO_3	1	-2	-3	-1	-3	-3	-3	-3	-2	-2	-3	-2	-2	-3	-1	-1	-3	-3	-3	5	8	2	-3
SO_4	1	8	-1	2	-2	-1	-2	2	2	-2	-1	1	-1	-1	1	0	-2	-2	-1	6	4	8	3
C	-2	-1	0	3	4	0	3	5	0	3	6	0	-3	-2	1	1	5	6	-2	-3	-3	2	8
rh	1	1	0	-1	-3	-1	-2	1	0	-1	-3	2	-2	1	0	1	1	-3	-3	2	3	2	-1
temp	0	0	0	-1	1	0	-2	-3	-1	-2	-2	0	2	1	-1	0	-3	-3	1	-2	-2	-2	0
wind	-1	-3	-1	-1	-1	-1	-4	-4	-2	-4	-4	-1	0	1	0	0	-5	-5	1	-3	-1	-1	-4
sr	-2	-3	-2	-2	-1	-2	-1	0	-2	-2	-1	-2	-2	-2	-2	-2	0	-2	0	-1	-3	-1	-1
rain	-3	-2	-3	-2	0	-3	-2	-2	0	-1	-2	-2	-1	0	0	-3	-2	-3	2	-2	-2	1	-2

TABLE 1. Thermometer Changes in Fine Data Output Values Due to Maximum Input Changes

OUT IN	P	S	Cl	K	Ca	Si	Ti	V	Cr	Mn	Fe	Ni	Cu	Zn	As	Se	Br	Pb	Na	NH4	NO3	SO4	C
P	8 (8)	(3)	3 (4)	(3)		2 (2)	2(4)	(2)			4	5 (5)	(2)	(2)		2 (3)			(2)				
S	3	8 (8)	(3)					4			3	2(4)								5 (2)		7 (8)	3
Cl	4 (3)	3	8 (8)	5 (3)		8 (8)	6 (7)	4 (2)	3	4 (4)	3	7 (5)	(4)	(3)		4 (3)		4	5 (5)				2
K	3		3 (5)	8 (8)	3 (2)	2 (4)	5 (5)	5 (4)		4 (5)	5 (6)	3 (5)	2 (4)		2 (3)	2 (4)	3 (3)	5 (5)	3 (3)			(2)	(3)
Ca				2 (3)	8 (8)		3 (2)	3 (2)		5 (4)	5 (5)						2 (3)	5 (4)					6 (4)
Si	2 (2)		8 (8)	4 (2)		8 (8)	6 (7)	2	3	3 (4)	2	5 (3)		3 (4)		2 (3)		3	3 (5)				
Ti	4 (2)		7 (6)	5 (5)	2 (3)	7 (6)	8 (8)	4 (5)	2	6 (6)	6 (5)	2	3			2	(6)	5 (3)	6				(3)
V	2	(4)	2 (4)	4 (5)	2 (3)	(2)	5 (4)	8 (8)	2	5 (5)	6 (7)	4			4 (3)	2	(4)	5 (4)		(2)		3 (2)	4 (5)
Cr			(3)			(3)	(2)	(2)	8 (8)	(2)	(2)		(3)			(2)						(2)	
Mn			4 (4)	5 (4)	4 (5)	4 (3)	6 (6)	5 (5)	2	8 (8)	7 (8)				5 (3)	2	5 (3)	7 (7)	5 (2)			2	5 (3)
Fe	(4)	(3)	(3)	6 (5)	5 (5)	(2)	5 (6)	7 (6)	2	8 (7)	8 (8)				2 (3)		6 (6)	7 (6)					7 (6)
Ni	4 (5)	4 (2)	5 (7)	5 (3)		3 (5)	(2)	(4)				8 (8)	5	5 (2)		2 (3)			3 (3)				
Cu	2		4	4 (2)		3	(3)		3			(5)	8 (8)		4	2			5 (4)				
Zn	2		3			4 (3)						2 (5)		8 (8)		5 (4)							
As				3 (2)				3 (4)		3 (5)	3 (2)		(4)		8 (8)			4 (4)	2 (2)			(2)	4
Se	3 (2)		3 (4)	4 (2)		3 (2)	(2)	(2)	2	(2)		3 (2)	(2)	4 (5)		8 (8)			(2)				
Br				3 (3)	2 (2)		6	4		3 (5)	6 (6)						8 (8)	7 (7)					2 (5)
Pb			(4)	5 (5)	4 (5)	(3)	3 (5)	4 (5)		7 (7)	6 (7)				4 (4)		7 (7)	8 (8)	5 (3)				6 (6)
Na	2		5 (5)	3 (3)		5 (3)	(6)			2 (5)		3 (3)	4 (5)		2 (2)	2		3 (5)	8 (8)				
NH4		2 (5)						2												8 (8)	5 (5)	4 (6)	
NO3																				5 (5)	8 (8)	2 (4)	
SO4		8 (7)		2				2 (3)	2	(2)										6 (4)	4 (2)	8 (8)	3 (2)
C		(3)	(2)	3	4 (6)		3	5 (4)		3 (5)	6 (7)				(4)		5 (2)	6 (6)				2 (3)	8 (8)

TABLE 2. Significant Thermometer Changes in Fine Data Output Values with Reciprocal Changes in Brackets

Input P	**Input S**	**Input Cl**	**Input K**	**Input Ca**	**Input Si**	**Input Ti**	**Input V**
P = 8	S = 8	P = 4	Cl = 3	K = 2	P = 2	P = 4	Cl = 2
Cl = 3	Ni = 2	Cl = 8	K = 8	Ca = 8	Cl = 8	Cl = 7	K = 4
Si = 2	NH4 = 5	K = 5	Ca = 3	Ti = 3	K = 4	K = 5	Ca = 2
Ti = 2	SO4 = 7	Si = 8	Si = 2	V = 3	Si = 8	Ca = 2	Ti = 5
Ni = 5	sr = -3	Ti = 6	Ti = 5	Mn = 5	Ti = 6	Si = 7	V = 8
Se = 2	*SOURCE 3*	V = 4	V = 5	Fe = 5	Mn = 3	Ti = 8	Mn = 5
sr = -3	**Row 2**	Mn = 4	Mn = 4	Br = 2	Ni = 5	V = 4	Fe = 6
SOURCE 2		Ni = 7	Fe = 5	Pb = 5	Zn = 3	Mn = 6	As = 4
Row 1	**Input Mn**	Se = 4	Ni = 3	C = 6	Se = 2	Fe = 6	Pb = 5
	Cl = 4	Na = 5	Cu = 2	rh = -2	Na = 3	Pb = 5	SO4 = 3
Input Cr	K = 5	sr = -3	As = 2	*SOURCE 1*	sr = -3	temp = -2	C = 4
Cr = 8	Ca = 4	rain -3	Se = 2	**Row 5**	rain = -3	wind = -3	temp = -5
wind = -2	Si = 4	*SOURCE 2*	Br = 3		*SOURCE 2*	rain = -2	wind = -2
sr = -3	Ti = 6	**Row 3**	Pb = 5	**Input Cu**	**Row 6**	*SOURCE 1*	*SOURCE 1*
Row 9	V = 5		Na = 3	K = 4		**Row 7**	**Row 8**
	Mn = 8	**Input Fe**	sr = -2	Cu = 8	**Input Zn**		
Input Br	Fe = 7	K = 6	*SOURCE 1*	Na = 5	Si = 4	**Input As**	**Input Se**
K = 3	As = 5	Ca = 5	**Row 4**	rh = -4	Ni = 2	K = 3	P = 3
Ca = 2	Br = 5	Ti = 5		temp = 3	Zn = 8	V = 3	Cl = 3
Mn = 3	Pb = 7	V = 7	**Input Ni**	sr = -3	Se = 5	Mn = 3	K = 4
Fe = 6	Na = 5	Mn = 8	P = 4	*SOURCE 2*	sr =-3	Fe = 3	Si = 3
Br = 8	C = 5	Fe = 8	S = 4	**Row 13**	*SOURCE 2*	As = 8	Ni = 3
Pb = 7	temp = -4	As = 2	Cl = 5		**Row 14**	Pb = 4	Zn = 4
C = 2	wind = -4	Br = 6	K = 5	**Input NO3**		Na = 2	Se = 8
temp = -5	sr = -3	Pb = 7	Si = 3	NH4 = 5	**Input SO4**	*SOURCE 1*	rain = -3
wind = -6	*SOURCE 1*	C = 7	Ni = 8	NO3 = 8	S = 8	**Row 15**	*SOURCE 2*
rain = -2	**Row 10**	rh = -4	Zn = 5	SO4 = 2	V = 2		**Row 16**
SOURCE 1		temp = -3	Se = 2	rh = 2	NH4 = 6	**Input C**	
Row 17	**Input Pb**	wind = -3	Na = 3	temp = -3	NO3 = 4	Ca = 4	
	K = 5	*SOURCE 1*	sr = -3	sr = -3	SO4 = 8	V = 5	
	Ca = 4	**Row 11**	rain = -3	rain = -3	C = 3	Mn = 3	
	Ti = 3		*SOURCE 2*	*SOURCE 3*	*SOURCE 3*	Fe = 6	
	V = 4	**Input Na**	**Row 12**	**Row 21**	**Row 22**	Br = 5	
	Mn = 7	Cl = 5				Pb = 6	
	Fe = 6	K = 3	**Input NH4**			SO4 = 2	
	As = 4	Si = 5	S = 2			C = 8	
	Br = 7	Mn = 2	NH4 = 8			wind = -5	
	Pb = 8	Ni = 3	NO3 = 5			*SOURCE 1*	
	Na = 5	Cu = 4	SO4 = 4			**Row 23**	
	C = 6	As = 2	*SOURCE 3*				
	rh = -3	Pb = 3	**Row 20**				
	wind = -2	Na = 8					
	sr = -3	rh =-3					
	rain = -3	*SOURCE 2*					
	SOURCE 1	**Row 19**					
	Row 18						

TABLE 3. Groups or Sources for the Fine Data with Thermometer Values

However, when phosphorus is the input value the resulting nickel output difference is 5. This indicates that there is a genuine positive relationship between nickel and phosphorus, and hence when Ni is the input, P is a valid output and equal to 4 thermometer units.

Each row of Table 2 represents a group of elements, and clearly more than one row may represent the same group. The significant elements for each row are shown in Table 3. A close inspection of the elements in rows 1 to 23 indicates that only three possible groups of pollutants can be identified. Rows 4, 5, 7, 8, 10, 11, 15, 17, 18 and 23 can be identified as belonging to the first group, rows 1, 3, 6, 12, 13, 14, 16 and 19 the second group and rows 2, 20, 21 and 22 the third group. Row 9 indicates that chromium (Cr) is not associated with any other element, but that its concentration decreases slightly due to increases in wind speed and solar radiation. Group 1 pollutants are associated with high concentrations of Fe, Pb, Mn and V, and low concentrations of Ni, Cu, Zn and Na. Group 2 pollutants are associated with the opposite composition of elements to group 1, and group 3 is associated with secondary pollutants such as sulphate (SO_4), ammonium (NH_4) and nitrate (NO_3) resulting from sulphur dioxide and nitrogen oxide emissions.

In factor analysis one of the most difficult steps is to ascertain how many factors or groups are required to represent the data. About five criteria are available in order to make this choice, one of which is to use factors with variations or correlation matrix eigenvalues greater than unity. Factors with eigenvalues of one or less are no better than single variables. For the fine data the eigenvalue plot for the 23 elements, according to Wormald (1994), suggested that there is only a 3 factor solution even though there were 5 eigenvalues greater than unity.

The first group of elements and meteorological variables are shown in Table 4. Thermometer concentrations of each significant element are shown for all the rows belonging to group 1. Since all of these rows represent input elements to the same group, we need to average out the values. The simplest and best way is to total the numbers and divide by the number of rows or inputs. When this is done, Table 5 gives the thermometer concentrations of each element representing group 1. Elements with concentrations less than 1.5 units have been omitted in Table 5. Table 4 also suggests that as the wind speed and temperature both increase the concentrations of several of the elements from group 1 decrease, especially bromine (Br) and manganese (Mn).

The second group of elements and meteorological variables are shown in Table 6. The significant elements representing group 2, after neglecting concentrations less than 1.5 units, are shown in Table 7. According to Table 6, several of group 2’s elements decrease slightly in concentration especially with increasing solar radiation, and to a lesser extent increased rainfall.

The third group of elements and meteorological variables are shown in Table 8. After neglecting concentrations less than 1.5 thermometer units, group 3 consists of the elements shown in Table 9. It can be seen from Table 8 that no specific meteorological change affects the elements from this group to any significant extent.

Output Input	**P**	**Cl**	**K**	**Ca**	**Si**	**Ti**	**V**	**Mn**	**Fe**	**Ni**	**Cu**	**As**	**Se**	**Br**	**Pb**	**Na**	**SO4**	**C**	**rh**	**temp**	**wind**	**sr**	**rain**
Row4 (K)	-	3	8	3	2	5	5	4	5	3	2	2	2	3	5	3	-	-	-	-	-	-2	-
Row5 (Ca)	-	-	2	8	-	3	3	5	5	-	-	-	-	2	5	-	-	6	-2	-	-	-	-
Row7 (Ti)	4	7	5	2	7	8	4	6	6	-	-	-	-	-	5	-	-	-	-	-2	-3	-	-2
Row8 (V)	-	2	4	2	-	5	8	5	6	-	-	4	-	-	5	-	3	4	-	-5	-2	-	-
Row10 (Mn)	-	4	5	4	4	6	5	8	7	-	-	5	-	5	7	5	-	5	-	-4	-4	-3	-
Row11 (Fe)	-	-	6	5	-	5	7	8	8	-	-	2	-	6	7	-	-	7	-4	-3	-3	-	-
Row15 (As)	-	-	3	-	-	-	3	3	3	-	-	8	-	-	4	2	-	-	-	-	-	-	-
Row17 (Br)	-	-	3	2	-	-	-	3	6	-	-	-	-	8	7	-	-	2	-	-5	-6	-	-2
Row18 (Pb)	-	-	5	4	-	3	4	7	6	-	-	4	-	7	8	5	-	6	-3	-	-2	-3	-3
Row23 (C)	-	-	-	4	-	-	5	3	6	-	-	-	-	5	6	-	2	8	-	-	-5	-	-
Total	4	16	41	34	13	35	44	52	58	3	2	25	2	36	59	15	5	38	-9	-19	-25	-8	-7
Average	**0.4**	**1.6**	**4.1**	**3.4**	**1.3**	**3.5**	**4.4**	**5.2**	**5.8**	**0.3**	**0.2**	**2.5**	**0.2**	**3.6**	**5.9**	**1.5**	**0.5**	**3.8**	**-0.9**	**-1.9**	**-2.5**	**-0.8**	**-0.7**

TABLE 4. Thermometer Values for Elements Belonging to Group 1 and Their Meteorological Variables

Element	**Therm. Conc. /8**	**Element**	**Therm. Conc. /8**
Chlorine (Cl)	**1.6**	**Iron (Fe)**	**5.8**
Potassium (K)	**4.1**	**Arsenic (As)**	**2.5**
Calcium (Ca)	**3.4**	**Bromine (Br)**	**3.6**
Titanium (Ti)	**3.5**	**Lead (Pb)**	**5.9**
Vanadium (V)	**4.4**	**Sodium (Na)**	**1.5**
Manganese (Mn)	**5.2**	**Carbon (C)**	**3.8**

TABLE 5. Thermometer Concentrations for the Principal Elements Representing Group 1 (Fine Data)

Output Input	**P**	**S**	**Cl**	**K**	**Si**	**Ti**	**V**	**Mn**	**Ni**	**Cu**	**Zn**	**As**	**Se**	**Pb**	**Na**	**rh**	**temp**	**sr**	**rain**
Row1 (P)	8	-	3	-	2	2	-	-	5	-	-	-	2	-	-	-	-	-3	-
Row3 (Cl)	4	-	8	5	8	6	4	4	7	-	-	-	4	-	5	-	-	-3	-3
Row6 (Si)	2	-	8	4	8	6	-	3	-	-	3	-	2	-	3	-	-	-3	-3
Row12 (Ni)	4	4	5	5	3	-	-	-	-	-	5	-	2	-	3	-	-	-3	-3
Row13 (Cu)	-	-	-	4	-	-	-	-	8	8	-	-	-	-	5	-4	3	-3	-
Row14 (Zn)	-	-	-	-	4	-	-	-	-	-	8	-	5	-	-	-	-	-3	-
Row16 (Se)	3	-	3	4	3	-	-	-	-	-	4	-	8	-	-	-	-	-	-3
Row19 (Na)	-	-	5	3	5	-	-	2	4	4	-	2	-	3	8	-3	-	-	-
Total	21	4	32	25	33	14	4	9	33	12	20	2	23	3	24	-7	3	-18	-12
Average	**2.6**	**0.5**	**4.0**	**3.1**	**4.1**	**1.8**	**0.5**	**1.1**	**4.1**	**1.5**	**2.5**	**0.3**	**2.9**	**0.4**	**3.0**	**-0.9**	**0.4**	**-2.3**	**-1.5**

TABLE 6. Thermometer Values for Elements Belonging to Group 2 and Their Meteorological Variables

Element	**Therm. Conc. /8**	**Element**	**Therm. Conc. /8**
Phosphorus (P)	**2.6**	**Nickel (Ni)**	**4.1**
Chlorine (Cl)	**4.0**	**Copper (Cu)**	**1.5**
Potassium (K)	**3.1**	**Zinc (Zn)**	**2.5**
Silicon (Si)	**4.1**	**Selenium (Se)**	**2.9**
Titanium (Ti)	**1.8**	**Sodium (Na)**	**3.0**

TABLE 7. Thermometer Concentrations for the Principal Elements Representing Group 2 (Fine Data)

Output Input	**S**	**V**	**Ni**	**NH4**	**NO3**	**SO4**	**C**	**rh**	**temp**	**sr**	**rain**
Row2 (S)	8	-	2	5	-	7	-	-	-	-3	-
Row20 (NH4)	2	-	-	8	5	4	-	-	-	-	-
Row21 (NO3)	-	-	-	5	8	2	-	2	-3	-3	-3
Row22 (SO4)	8	2	-	6	4	8	3	-	-	-	-
Total	18	2	2	24	17	21	3	2	-3	-6	-3
Average	**4.5**	**0.5**	**0.5**	**6.0**	**4.3**	**5.3**	**0.8**	**0.5**	**-0.8**	**-1.5**	**-0.8**

TABLE 8. Thermometer Values for Elements Belonging to Group 3 and Their Meteorological Variables

Element	**Therm. Conc. /8**
Sulphur (S)	**4.5**
Ammonium (NH4)	**6.0**
Nitrate (NO3)	**4.3**
Sulphate (SO4)	**5.3**

TABLE 9. Thermometer Concentrations for the Principal Elements Representing Group 3 (Fine Data)

Averaged Fine Fraction Results

Since two independent sets of results were obtained for each of groups 1, 2 and 3, with and without meteorological data, we need to compare and average their results. Table 10 shows the concentrations for the elements in group 1, from Table 5 and Results 2 (not tabulated), and their average concentrations in thermometer units and ng/m^3. Sodium (Na) with an average concentration of less than 1.5 thermometer units has been omitted. Likewise, Table 11 shows the concentrations for the elements in group 2, from Table 7 and Results 2, and Table 12 the concentrations for the elements in group 3, from Table 9 and Results 2. The results from training the neural networks, with and without meteorological variables, are very similar. This indicates that firstly, the results can be relied upon, and secondly, that the meteorological conditions do not influence the composition of elements from each group.

Element	Results 1 TABLE 5 (/8)	Results 2 (/8)	Average Conc. (/8)	Average Conc. (ng/m^3)
Chlorine (Cl)	1.6	1.5	1.55	520
Potassium (K)	4.1	3.5	3.80	220
Calcium (Ca)	3.4	3.1	3.25	78
Titanium (Ti)	3.5	3.6	3.55	16
Vanadium (V)	4.4	3.7	4.05	15
Manganese (Mn)	5.2	4.9	5.05	170
Iron (Fe)	5.8	5.0	5.40	520
Arsenic (As)	2.5	1.8	2.15	4.8
Bromine (Br)	3.6	2.5	3.05	41
Lead (Pb)	5.9	5.7	5.80	420
Carbon (C)	3.8	4.0	3.90	3100

TABLE 10. Average Concentrations for Elements Representing Group 1 (Fine Data)

Element	Results 1 TABLE 7 (/8)	Results 2 (/8)	Average Conc. (/8)	Average Conc. (ng/m^3)
Phosphorus (P)	2.6	2.6	2.60	140
Chlorine (Cl)	4.0	3.9	3.95	1300
Potassium (K)	3.1	1.6	2.35	150
Silicon (Si)	4.1	3.5	3.80	12
Titanium (Ti)	1.8	1.9	1.85	8.9
Nickel (Ni)	4.1	4.1	4.10	20
Copper (Cu)	1.5	1.5	1.50	140
Zinc (Zn)	2.5	1.6	2.05	2700
Selenium (Se)	2.9	2.6	2.75	2.9
Sodium (Na)	3.0	3.1	3.05	430

TABLE 11. Average Concentrations for Elements Representing Group 2 (Fine Data)

Element	Results 1 TABLE 9 (/8)	Results 2 (/8)	Average Conc. (/8)	Average Conc. (ng/m^3)
Sulphur (S)	4.5	4.8	4.65	2400
Ammonium (NH4)	6.0	5.0	5.50	3700
Nitrate (NO3)	4.3	3.3	3.80	6000
Sulphate (SO4)	5.3	4.8	5.05	7900

TABLE 12. Average Concentrations for Elements Representing Group 3 (Fine Data)

Comparison of Fine Fraction Results with Factor Analysis

The results from the factor analysis study by Wormald (1994) applied to the same data, excluding the meteorological variables, is compared with the neural network results in Table 13. In factor analysis the groups are called factors. In the factor analysis elements with loadings less than 0.25 are not shown and in the neural analysis elements with concentrations less than 1.5 thermometer units are not shown. Also, the neural analysis was based on three groups whereas the factor analysis was based on five factors. Comparison of the results shows that agreement for the first three groups is very good. This is surprising considering the many stages, in both analyses, which involved subjectivity by the modeller. The very low loadings for factors four and five indicate that a three factor model, as suggested previously, would have been more appropriate. It should be noted that the loadings given by a factor analysis study cannot be used to predict concentrations, so identifying pollution sources using this method is far from satisfactory.

Element	**Factor 1 (Loading)**	**Group 1 (Conc. /8)**	**Factor 2 (Loading)**	**Group 2 (Conc. /8)**	**Factor 3 (Loading)**	**Group 3 (Conc. /8)**	**Factor 4 (Loading)**	**Factor 5 (Loading)**
Lead	**0.92**	**5.8**						
Iron	**0.90**	**5.4**						
Carbon	**0.85**	**3.9**						
Manganese	**0.81**	**5.1**						
Titanium	**0.70**	**3.6**	**0.51**	**1.9**			**0.33**	
Bromine	**0.69**	**3.1**						**0.31**
Potassium	**0.63**	**3.8**	**0.40**	**2.4**			**0.32**	
Calcium	**0.63**	**3.3**					**0.37**	
Vanadium	**0.61**	**4.1**	**0.37**		**0.47**			
Arsenic	**0.45**	**2.2**			**0.31**			
Zinc			**0.86**	**2.1**				
Nickel			**0.85**	**4.1**				
Chlorine	**0.31**	**1.6**	**0.84**	**4.0**			**0.29**	
Silicon	**0.29**		**0.84**	**3.8**			**0.29**	
Copper			**0.71**	**1.5**			**0.33**	
Selenium			**0.63**	**2.8**			**-0.27**	**0.39**
Phosphorus			**0.57**	**2.6**	**0.41**			**-0.39**
Ammonium					**0.92**	**5.5**		
Sulphate					**0.91**	**5.1**		
Sulphur					**0.87**	**4.7**		
Nitrate					**0.67**	**3.8**	**0.31**	
Sodium			**0.38**	**3.1**			**0.70**	
Chromium								**0.85**

Note: In factor analysis loadings less than 0.25 are not shown.
In neural analysis thermometer concentrations less than 1.5 are not shown.

TABLE 13. Comparison of Results with Factor Analysis for the Fine Data

Identification of Air Pollution Sources in the Birmingham Region

The principal reason for doing a factor analysis or neural network study on measured aerosol data is to identify possible sources of pollution from the factors or groups of elements found. Several references are available to help in the interpretation of factors giving information on source composition, usually concentrations obtained close to the source. For example, Hopke (1985), and Kowalczyk et al. (1978).

Source Identification using Neural Network Concentrations

The neural network analyses, applied to the fine and coarse fraction data, identified various groups of elements with their concentrations expressed in thermometer units and ng/m^3. The relative concentrations of these elements is the key to source identification since, according to Houck (1991), each source of pollution has a unique "fingerprint". Table 14 shows the possible sources of pollution which could contribute to the Birmingham aerosol, together with their source concentrations expressed in mg/g. This data was taken from the "Selected Source Profiles" appendix of the book by Hopke (1985). Average values were used where more than one set of concentrations were given for the same source.

The possible sources of pollution were selected on the basis of known industries or emissions in the Birmingham region, but the list could be enlarged to make it more comprehensive. Since Birmingham is well known for its metal industries, it is reasonable to include sources such as steel furnace, powder metallurgy, aluminium furnace and copper smelter. Also, there are coal and oil fired power plants in Birmingham plus many oil fired furnaces. Road dust, car emissions and diesel truck emissions are definitely present plus road salt (marine) since the data was collected during the winter months. There is also in Birmingham the Brierley Hill and Chance's glassworks, the Dunlop Tyre and Rubber Company, the Rugby Portland Cement works near Southam south-east of Birmingham, and asphalt plants. Also, there are incinerators belonging to refuse, sludge and car recycling plants. Finally, a rural soil source has been added since the City is surrounded by market gardens.

With the source profiles entered into a database, the next step is to train a neural network so that it can learn the source profile of each pollutant. The neural shell is loaded with the database containing the information shown in Table 14. This information consists of eighteen sources each identified by the mass concentrations in mg/g of twenty one different elements. After shuffling the data twice the neural network was defined as having 21 input neurons, one for each element, and 18 output neurons, one for each source. With one hidden layer containing 21 neurons, a learning rate equal to 1.0 and a training tolerance constant equal to 0.10, the network was trained and took 308 runs for it to match all the input/output pairs.

ELEMENT SOURCE		S	Cl	K	Ca	Si	Ti	V	Cr	Mn	Fe	Ni	Cu	Zn	As	Se	Br	Pb	Na	NO_3	C
Rural Soil	0	0.28	0.10	16.2	10.6	290	7.6	0.22	0.18	0.13	55.3	0.04	0.09	0.15	0	0	0.05	0.23	8.9	0	2.7
Road Dust	4.2	5.7	2.0	20	33	220	5.5	0.3	0.3	1.0	61	0.13	0.25	2.4	0.3	0	0.10	3.6	2.2	0	21
Refuse Incinerator	1.0	32	200	10	35	24	3.6	0.02	0.87	0.41	7.1	0.17	1.5	4.9	0.31	0.04	0.92	3.0	2.5	40	36
Asphalt Plant	0	1.5	0.1	8.5	26	259	8.3	0.28	0.73	1.9	55	0.09	0.9	2.2	0	0	0.14	0.15	16.5	1.0	9.4
Cement Works	59	0	0	20	207	10	0.5	0	0.5	0.5	5.5	0.5	0.5	0.5	0	0.5	0	0.5	0	5.5	0
Coal Power Plant	0	200	11	10	23	5	2.6	0.13	0.4	0.31	18	0.18	1.3	3.0	0.26	0	0.16	2.0	9.7	0	0
Glass Manufacture	0	57	6.8	22	35	11.5	0	0.11	0.19	0.37	13	0.55	0	14	2.5	7.5	0.05	13	50	0	0
Oil Power Plant	0	0	37	1.3	250	0	0.08	70	0.18	0.3	8.4	12	2.5	4.9	0.08	0.1	0.16	1.2	37	0	0
Oil Furnace	0	150	0.57	0.63	1.3	0	0.27	1.4	0.56	0.28	11	2.4	2.8	0.8	0.12	0.04	0.02	2.1	2.9	5	0
Steel Furnace	0	84	42	15	22	0	0.92	0.06	3.0	40	270	3.7	5.3	210	0.18	0.05	0.28	52	20	0	0
Aluminium Furnace	0	180	30	42	3.9	2.2	0.58	0.32	0.51	0.15	15	4.6	3.6	7.5	0.09	0	0.20	6.0	30	5.5	0
Copper Smelter	0	50	0	47	11.7	0	3.6	0.07	0.07	0.19	53	0	5.5	0.54	0	0.01	0	0.05	17	0	0
Tyre Manufacture	0	26	13	7.6	3.6	78	0.24	0.10	0.07	0.32	1.5	0.10	0.37	0.07	0	0	0.14	0.1	1.9	8	395
Car Recycle Plant	0	2.4	47	1.4	30	4.4	0.3	0	1.0	0.43	47	0.16	5.5	15	0	0	44	2.5	0.9	0	23
Powder Metallurgy	0	240	32	0	6.3	0	0.53	0.02	1.2	0.1	8.9	0	0	0.43	0	0.04	0	3.6	13	0	0
Car Emissions	0	11.1	8.8	0.58	2.1	0	0.45	0	0.15	0.05	3.3	0.1	0.29	0.27	0	0	14.4	73	0	9.0	31
Diesel Trucks	0	7.6	17	0.11	12.7	3.0	2.0	0.1	0.2	0.18	15.3	0.2	5.2	1.5	0	0	0.3	1.0	5.5	7.2	460
Marine	0	12.7	150	13	10.9	49	1.5	0	0.55	0.33	7.6	0	0	0.68	0	0	0.75	0.02	109	37.5	0

TABLE 14. Probable Sources of Pollution Together with Their Elements Source Concentrations in mg/g

With the network trained the edit input facility could be entered. By setting each of the 21 elements concentrations to zero, the output section displayed a zero unit of association for every source. This is to be expected, since if there is zero concentration for each element, there should not be any sources indicated by the model. By entering the concentrations in thermometer units for the elements representing group 1 of the fine fraction data shown in Table 13, the output section showed the level of association or source strengths for the possible sources. The inputs entered, to the nearest whole number are: Pb = 5.8, Fe = 5.4, C = 3.9, Mn = 5.1, Ti = 3.6, Br = 3.1, K = 3.8, Ca = 3.3, V = 4.1, As = 2.2 and Cl = 1.6. The results show that possible sources are steel furnace (association strength = 1), car emissions (association strength = 4), and road dust (association strength = 7), there are no other sources.

The source profiles being used have element concentrations measured in mg/g and were recorded within a few tens of metres of the source. On the other hand the receptor model recorded the concentrations of the elements, in ng/m^3 of air, several hundreds or thousands of metres from the sources. This huge difference in concentration for each element, makes it impossible to use the receptor concentrations themselves for direct comparison with the source concentrations. However, by using thermometer unit concentrations, at least the maximum receptor and source concentrations for each element are assigned the same value, that is eight thermometer units. Therefore, by using thermometer units in the neural edit facility rather than numbers in ng/m^3, it is possible to extract meaningful results from the neural networks. Clearly this process is not perfect, especially since some sources have similar profiles, so it is necessary to repeat each neural network using different network parameters in order to obtain alternative interpretations of the same data.

For each group of elements found from the fine fraction data, their source strengths in thermometer units, i.e. /8, were found for each of eight different networks a, b, c, d, e, f, g and h trained. For group 1 it was deduced that the main source is road dust, followed by car emissions and oil power plant etc. For group 3 elements, both fine and coarse fractions, there are secondary pollutants such as nitrate (NO_3), which originates from nitrogen oxide emissions mainly from vehicles, sulphate (SO_4) originating from sulphur dioxide emissions as a result of oil and coal combustion, and ammonium (NH_4) from fertiliser plants or livestock.

Finally, a summation of all the source strengths, both fine and coarse fractions showed that air pollution originating from motor vehicles and oil combustion are the two dominant sources of pollution in Birmingham. These are followed, to a much lesser extent, by tyre manufacture, cement production, glass manufacture, road salt, steel furnaces, coal power plants and powder metallurgy. The reasons why pollution from the iron and steel industry was much less in the early nineties than one might have expected is partly due to better emission controls but primarily due to the closure of many steelworks around Birmingham belonging to British Steel.

Discussion of Results

The neural network computations used to find groups of elements from the fine and coarse fraction data produced results that were not too dissimilar from the factors obtained by factor analysis. This suggests that this new methodology is very useful for receptor modelling especially as it gives the elements concentrations rather than the less useful loadings produced by factor analysis. One of the problems associated with the receptor modelling of a complicated aerosol, which has nothing to do with the method of calculation used, is the fact that each group of elements or factors found, can have several sources of pollution. Because of this another analysis has to be made in order to compute the relative strengths of these sources. The previous section illustrated how a neural network solution could be used to solve this problem, which previously had been addressed in a qualitative rather than a quantitative manner.

Overall, the primary purpose of receptor modelling is to identify sources of pollution so that steps can be taken to improve the ambient air quality. From the knowledge gained by analysing the Birmingham aerosol data, it can be concluded that if the region’s air quality is to be improved then firstly, not only automobile emissions need to be reduced, but also the number of vehicles on the roads. Secondly, industries and plants using oil-fired boilers, furnaces etc., which are in abundance in the Birmingham region, need stricter emission controls.

Acknowledgements

The writer is registered for a Ph.D. degree in Environmental Sciences at the University of Greenwich. The support of Professor Bernard Fisher of the University of Greenwich and Professor Shafik Al Naib of the University of East London is appreciated.

APPENDIX I - References

Hopke, P.K. (1981). “The application of factor analysis to urban aerosol source resolution.” *Atmospheric Aerosol*, American Chemical Society Symposium Series 167, 21-49.

Hopke, P.K. (1985). *Receptor Modeling in Environmental Chemistry*. John Wiley & Sons, New York.

Hopke, P.K. (1991). “An introduction to receptor modeling.” in *Receptor Modeling for Air Quality Management*. ed. P.E. Hopke, Elsevier, 1-8, Amsterdam.

Houck, J.E. (1991). “Source sampling for receptor modeling.” in *Receptor Modeling for Air Quality Management*. ed. P.E. Hopke, Elsevier, 45-78, Amsterdam.

Kowalczyk, G.S., Choquette, C.E., and Gordon, G.E. (1978). “Chemical element balances and identification of air pollution sources in Washington, D.C.” *Atmospheric Environment*, 12, 1143-1153.

Lukana, L., and Harrison, R. (1991). Communications to Dr. R. Sokhi, University of Greenwich, Deptford Campus.

Wormald, N.B. (1994). “Use of receptor modelling in urban aerosol source resolution.” M.Sc. dissertation in Scientific and Engineering Software Technology, University of Greenwich.