APPLICATION AND EVALUATION OF FOUR REGRESSION TECHNIQUES FOR A CHEMICAL MASS BALANCE RECEPTOR MODEL

By

Michael I. Holzman
Dames & Moore
1550 Northwest Highway
Park Ridge, Illinois 60068

Peter A. Scheff
Pritzker Department of Environmental Engineering
Illinois Institute of Technology
Chicago, Illinois 60616

Mohd Rashid Mohd Yusoff
Department of Chemical Engineering
University of Technology Malaysia
Kuala Lumpur, Malaysia

APPLICATION AND EVALUATION OF FOUR REGRESSION TECHNIQUES FOR A CHEMICAL MASS BALANCE RECEPTOR MODEL

INTRODUCTION

Chemical mass balance (CMB) receptor models have evolved over the past 15 years as a potential alternative to dispersion models for assessing the source contributions of pollutants and aerosols in the atmosphere. Unlike dispersion models, which require a detailed inventory of emission rates and stack parameters from major sources in addition to meteorological data and empirical dispersion factors, receptor models need only information about the characteristics of samples collected at a site and the chemical composition of source categories.

There are many advantages and potential applications of receptor models to contemporary air pollution problems. Their relatively simple mathematics compared to source-oriented dispersion models results in a less time- and cost-intensive method of source apportionment. Fugitive and area source contributions to ambient aerosol samples can be predicted without the need to develop emission factors. A major application for receptor models is in the area of criteria pollutant standards attainment, where they can be used to determine the major contributing sources to regional air pollutant levels. State implementation plans can then be created to regulate those sources.

Source impacts are estimated with CMB receptor models through application of different regression techniques to solve simple mass balance equations. Variations of ordinary least squares regression, both unbiased and biased techniques, have been used in source apportionment studies in major urban airsheds across the country. Unbiased techniques include weighted least squares (WLS) and effective variance weighted least squares (EVWLS) regression. Biased techniques that have been considered include ridge regression (RR), principal components, and latent root regression. Studies have also been published to directly intercompare the different methods.[1,2] For the purposes of this study, four different solutions to the CMB receptor model have been developed and evaluated for an environmental data set: two unbiased techniques (WLS and EVWLS) and two biased techniques (RR weighted by the measurement variance of the receptor data and RR weighted by the effective variance). These four solutions were then evaluated and intercompared through statistical analysis and physical validation techniques.

MODEL DEVELOPMENT

A simple mass balance forms the basis of all CMB receptor models for particulate source apportionment. In matrix form, the CMB equation is:

$$C = AS \qquad [1]$$

where C is the n x 1 vector of observed elemental concentrations, A is the n x p matrix of source compositions, and S is the p x 1 vector of source contributions. The equation can be solved for source contribution vector S using least squares statistics.

Weighted Least Squares (WLS) Regression

WLS is a transformation of the ordinary least squares regression technique in which the most probable values of the source impact coefficients, S_j, are predicted by minimizing the sum of squares of the errors weighted by the inverse of the measurement variance.[3] This transformation has the effect of giving equal importance to each tracer element in the solution as well as allowing the less precise measurements, defined by large measurement variance, to have less influence than more precise measurements.[4] The solution is written in matrix form as:

$$S = (A^tWA)^{-1}A^tWC \qquad [2]$$

where W is a diagonal matrix with inverse measurement variance on the diagonal, superscript "t" is the matrix transpose, and superscript "-1" is the matrix inverse.

Effective Variance Weighted Least Squares (EVWLS)

The EVWLS procedure has been applied to the CMB equation.[5,6] In this variation of the least squares solution, the weighting term is the effective variance, defined as:

$$V_e = V_c + \sum_j V_{A_j} S^2{}_j \qquad [3]$$

where V_c is the measurement variance of vector C, V_{A_j} is the variance in the emissions from source A_j, and S_j is the contribution from source j. Substituting V_e for the measurement variance in equation [2] gives the effective variance solution. Since the EVWLS solution depends on the values of the source contributions, S_j, an iterative procedure is required. The major advantage of adding the uncertainties associated with both of the measured input variables is that better estimates of the values of S_j and the standard error of S_j may be obtained given measurements of V_A.[7]

Ridge Regression (RR)

One assumption of the least squares solution to the CMB model that is not always satisfied is that the source compositions are linearly independent of each other. In statistical terms, this is referred to as the problem of multicollinearity in the source profile matrix. When one source signature is nearly a linear combination of any subset of the other signatures, it can result in unstable parameter estimates, indicated by large standard errors in the estimation of S_j, estimates that are larger than physically reasonable, failure to resolve sources with similar profiles, or meaningless negative source coefficients.[8,9,10]

Hoerl and Kennard first suggested the ridge regression procedure of biased estimation for problems affected by multicollinearities.[11] The ridge estimates for the source coefficients, S_j, are defined as a function of a ridge parameter, k, which is introduced into the least squares solution:

$$S_j(k) = (A^tWA + kI)^{-1}A^tWC \qquad [4]$$

where I is the identity matrix and A, C, and W have been defined previously. The weighting factor can be either the inverse of the receptor data variance or the total effective variance. When $k = 0$, the above equation reduces to the ordinary WLS or EVWLS solution, depending on the choice of W. If $k > 0$, a bias or systematic error is introduced into the regression coefficients with the objective of trading a large reduction in the estimates' variance for a small amount of bias. For the biased solution, the total variance or mean square error (MSE) is the cumulative effect of the variance and the bias:

$$\text{MSE} = \text{Variance} + (\text{Bias})^2 \qquad [5]$$

Since the variance is a decreasing function of k and the bias term is an increasing function of k, the MSE of the estimator or regression coefficient will decrease to a minimum and then increase with increasing bias (k). The objective, then, is to find a value of k that gives a stable set of coefficients with smaller MSE than the WLS solution, while minimizing the amount of bias introduced.

METHODOLOGY

The data set used in evaluating the four regression techniques contained samples of atmospheric inhalable particulate matter collected over a 4-month period during the late spring and summer of 1984. Ambient 24-hour average samples were collected on 6-day intervals from May 18 to August 22 at a sampling site located approximately 2 miles south of downtown Chicago. The elemental analysis was performed using non-destructive wavelength dispersive X-ray fluorescence spectrometry at the Illinois Institute of Technology Research Institute (IITRI). The concentrations of 19 elemental components were determined with this technique by calibrating the spectrometer with thin film standards prepared at different concentrations for each component of interest. In this manner, an estimate of the measurement error was calculated from the regression between concentration of standard and instrument response, and these were used directly in all CMB calculations.

Based on previous studies done in the Chicago area and other similar urban environments, seven sources were identified and used in the modeling.[3,12,13,14] These were coal combustion, limestone and cement, mobile, refuse incineration, soil, iron and steel industry, and grain handling sources. Most of the source signatures (a_{ij}) were developed from the Receptor Model Source Composition Library, a compilation of source profile data that is rated according to its judged usefulness for receptor modeling.[15] A more complete description of the sampling procedure, elemental analysis, and source profile development can be found in the theses of Yusoff[12] and Holzman.[16]

An interactive FORTRAN computer program developed under contract for the U.S. Environmental Protection Agency (EPA) was used to apply all four regression techniques to the CMB model.[17] To find the "best fit" regression model for each day of data, the model was first run with all seven sources. Any negative source coefficients were then removed one at a time in order of decreasing variance inflation factors until the final "best" model contained all positive source coefficients. For the ridge regression solution, the ridge parameter, k, was selected by an automatic procedure that selects an optimum k value by optimizing the agreement between the observed and calculated aerosol concentrations.

The four regression solutions to the CMB equation were intercompared through evaluation of statistical parameters and by physical validation techniques utilizing emission inventories and meteorology.

RESULTS

Table I summarizes the results of the four solutions to the CMB equation. The mean source coefficients are the average of the regression coefficients for each of the 17 days for each regression technique. The root mean square standard errors (a measure of average statistical uncertainty) are also presented, along with the percentage of total particulate concentration predicted by each of the modeled sources. The unexplained particulate is the difference in particulate concentration between the total measured and that apportioned by the models.

DISCUSSION

Comparison of Methods on Initial Runs

The solutions of the four regression techniques for all seven sources on three selected days are shown in Table II. These days were chosen to represent a range of extreme problems in the data that cause the ordinary WLS solution to produce from zero to four negative source coefficients. On the first day (7/11/84), the WLS solution results in four negative source coefficients. Variance inflation factors for both the soil and grain sources are extremely high, indicating multicollinearity problems. The EVWLS solution also gives four negative coefficients in this case. However, the mobile source is now negative instead of the steel source in the WLS solution. Variance inflation factors are about half the magnitudes of the values in the WLS solution, and standard errors of the coefficients from the EVWLS solution are less. Both ridge regression methods introduce a large bias into the solution to reduce the amount of negative coefficients. This is evident from the relatively high ridge parameters (k values) used.

On the second day (7/5/84), the WLS solution resulted in two negative source coefficients. The EVWLS solution was very similar to WLS except for the mobile source coefficient, which was lower by about a factor of 10. Variance inflation factors were also much lower in the EVWLS solution. Both RR methods result in no negative coefficients; however, the solutions are significantly different from one another and from the unbiased least squares methods. The RR solution, weighted only with the concentration matrix standard errors, uses a very high k value of 0.72, compared with 0.14 used in the RR/EV method. A large amount of bias is, therefore, introduced into the RR solution.

On the third sample day (5/18/84), no negative sources were present in the WLS solution, and the coefficients from the WLS, EVWLS, and RR/EV solutions were all similar. The k value of 0.02 in the RR/EV solution indicated that relatively little bias was used, and the solution thus practically reduces to the EVWLS solution. The standard errors of the EVWLS and RR/EV coefficients were equivalent except for the steel source in the EVWLS solution, which was much higher. The RR solution resulted in a mobile source coefficient that was much higher than the other three methods. When compared with other studies done in the Chicago area, this number seems to be unreasonably high.[3,13] In fact, the RR solution consistently predicted high mobile source coefficients throughout the study.

Comparison of Mean Source Coefficients

As a first level of comparison, the mean source coefficients (Table I) do not differ significantly for lime, refuse, soil, steel, and grain sources. However, coal and mobile sources do vary significantly between the models. The differences in the coal source coefficients may be attributable to the large number of zero coefficients resulting from multicollinearity problems in the WLS and EVWLS solutions. Yet for the mobile source, negative coefficients were not as large a problem as with the coal source. For the RR solution, the mobile source coefficient was about five times higher than any of the other model predictions. As stated above, this high value seems unrealistic when compared with other relevant studies.

A qualitative statistical comparison can be made between the mean source coefficients of the different models by comparing the standard errors of the coefficients. These were calculated by taking a root mean square average of the daily standard errors. In general, these are smaller for the EVWLS and RR/EV methods than for the RR and WLS methods. With smaller standard errors, the statistical signal to noise ratio is increased, meaning that the source coefficients can be more accurately determined with respect to predicting measured concentrations from source profile information. This does not imply, however, that the regression coefficients, when interpreted as source apportionment coefficients, are better in a physical validation sense.

Also, the number of negative or zero source coefficients can be compared. In general, both ridge regression methods result in significantly fewer zero coefficients than WLS and EVWLS. Ridge regression, thus, has better resolution of sources.

Finally, the amount of particulate concentration that is unexplained by the models can be compared. The average unexplained particulate concentration is seen to be consistent at about 15 $\mu g/m^3$ for all but the RR solution. As discussed above, the RR solution tends to reduce the amount from unexplained particulate sources by inflating the mobile source impact.

Correlation of Daily Coefficients Among Models

Comparison of the mean source coefficients of the four models can only be considered as a first level of comparison. This method reveals nothing about how the daily coefficients vary with daily changes in meteorology and emissions. A better method of comparison is to examine whether the source coefficients have the same daily variation between the models to determine whether they are indeed predicting the same system.

Table III presents a correlation matrix for each of the seven sources. In each table, the coefficients of each of the four methods are correlated against each other. Lime, soil, steel, and grain sources are all significantly correlated among all four models. All models are able to predict the same daily variation in these source impacts with a significantly high probability. For the coal source, EVWLS, RR, and RR/EV are significantly correlated with each other, but the WLS coal coefficient is not correlated with any of the other three methods' coefficients. This is most likely due to the high number of zero coal impact coefficients predicted by the WLS model. The mobile source coefficients are not correlated between the EVWLS and RR/EV and between the RR and RR/EV methods. But between WLS, EVWLS, and RR and between WLS and RR/EV, the correlation coefficients are all approximately 0.5. For the refuse source coefficient, all methods except WLS are significantly correlated.

Quality of Fit

One criterion of goodness of fit that has been used in receptor modeling studies is an observed to predicted elemental ratio in the range of 0.5 to 2.0. In general, all of the regression techniques are able to meet this criterion for more than half of the 19 elements. WLS accurately fits nine out of 19 elements, while EVWLS and RR/EV fit 10. RR was able to predict 12 of the 19 elements within a factor of two. The same elements that are predicted well with WLS are also predicted accurately with EVWLS, RR, and RR/EV methods. Ti, Co, and As are not predicted well with any of the models. This analysis, however, also does not directly evaluate the physical significance of the predictions.

Physical Validation

Both point source and area source validation techniques were used to intercompare the models. To accomplish the point source validation, an emission inventory was obtained from the EPA for all major (greater than 50 tons/year) point sources in the Chicago area for the following categories: coal burning, iron and steel industry, refuse incineration, and lime and cement industry.[18] The sources of interest were all from Cook and Lake Counties in Illinois and Lake County in Indiana. One important finding from the emission inventory was that there were no major (greater than 50 tons/year) refuse incineration sources. Because of this, the CMB models were rerun without refuse for all subsequent comparisons with the emission inventory impacts. To complete the data requirements for the point source validation, meteorological parameters were taken from National Weather Service (NWS) stations at O'Hare Airport and at the lakefront in downtown Chicago.[19]

Multiple point source dispersion models were originally considered for the receptor model validation. However, only two out of the 62 coal, steel, and lime industry sources identified in the emission inventory were less than 10 kilometers from the receptor site. Most were in the range of 20 to 30 kilometers away. A major caution on the use of any of the Gaussian dispersion models is that estimated impacts are unreliable for distances over 10 kilometers from the source.

For this reason, a different point source validation technique was applied that also makes use of the emission inventory and hourly wind direction data.[3] For each hour of the study period days, the wind direction was used to estimate the total particulate emission rate from upwind sources in each of the three point source categories. All of the plant emission stacks that had coordinates locating them within 45 degrees (±22.5 degrees) of the hourly wind direction vector were summed for each hour and each source to arrive at upwind daily source impacts. These impacts were then compared with the receptor model impacts. A 22.5-degree (±12.25 degrees) plume is consistent with the narrow plume concept of lateral dispersion of point source plumes. However, since the wind direction vectors were only known to within ±22.5 degrees, a 45-degree plume concept was used in this study.

Plots comparing the WLS, EVWLS, and RR CMB models to the emission inventory impacts for the coal and limestone sources are shown in Figures 1 and 2. For the limestone industry source category, the daily coefficients from all four solutions were significantly correlated with the emission inventory impacts. For the coal source impacts, significant relationships were found for the WLS, EVWLS, and RR/EV methods. The RR solution to the receptor model did not show the same association with the emission inventory impacts that the other methods did for the coal source. A positive correlation is seen in the plot up to a point, and then a negative association is seen. For the steel industry source category, none of the models exhibited a positive association between the receptor and emission inventory impacts. Due to the similarity of the EVWLS and RR/EV solutions, the validation plots for the RR/EV solution were virtually identical to the EVWLS plots and are, therefore, not presented here.

Area Source Validation

To evaluate the area source coefficients, the relationships between particulate concentration and wind speed were examined. Soil particulate concentrations were expected to increase with wind speed due to higher entrainment, and mobile source impacts should be diluted due to greater advection of "clean" air at higher wind speeds. All four solutions to the CMB exhibited a significant correlation (0.7 to 0.82) between soil impact coefficients and wind speed, but none were able to detect a strong negative association between mobile impacts and wind speed. The same results have been found in two previous studies.[3,12] The plots of soil impacts versus wind speed are shown in Figure 3.

CONCLUSIONS

1. In cases where the least squares regression solutions to the CMB model are affected by multicollinearity problems, the biased ridge regression techniques do result in fewer negative coefficients and smaller standard errors. However, the appropriateness of the biased solutions cannot be judged by this alone.

2. The EVWLS solution exhibits a better signal to noise ratio and fewer negative coefficients than the WLS solution.

3. The RR solution, weighted by the receptor concentration variance only, seems to give unreasonable results by more than one measure. As evidenced by high ridge parameters (k values) used in the majority of the daily solutions, a large amount of bias is introduced. Mobile source coefficients are significantly overestimated relative to the other three models used in this study and the results of other Chicago and urban studies. Relatively high steel source

impacts are estimated to occur from the northwest direction, where no major steel mills are located. Finally, the positive association between coal source impacts estimated by the receptor model and those from an emission inventory, which is seen by the WLS, EVWLS, and RR/EV models, is not seen with the RR solution.

4. The RR/EV solution, in which the weighting factor is the effective variance, does not suffer from the same problems as the RR solution. This is most likely due to the fact that much lower ridge parameters are used, and as a consequence, less bias is introduced than in the RR solution. In fact, the RR/EV solution reduces to the EVWLS solution in 6 out of 17 days, and the k value was less than 0.01 in 10 out of 17 days. There was, therefore, very little difference between the two solutions.

5. For the WLS, EVWLS, and RR/EV solutions to the CMB receptor model, the mean source impact coefficients have similar magnitudes. The RR solution overpredicts mobile and coal source impacts relative to the other three models.

6. In general, the daily source impacts are significantly correlated between the models. This indicates that the models have the same relative response to daily variations in source strength and meteorology.

7. The refuse source profile has not been characterized accurately enough for any of the models to predict minor impacts when no major sources exist. Severe overprediction resulted.

8. Both EVWLS and RR/EV predicted lower standard errors for the source coefficients than the WLS solution. This does not imply that the EVWLS and RR/EV coefficients are better estimates of the physical system, since the daily source coefficients are generally consistent among the three models.

9. While the statistical evaluation of the models suggests an advantage for the effective variance solutions when compared with the weighted least squares solutions, this advantage was not seen when the physical predictions were evaluated. In fact, the physical evaluation of the results shows a clearer linear relationship between wind speed and CMB soil and between CMB coal and point source coal impact for WLS than EVWLS solutions.

REFERENCES

1. Watson, J.G. 1984. A Comparison of Results from Source and Receptor Model Evaluation and Validation Studies. Presented at the 77th Air Poll. Control Assoc. Meeting, San Francisco. In Transcript.

2. Dzubay, T.G., and Stevens, R.K. 1982. Intercomparison of Results of Several Receptor Models for Apportioning Houston Aerosol. In Receptor Models Applied to Contemporary Pollution Problems. Air Pollution Control Assoc., Pub. SP-48, pp. 60-71.

3. Scheff, P.A., Wadden, R.A., and Allen, R.J. 1984. Development and Validation of a Chemical Mass Balance for Chicago. Envir. Sci. and Technol. 18:923-931.

4. Watson, J.G. 1982. Overview of Receptor Model Principles. In Receptor Models Applied to Contemporary Pollution Problems. Air Poll. Contr. Assoc. Pub. SP-48, pp. 6-16.

5. Britt, H.I., and Luecke, R.H. 1973. The Estimation of Parameters in Non-Linear, Implicit Models. Technometrics 15:233-247.

6. Watson, J.G., Cooper, J.A., and Huntzicker, J.J. 1984. The Effective Variance Weighting for Least Squares Calculations Applied to the Mass Balance Receptor Model. Atmos. Envir. 18:7, pp. 1347-1355.

7. Henry, R.C., and Hidy, G.M. 1979. Multivariate Analysis of Particulate Sulfate and Other Air Quality Variables by Principal Components. Atmos. Envir. 13:1581-1596.

8. Draper, N.R., and Smith, H. 1981. Applied Regression Analysis. 2nd edition, John Wiley & Sons, New York.

9. Marquardt, D.W., and Snee, R.D. 1975. Ridge Regression in Practice. The American Statistician 29:5-20.

10. Williamson, H.J., Balfour, W.F., and Schmidt, C.P. 1982. Source Apportionment through Weighted Ridge Regression. Radian Corporation, Austin, TX.

11. Hoerl, A.E., and Kennard, R.W. 1970. Ridge Regression: Biased Estimation for Non-Orthogonal Problems. Technometrics 12:55-67.

12. Yusoff, M.R.M. 1984. Elemental Composition and Sources of Aerosol in Chicago by X-Ray Fluorescence and Chemical Element Balances. M.S. Thesis, Illinois Institute of Technology.

13. Gatz, D.F. 1975. Relative Contributions of Different Sources of Urban Aerosols: Applications of a New Estimation Method to Multiple Sites in Chicago. Atmos. Envir. 9:1-18.

14. Kowalczyk, G.S., Choquette, C.E., and Gordon, G.E. 1978. Chemical Element Balances and Identification of Air Pollution Sources in Washington, D.C. Atmos. Envir. 12:1143-1153.

15. Core, J.E., Shah, J.J., and Cooper, J.A. 1984. Receptor Model Source Composition Library. EPA-450/4-85-002. U.S. Environmental Protection Agency.

16. Holzman, M.I. 1985. Evaluation of Four Chemical Mass Balance Receptor Models for the Source Apportionment of Chicago Aerosol. M.S. Thesis, Illinois Institute of Technology.

17. Williamson, H.J., and Dubose, D.A. 1983. User's Manual for CMB Model. EPA-450/4-83-014. USEPA.

18 U.S. Environmental Protection Agency. 1985. Emission Inventory for the Chicago Area. Computer printout.

19. National Climatic Data Center. 1984. Meteorological Data for Chicago O'Hare Airport Station. Computer printout.

Table I. Summary of Source Coefficients and Standard Errors[a]

Regression Method	Mean Source Coefficient ± Standard Error ($\mu g/m^3$) and % of Total							
	Coal	Lime	Mobile	Refuse	Soil	Steel	Grain	Unexplained[b] Particulates
WLS	0.47 ± 1.84 (0.95)	5.77 ± 0.61 (11.67)	2.41 ± 6.70 (4.87)	7.62 ± 2.28 (15.41)	8.52 ± 4.12 (17.23)	5.92 ± 1.40 (11.97)	3.00 ± 3.67 (6.07)	15.75 ± 15. (31.84)
No. of Zeros	11	1	5	0	1	2	12	
EVWLS	0.42 ± 1.25 (0.85)	6.03 ± 0.80 (12.19)	0.44 ± 2.78 (0.89)	8.61 ± 1.57 (17.41)	7.77 ± 4.17 (15.71)	6.79 ± 6.89 (13.73)	3.18 ± 3.11 (6.43)	16.22 ± 15. (32.79)
No. of Zeros	10	0	4	0	3	1	7	
RR	1.59 ± 0.76 (3.21)	4.52 ± 0.60 (9.14)	10.93 ± 4.40 (22.10)	6.31 ± 1.84 (12.76)	6.88 ± 0.82 (13.91)	5.35 ± 1.43 (10.82)	5.12 ± 3.06 (10.35)	8.76 ± 12.88 (17.71)
No. of Zeros	3	0	1	0	0	2	3	
RR/EV	0.73 ± 0.87 (1.48)	5.80 ± 0.68 (11.73)	1.65 ± 2.68 (3.33)	7.58 ± 1.47 (15.33)	6.48 ± 2.11 (13.10)	7.11 ± 1.21 (14.38)	5.80 ± 2.32 (11.73)	14.32 ± 13.09 (28.95)
No. of Zeros	6	0	3	0	2	0	0	

[a]Standard errors are root mean square average of 17 daily standard errors of the estimate S_j.

[b]Unexplained = average particulate concentration - sum of seven sources.

Table II. Comparison of Methods on Three Sample Days

	Sample Day #1 (7/11)											
	$R^2 = 0.693$			$R^2 = 0.893$			$R^2 = 0.62$			$R^2 = 0.767$		
	WLS	Standard Error	VIF[a]	EVWLS	Standard Error	VIF[a]	RR	Standard Error	K[b]	RR/EV	Standard Error	K[b]
Coal	-2.083	±2.121	6.94	-3.566	±0.978	9.35	0.241	±0.725	0.28	0.156	±0.247	0.48
Lime	-1.809	±3.284	4.44	-0.946	±0.899	2.97	0.434	±1.494		0.559	±0.586	
Mobile	4.105	±7.879	3.69	-1.348	±1.771	1.11	1.195	±3.619		0.207	±2.768	
Refuse	9.416	±2.248	2.01	12.941	±1.612	2.93	5.671	±1.423		6.186	±1.087	
Soil	10.298	±10.471	108.95	8.410	±3.461	50.69	0.682	±0.490		0.513	±0.184	
Steel	-2.111	±1.896	6.48	0.785	±0.893	1.94	-0.037	±0.651		0.660	±0.697	
Grain	-6.595	±11.611	62.06	-3.065	±3.418	21.60	0.871	±1.020		0.875	±0.373	
	Sample Day #2 (7/5)											
	$R^2 = 0.989$			$R^2 = 0.953$			$R^2 = 0.933$			$R^2 = 0.948$		
	WLS	Standard Error	VIF[a]	EVWLS	Standard Error	VIF[a]	RR	Standard Error	K[b]	RR/EV	Standard Error	K[b]
Coal	-3.999	±2.244	13.93	-3.214	±2.076	5.96	2.007	±0.643	0.72	1.953	±0.998	0.14
Lime	1.339	±1.536	32.00	1.475	±1.463	5.00	2.605	±0.348		4.717	±0.928	
Mobile	2.566	±5.829	2.78	0.288	±3.136	1.12	20.343	±4.615		2.311	±4.548	
Refuse	19.212	±2.519	1.52	14.051	±3.248	2.52	8.408	±2.988		8.501	±2.533	
Soil	26.760	±3.746	187.23	27.792	±7.381	16.15	4.264	±0.469		5.388	±0.825	
Steel	11.605	±2.093	3.03	7.061	±2.079	1.59	7.579	±1.620		7.838	±7.325	
Grain	-18.782	±8.905	73.69	-18.786	±8.178	4.64	2.951	±1.040		2.239	±1.781	
	Sample Day #3 (5/18)											
	$R^2 = 0.999$			$R^2 = 0.999$			$R^2 = 0.991$			$R^2 = 0.995$		
	WLS	Standard Error	VIF[a]	EVWLS	Standard Error	VIF[a]	RR	Standard Error	K[b]	RR/EV	Standard Error	K[b]
Coal	1.288	±1.514	12.03	1.844	±1.028	3.82	5.490	±1.163	0.2	2.372	±0.854	0.02
Lime	19.870	±1.046	6.12	20.018	±1.422	1.88	17.421	±1.040		20.064	±1.304	
Mobile	2.905	±2.980	1.23	1.683	±1.742	1.03	11.690	±6.478		1.800	±1.741	
Refuse	7.091	±0.915	1.19	7.319	±1.116	1.43	5.877	±1.973		7.009	±1.049	
Soil	21.839	±5.378	101.13	21.947	±3.810	11.38	13.695	±0.739		19.822	±2.796	
Steel	19.003	±1.386	2.41	16.311	±1.558	1.52	17.077	±2.417		16.260	±1.514	
Grain	10.751	±5.194	41.30	11.236	±4.220	4.51	15.470	±1.852		13.218	±3.310	

[a]Variance Inflation Factor.

[b]Ridge parameter.

Table III. Correlation of Coefficients Between Models

	Coal1	Coal2	Coal3	Coal4
Coal1		0.15	0.17	0.24
Coal2	0.15		0.51	0.84
Coal3	0.17	0.51		0.62
Coal4	0.24	0.84	0.62	

	Lime1	Lime2	Lime3	Lime4
Lime1		0.99	0.97	0.98
Lime2	0.99		0.98	0.99
Lime3	0.97	0.98		
Lime4	0.98	0.99	0.97	

	Mobile1	Mobile2	Mobile3	Mobile4
Mobile1		0.49	0.50	0.46
Mobile2	0.49		0.43	0.20
Mobile3	0.50	0.43		-0.13
Mobile4	0.46	0.20	-0.13	

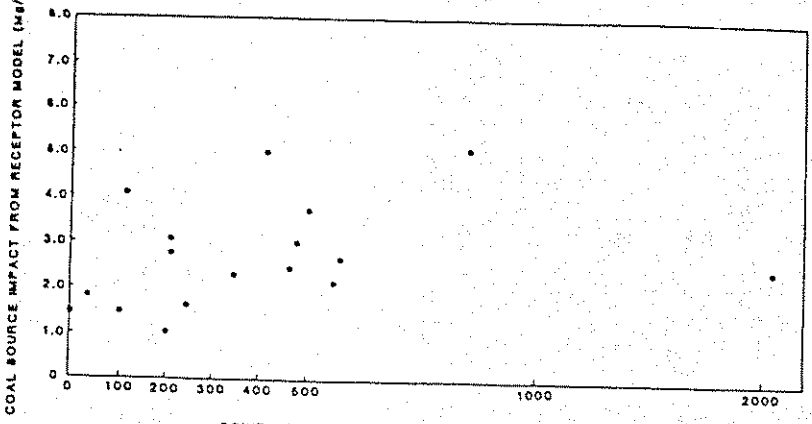
	Refuse1	Refuse2	Refuse3	Refuse4
Refuse1		0.35	0.54	0.50
Refuse2	0.35		0.73	0.90
Refuse3	0.54	0.73		0.72
Refuse4	0.50	0.90	0.72	

	Soil1	Soil2	Soil3	Soil4
Soil1		0.96	0.91	0.97
Soil2	0.96		0.89	0.99
Soil3	0.91	0.89		0.92
Soil4	0.97	0.99	0.92	

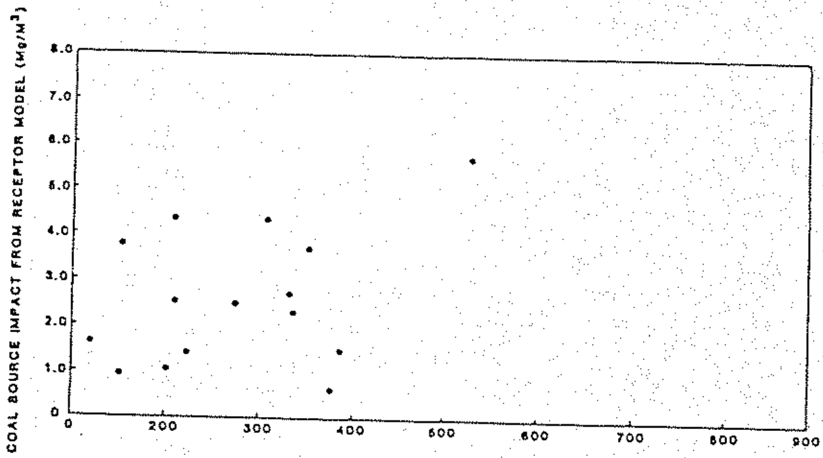
	Steel1	Steel2	Steel3	Steel4
Steel1		0.65	0.94	0.95
Steel2	0.65		0.58	0.73
Steel3	0.94	0.58		0.93
Steel4	0.95	0.73	0.93	

	Grain1	Grain2	Grain3	Grain4
Grain1		0.54	0.95	0.96
Grain2	0.54		0.54	0.63
Grain3	0.95	0.54		0.97
Grain4	0.96	0.63	0.97	

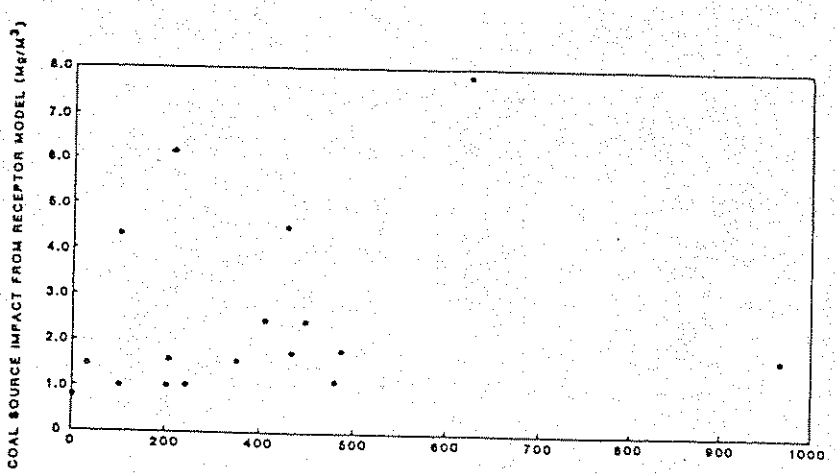
Note: 1 = WLS; 2 = EVWLS; 3 = RR; 4 = RR/EV.



VALIDATION OF WLS COAL SOURCE COEFFICIENT

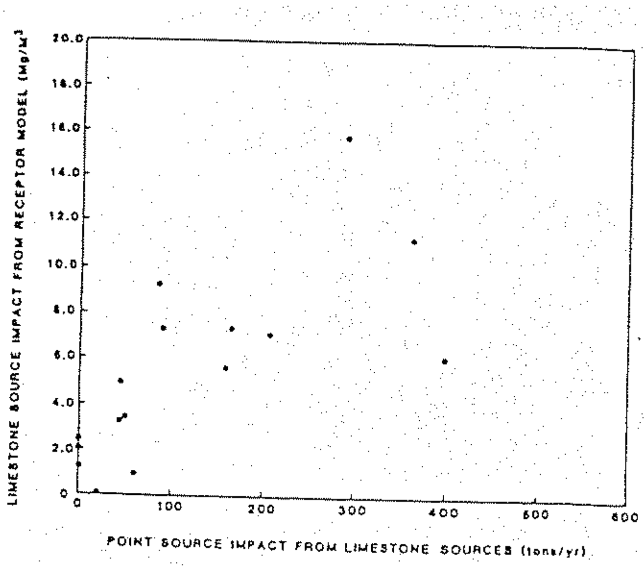


VALIDATION OF EVWLS COAL SOURCE COEFFICIENT

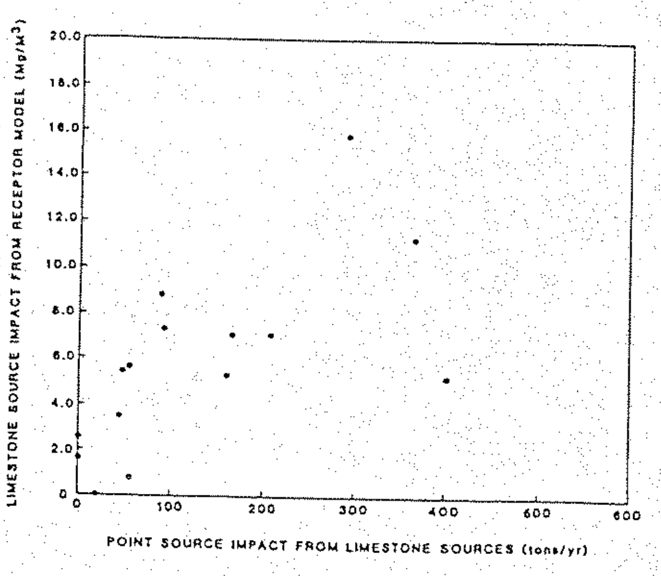


VALIDATION OF RR COAL SOURCE COEFFICIENT

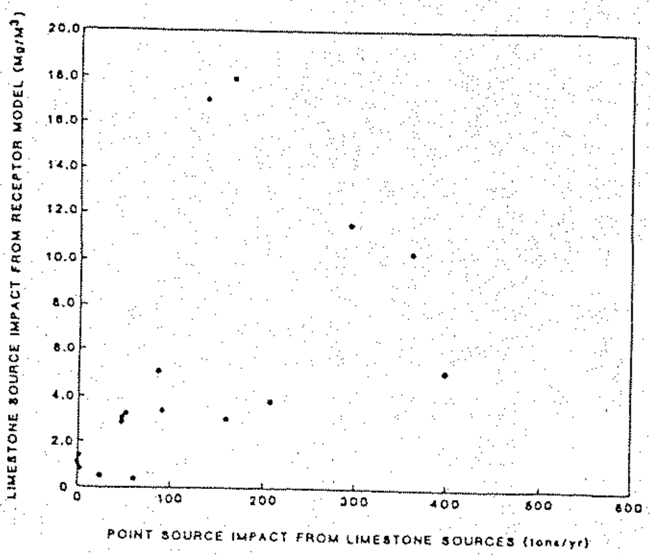
FIGURE 1 VALIDATION OF COAL SOURCE COEFFICIENT



VALIDATION OF WLS LIMESTONE SOURCE COEFFICIENT

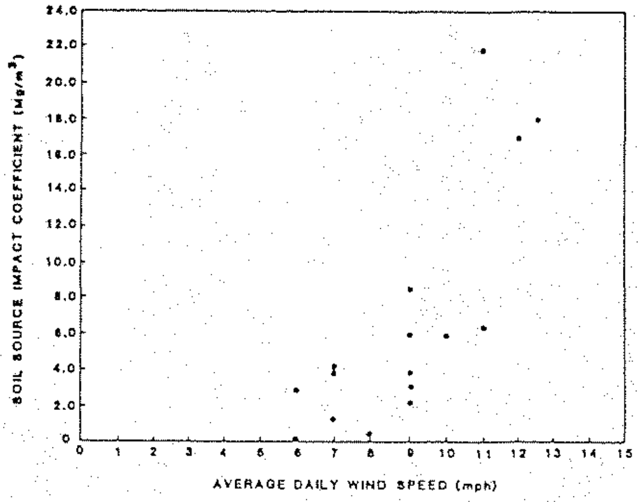


VALIDATION OF EVWLS LIMESTONE SOURCE COEFFICIENT

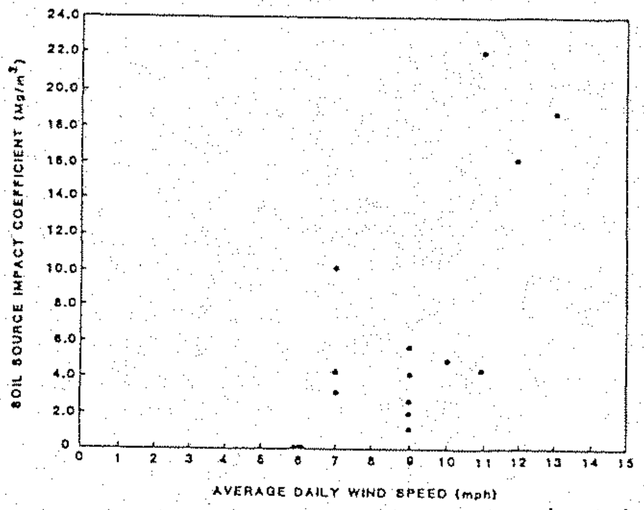


VALIDATION OF RR LIMESTONE SOURCE COEFFICIENT

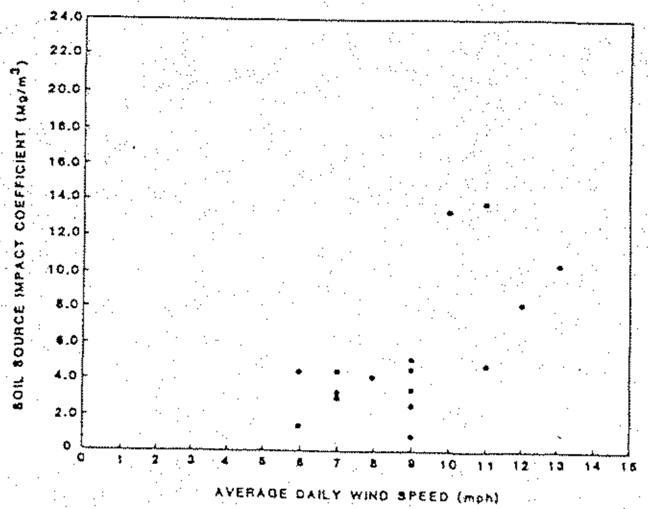
FIGURE 2 VALIDATION OF LIMESTONE COEFFICIENT



PLOT OF WLS SOIL SOURCE COEFFICIENT VERSUS WIND SPEED



PLOT OF EVWLS SOIL SOURCE COEFFICIENT VERSUS WIND SPEED



PLOT OF RR SOIL SOURCE COEFFICIENT VERSUS WIND SPEED

FIGURE 3 RELATIONSHIP BETWEEN SOIL SOURCE COEFFICIENT AND WIND SPEED