ANALYZING DIOXIN MEASUREMENT DATA IN MUNICIPAL WASTE INCINERATION USING MULTIPLE REGRESSION ANALYSIS

H.Huang and A.Buekens
Department of Chemical Engineering and Industrial Chemistry, Free University of Brussels, Pleinlaan 2, 1050 Brussels, Belgium

Introduction

The experimental study of dioxin emissions from incineration facilities usually yields a data set containing the measurements of dioxins and other process parameters. In analyzing the experimental data the objectives are usually: (1) to identify the process parameters affecting dioxin emissions, (2) to describe the general trends about how the dioxins are influenced by these parameters, and (3) to obtain the possible quantitative correlations. Statistical techniques can serve as useful tools in this data analysis process. In the past the regression method has been utilized in a number of studies on dioxin emissions. In the Environment Canada study the data are analyzed using the Pearson correlation coefficients,[1-2] and later by two-parameter models[3] and also by multiple linear models.[2,4] Recently, a model involving the interactive, square and logarithm transformation terms of process parameters has been used to correlate the dioxin emission data from a furnace reactor.[5] In this paper we use the regression method to analyze one dioxin data set for large-scale, grate-type incinerators and construct the various single-variable/multi-variable, linear/quadratic models for this data set, and then compare these models and discuss the adequacy or inadequacy, application and limitation of the method.

Multiple Regression Analysis

Multiple regression analysis is the calculation and testing of empirical relationships or regression equations between a response variable and some independent variables from an existing data set. The calculation of regression equations is usually based on the "least square" principle. Statistical parameters measuring the fitting of regression equations to a data set are the square of the multiple correlation coefficient R^2 and the F ratio. The R^2 ranges from 0 to 1. A higher value indicates better correlations. If there is only one independent variable, then the multiple correlation coefficient is simplified to the Pearson correlation coefficient, which is indicated customarily by r^2. The F ratio is used to compare with the critical F value for different size of data set. When the F ratio is larger than the critical F value, the regression equation is statistically significant. Details of computation and interpretation of multiple regression analysis can be found in some statistics handbooks.[6-8]

Descriptive Statistics of the Dioxin Data Set

The dioxin data set analyzed in this paper consists of 47 measurements at 11 large-scale, grate-type municipal waste incinerators performed in a Dutch study programme.[9] The flue gas composition and dioxin data are chosen from the original report and examined for their possible correlationships. These parameters and their descriptive statistics are shown in table 1.

Table 1. Descriptive statistics of the dioxin data set (number of measurements n=47)

	Min.	Max.	Arithmetic mean m_i	Standard deviation s_i
O_2	8.3	17.1	11.9	1.84
CO	5	3199	295.4	682.5
NO_x	227	497	352.4	69.9
SO_2	38	468	171.6	94.6
HCl	0.3	1065	387.8	336.3
F.A.	0.6	156	45.3	39.6
Dioxin	1.2	123	31.7	35.5

Note: F.A.=fly ash. The unit of O_2 is vol%, dioxin is in ng-TE/Nm^3 at 11%O_2, dry basis, all others are in mg/Nm^3 at 11%O_2, dry basis.

The Single-variable Correlations

The linear correlations between dioxin and other parameters pairwisely are of the form:

$$\text{Dioxin} = a + b\, x_i \pm e \qquad (1)$$

where, x_i are O_2, CO, NO_x, SO_2, HCl and F.A., respectively, a and b are regression coefficients, e is standard error, dioxin and x_i have the same units as in table 1. The equation coefficients estimated from regression analysis are listed in table 2.

Table 2. The single-variable linear correlations

	r^2	a	b	e	F
O_2	0.116$^+$	-46.6	6.58	33.8	5.9$^+$
CO	0.004	32.7	-0.003	35.9	0.18
NO_x	0.024	59.6	-0.079	35.5	1.1
SO_2	0.07	14.6	0.099	34.7	3.4
HCl	0.48$^+$	3.33	0.073	25.9	41.4$^+$
F.A.	0.205$^+$	13.3	0.407	32	11.6$^+$

Note: The significant statistical parameters are indicated by $^+$. The critical F value is $F_{0.95}(1,45)=4.06$.

From table 2 it can be seen that O_2, HCl and F.A. are significant variables as their F values are higher than the critical F value. Their r^2 indicate that about 48%, 20.5% and 11.6% of the variation of dioxins can be explained by the variations of HCl, F.A. and O_2, respectively. The positive signs of their regression coefficients b indicate that they are all positively correlated with dioxins.

The quadratic correlations between dioxin and other parameters pairwisely are:

$$\text{Dioxin} = a + b\, x_i^* + c\, (x_i^*)^2 \pm e \qquad (2)$$

where, x_i^*: standardized variables, $x_i^*=(x_i-m_i)/s_i$, m_i and s_i are listed in table 1. Standardized variables are customarily utilized in regression analysis involving many terms. The equation coefficients are given in table 3.

Table 3. The single-variable quadratic correlations

	R^2	a	b	c	e	F_b	F_c	F
O_2^*	0.242^+	39.3	17.7	-7.8	31.6	11.7^+	7.3^+	7.04^+
CO^*	0.004	32.5	0.16	-0.77	36.3	0	0.2	0.1
NO_x^*	0.027	30	-5.82	1.73	35.8	1.19	0.14	0.62
SO_2^*	0.188^+	40.2	22.5	-8.67	32.8	10^+	6.37^+	5.08^+
HCl^*	0.552^+	21.2	21.7	10.7	24.3	33.5^+	7.13^+	27.1^+
F.A.*	0.243^+	38.4	24.6	-6.86	31.6	11^+	2.19	7.05^+

Note: F_b and F_c are the F testing of linear and quadratic terms. F is the testing of the overall correlations. The critical values are $F_{0.95}(1,44)=4.06$ for F_b and F_c, $F_{0.95}(2,44)=3.21$ for F.

From table 3 it can be seen that when the quadratic effects are considered, O_2, SO_2, HCl and F.A. are significant variables and their R^2 indicate that about 55.2%, 24.3%, 24.2% and 18.8% of the variation of dioxins can be explained by the variations of HCl, F.A., O_2 and SO_2, respectively. The correlation between O_2^* and dioxin is: dioxin=39.3+17.7 O_2^*-7.8$(O_2^*)^2$. It gives a maximum dioxin point at O_2^*= 17.7/2X7.8=1.13. Noticing from table 1 that for variable x_i, most of the original data points fall within the interval: (m_i-s_i, m_i+s_i), and x_i^* is within (-1,1), so that the results of regression analysis is reliable only within this interval. Therefore, a maximum dioxin point at O_2^*=1.13 means that within the range of investigation, dioxins increase with the increase of O_2 and because of the quadratic effect, the increase is more rapid at low O_2 than at high O_2. Similar analyses of other correlations in table 3 show that dioxins increase with the increase of SO_2 and HCl quadratically, but with F.A. almost linearly.

<u>The Multiple Correlations</u>

The multiple linear correlation between dioxin and other parameters is of the form:

$$\text{Dioxin} = b_0 + \Sigma\, b_i x_i^* \pm e \qquad (3)$$

From regression analysis of the data set the correlation is estimated as:

$$\text{Dioxin}=31.7+3.7\, O_2^*-6.1\, CO^*-3.8\, NO_x^*+14.9\, SO_2^*+22.9\, HCl^*-0.73\, F.A.^* \qquad (4)$$

For this correlation the R^2 is 0.611, e is 23.8, the overall F is 10.5, the critical F is $F_{0.95}(6,40)=2.34$.

The multiple quadratic correlation between dioxin and other parameters with interactive terms is of the form:

$$\text{Dioxin} = b_0 + \Sigma\, b_i x_i^* + \Sigma\, b_{ii}(x_i^*)^2 + \Sigma\, b_{ij} x_i^* x_j^* \pm e \qquad (5)$$

The resulting correlation from regression analysis of the data set is:

$$\text{Dioxin}= 35.7 +10.1\, O_2^* -14.7\, (O_2^*)^2 +11.9\, CO^* -10.6\, NO_x^* +7.3\, (NO_x^*)^2 +16.2\, SO_2^* -7.6\, (SO_2^*)^2 +20.9\, HCl^* -6.3\, F.A.^* +7.9\, (F.A.^*)^2 +10.7\, O_2^*\, HCl^* +11.2\, SO_2^*\, HCl^* \qquad (6)$$

For this correlation the R^2 is 0.891, e is 13.7, the overall F is 23.1, the critical F is $F_{0.95}(12,34)=2.05$. All terms in equation (6) are significant in F testing. Other quadratic terms not present in the equation as well as two interactive terms O_2^* F.A.* and HCl* F.A.* are insignificant. The interactive terms involving CO* and NO_x^* have not been included in the regression analysis.

From equation (6) the influence of flue gas composition on dioxins can be identified as follows: (1) O_2: a maximum point exists at O_2^*=(10.1+10.7 HCl*)/ 2X14.7. When HCl* is 1, O_2^* is 0.7, but when HCl* is -1, O_2^* is 0 and O_2 is 12%. Thus very approximately, the influence of O_2 on dioxins is correlated with HCl, at high HCl dioxins increase with the increase of O_2, at low HCl a maximum dioxin point is found at 12%O_2; (2) CO: Dioxins are positively linearly correlated with CO; (3) NO_x: Dioxins decrease with the increase of NO_x; (4) SO_2: The effect of SO_2 on dioxins is correlated with HCl and is similar to O_2; (5) HCl: The effect of HCl is correlated with O_2 and SO_2. When both O_2 and SO_2 are at low level, HCl has little effect on dioxins, but when both O_2 and SO_2 are at high level, dioxins increase with the increase of HCl; (6) F.A.: A minimum point is found at F.A.*=0.4 or F.A.=60 mg/Nm^3.

Discussions

The correlations between flue gas composition and dioxins obtained above using regression method are generally in agreement with the understanding of the formation mechanisms of dioxins. The entering of such terms as O_2, HCl, F.A., SO_2 into the model is justified by the possibility of their participating in dioxin formation reactions. CO is an indicator of incomplete combustion and may also enter the model. The role of NO_x is less clear. But in previous studies NO_x is also found to be a significant parameter[2)]. The positive correlation between O_2 and dioxins can be explained by *de novo* synthesis of dioxins which is shown to be increased by a higher flue gas O_2 concentration in laboratory experiments. This observation suggests that in order to reduce dioxin formation too high a flue gas O_2 level should be avoided and the optimal O_2 level may range from 6% to 9% corresponding to an excess air ratio of less than 100%. In the following several questions regarding the data analysis method are discussed:

(1) The Pearson correlation coefficient: The r^2 listed in table 2 can be calculated without actually doing the regression analysis and are commonly used as convenient indication of pairwise correlations. The underlying assumption of the r^2 is the first order assumption over the range of investigation as the r^2 is essentially a measure of the total variance explained by a linear correlation between two parameters. If the r^2 is higher than 0.25, it means strong linear correlation. If the r^2 is less than 0.25, it means no correlation at all or strong quadratic correlation. For the present wide range of parameters studied as shown in table 1 the first order assumption may not apply as O_2 and SO_2 have significant quadratic effects as shown in table 3. Using the Pearson correlation coefficient alone cannot detect the quadratic effects and thus may sometimes lead to under-estimation of the influence of some variables.

(2) Single-variable or multi-variable analysis? The single-variable correlations given in table 2 and 3 can also be obtained by plotting the data in an X-Y coordinate and doing a curve-fitting. The advantage of regression method is that statistical parameters can be calculated and the significance of regression equations be judged objectively. The problem of such a single-variable analysis is, however, that when studying the effect of one variable in an X-Y coordinate, other variables are not held constant and the observations regarding this variable may be offset by other changing variables as dioxins are essentially influenced by many process variables simultaneously. Therefore, single-variable analysis may not produce consistent and reliable results in some cases. Good pairwise correlations

in table 2 and 3 indicate possible strong effects, but a lack of correlation may be due to the effects of other changing variables and hence no simple conclusions can be drawn.

The multiple regression equations (3) to (6) provide correlations of the influence of several variables simultaneously. The computation of these equations is based on the "least square" method and is in principle similar to fitting a straight line to some data points in a two-dimensional coordinate or fitting a plane to some data points in a three-dimensional space and may be regarded as the extension to an n-dimensional space ($n>3$). Comparing the results from single-variable and multiple correlations it can be seen that more elaborate descriptions of the relationships between dioxins and other parameters are given in the multiple correlations.

(3) Linear or quadratic model? In the Environment Canada study[2)], multiple linear model of the form of equation (3) has been used to correlate the dioxin data and the R^2 of the correlation reached 0.89. The good linear correlation suggests that within the range of study the system can be approximated well by a first-order assumption. But in the present study the multiple linear model of equation (4) has an R^2 of 0.611 only. This poor correlation indicates that first-order assumption does not hold. When the model is extended to include quadratic and interactive terms in equation (5) and (6), the R^2 reaches 0.891. This suggests that quadratic and interactive effects are important and multiple quadratic model of the form of equation (5) is necessary for good representation of the data set. A requirement for computation of regression equations (3) and (5) is that the number of experimental measurements should be at least larger than the number of terms in the regression equations. The computation of multiple quadratic model needs therefore much more experimental runs than multiple linear model.

(4) Higher order terms? Equation (6) has an R^2 of 0.891, which is sufficient for general trends analysis, but may not be sufficient for quantitative prediction purpose. If higher order terms and other transformation terms are included, the R^2 can be increased to more than 0.95 and model predictions are further improved. A model of this type is given by Gullett et.al.[5)] However, the mechanistic implications of these models are not very clear and no reliable computational algorithm for a search of suitable transformation terms is available.

(5) Model applicability: The application of empirical models based on regression analysis of some experimental data is limited to the specific system and conditions from which the experimental data are obtained. For incinerators of similar design the models may be similar qualitatively. For different types of incinerator the models may differ substantially because of the different combustion characteristics and thus need to be developed for individual cases.

Conclusions

Dioxin emissions from incineration facilities are influenced by a variety of process parameters simultaneously, and quadratic and interactive effects are important, so that multivariate statistical method is necessary for analyzing the experimental data to obtain meaningful correlations. The analysis of dioxin measurement data using various regression methods in this paper demonstrates that: (1) Some methods widely used in initial data exploration such as the calculation of the Pearson correlation coefficient and plotting and studying the data in an X-Y coordinate may not be sufficient for studying the complex relationship between dioxins and other process parameters; (2) Multiple linear model is adequate for correlating dioxin emission data only if the system can be approximated well by a first-order assumption within the range of operating conditions studied; (3) Multiple quadratic model with interactive terms can provide good correlations of dioxin emission data with reasonable model complexity, and may be used as a statistical tool in general trends analysis, developing dioxin prediction and control models and optimizing the combustion conditions to minimize dioxin emissions.

References

1. Environment Canada (1985), National Incinerator Testing and Evaluation Programme: Two-stage Combustion, Report EPS 3/UP/1, Sept. 1985
2. Environment Canada (1988), National Incinerator Testing and Evaluation Programme: Environmental Characterization of Mass Burning Incinerator Technology at Quebec City, Report EPS 3/UP/5, June 1988
3. Barton R.G., W.D. Clark, W.S. Lanier, W.R. Seeker (1990), Dioxin emissions from waste incinerators, *Chemosphere*, 20, 1825-32
4. Chang N.-B., S.-H.Huang (1994), A chemometric approach for the verification of the formation mechanism of dioxin/furan in municipal waste incinerators, *Organohalogen Compd.*, 20, 419-24
5. Gullett B.K., P.M. Lemieux, J.E. Dunn (1994), Role of combustion and sorbent parameters in prevention of polychlorinated dibenzo-p-dioxin and polychlorinated dibenzofuran formation during waste combustion, *Environ. Sci. Technol.* 28, 107-18
6. Flury B., H.Riedwyl (1988), Multivariate Statistics: a Practical Approach, Chapman & Hall, London
7. Cooke D., A.H.Craven, G.M.Clarke (1985), Statistical Computing in Pascal, Edward Arnold, London
8. Box G.E.P., N.R,Draper (1987), Empirical Model-building and Response Surfaces, Wiley, New York
9. Ministry of Housing, Planning and Environment (1993), The Combustion of Municipal Solid Waste in the Netherlands: Emission Occurring During Combustion, Dispersal of Dioxin and the Associated Risks, Report 730501052, Feb. 1993, The Netherlands