Fitting of Analytical Functions with Digital Computers in Spectrochemical Analysis

Marvin Margoshes[1] and Stanley D. Rasberry

Spectrochemical Analysis Section, National Bureau of Standards, Washington, D. C. 20234

Two procedures are described for fitting spectrometric analytical functions with a digital computer and conversion of instrument readings to element concentrations. The first procedure is intended for computers giving rapid response, and particularly those which can operate in the conversational mode. During the running of the program, the user decides on the equation to be used and whether data for any standards are to be rejected. The second procedure is intended for use on batch-loading computers; these decisions are incorporated in the program. Results are given of the use of both procedures. The occasional failures of the procedures are discussed, and methods are described for detecting such failures. A discussion is given of a method for including an estimate of the uncertainty of the analytical function in the estimated analytical error.

THERE ARE obvious advantages in using digital computers to convert spectrometer readings to element concentrations when a large amount of data must be processed. A key problem in programming a computer for this task is the determination of a functional relation between the instrument readings and the concentrations of the analytes. When the relation is determined graphically, it is called the analytical curve; when mathematical methods are employed, the relation may properly be called the analytical function.

This paper emphasizes two aspects of the problem of fitting analytical functions to spectrometric data. The first is the selection of one function from several that might possibly be used. The second is recognition and rejection of data from any standard or standards that might deviate considerably from a function which gives a good fit to the data from the other standards.

Such large deviations might be caused by erroneous spectrometer readings, erroneous analytical data for the standards, systematic influences such as interelement or other matrix effects, or even simply operator error.

Analytical curves are frequently applied in spectrochemical analysis without conscious thought being given to the decisions made during the graphical construction and application of the curves. When a program is written for fitting analytical functions, it becomes necessary to give attention to these decisions. Although there has been some discussion of the theory of analytical functions (for example, *1–4*), in our opinion the existing theory is not adequate to permit straightforward programming. We have chosen to employ empirical approaches which mimic, to some extent, the graphical method.

(1) J. Sherman, *Spectrochim. Acta*, **7**, 283 (1955).
(2) L. R. Pitwell, *Appl. Spectrosc.*, **16**, 77 (1962).
(3) L. R. Pitwell, *Can. Spectrosc.* **11**, 131 (1966).
(4) H. Kaiser, *Optik*, **21**, 309 (1964).

When a curve is constructed graphically, one recognizes that it may not pass exactly through all of the points, because of systematic and random errors in the instrument readings and uncertainty in the concentrations of the standards. The scientist will draw a smooth curve representing a perceived pattern of the points, occasionally rejecting the points for one or more standards which do not conform to the pattern. Unfortunately, pattern recognition is still quite difficult on digital computers, so that a direct analog of graphical methods does not seem possible at this time. Also, in the graphical methods, decisions are sometimes made for reasons that are difficult to program, such as the expectation that self-absorption might cause bending of the analytical curve for a given series of standards.

In a laboratory handling a small range of analytical problems, such as a quality-control laboratory, it is often possible to take full advantage of past experience in fitting analytical functions on a computer. It is then feasible to devote considerable effort to defining the form of the analytical function and assuring that only those standards are employed for which the function fits well to the data. A restricted number of analytical functions are employed, and the computer may be used only to determine the coefficients of the function to give the best least-squares fit of the known equation to calibration data obtained at that time.

This approach may not be feasible for a laboratory with constantly changing analytical requirements. The spectrometrist is often faced with the need to provide an analysis within a limited time and at a limited cost. He must select a set of standards on the basis of whatever information is available on the probable or possible composition of the sample. In making this selection, he may sometimes include one or more standards which he has reason to believe may not give as good a fit to the analytical curve as the other standards, in the hope that this standard might give some additional information. After the standard is run, he will plot the data and decide on the curve to be drawn. Much of the potential saving of time by the use of the computer will be lost if the spectrometrist must first make a plot of the data in order to give information to the computer on the form of the analytical function and the points to be included.

This paper describes two procedures for processing analytical data with digital computers without the need for preliminary plotting. Procedure I is intended for use with a computer system giving rapid response, and preferably one permitting conversational-mode operation. It employs the computer to perform the computations rapidly, but reserves decision making for the scientist. In Procedure II nearly all decisions are programmed in addition to the necessary arithmetic computations. It is thus not restricted to computer systems which provide rapid response, but it does require considerably more computation time.

In both procedures, polynomial equations are employed to represent the analytical functions. Polynomial equations have been employed by others for this purpose (see for ex-

ample, Refs. 5–7), although there is no theoretical reason for their use. We do not know of any theoretically derived equation which will fit better to analytical calibration data than will the polynomial equations.

The polynomial equations employed have the forms

$$C = f(R) \tag{1}$$

$$\log(C) = f[\log(R)] \tag{2}$$

where C is the element concentration or, if the variable internal standard technique is employed, it is the concentration ratio, and R is the instrument response. In emission spectrometry, R may be the line intensity or intensity ratio, measured photographically or photoelectrically. In X-ray spectrometry, R may be the number of counts per unit time or the time required to accumulate a selected number of counts. In atomic absorption spectrometry, R is the absorbance. When each standard is run more than once, R is the average response.

In some cases, the instrument response may be a function of the analytical signal other than a linear proportionality. For example, one make of emission spectrometer gives a response which is inversely proportional to the logarithm of the intensity ratio. In this case, the function fitted is

$$\log(C) = f(R). \tag{3}$$

This is equivalent to Equation 2, since the instrument response is a logarithmic function.

EXPERIMENTAL

The computation procedures were developed and tested on a time-sharing computer system (8, 9), using the BASIC language (9, 10). Procedure II has also been programmed in FORTRAN IV for a batch-loading computer.

In programming the computations, we have taken advantage of subroutines which were available to us for least-square fitting of polynomials. Two different subroutines from the library of the time-sharing computer system were employed for the BASIC programs, and for the FORTRAN program we employed the subroutine ORTHO, which was written by P. Walsh of the National Bureau of Standards.

The accuracies of the BASIC subroutines were evaluated comparatively. ORTHO has been thoroughly evaluated by its author and by statisticians at NBS. The same analytical calibration data have been processed through each of the subroutines, and the coefficients of the equations fitted were compared, as well as the concentrations found from the average instrument reading for each standard and the equation which was fitted. In all cases, the three subroutines gave concentrations found which agreed within one part or less in 100,000. Such differences are insignificant for this application.

The effect of round-off error was evaluated in two ways. The BASIC subroutines were run on two different computers, one of which carries along approximately 9 decimal digits while the other carries about 12. The results agree again to within one part or less in 100,000. Round-off error with ORTHO was evaluated by comparing the results obtained when the same data were processed in single- and double-

precision mode. Again, the differences in the answers were negligible for this application.

The methods of calculation were tested with analytical data developed routinely in this laboratory, as well as some taken from the literature or provided by other laboratories. Emission, X-ray, and atomic absorption spectrometric data were included.

The results were evaluated by plotting the calibration functions derived by the computer programs and comparing these graphs with curves drawn by experienced spectrometrists using the same data. This approach was taken because we do not know of an objective method for evaluating the results of fitting analytical calibration data. Indeed, two experienced spectrometrists will normally construct somewhat different curves from the same calibration data. The only realistic requirement that can be made for the computer methods is that they give results that are deemed reasonable by a knowledgeable scientist.

PROCEDURE I

Description of Procedure. The data provided to the computer include the number of standards and, for each standard, the average instrument reading and the nominal concentration. One additional parameter is entered to instruct the computer to fit either Equation 1, 2, or 3 to the data.

A series of polynomial equations is fitted, starting with the first degree and continuing to higher degrees. The mathematical restraint that there must be at least one data point for each coefficient in the equation places a limit on the highest degree of polynomial that can be fit. The program was initially written to go to as high as the eleventh degree if there were twelve or more standards. Polynomials of such high degree are almost certainly not justified as representations of analytical data, and the program has since been changed to limit polynomials to the fifth degree.

The initial output of the program is printed at this point. It is a table listing the degree of polynomial and, for each degree, the standard error of the concentration and the index of determination. The standard error of the observed concentration is given by $[(\Sigma\Delta^2/(N - D - 1)]^{1/2}$, where Δ is the difference between the concentrations given and found for a standard, N is the number of standards, and D is the degree of polynomial. The index of determination is found by dividing the square of the standard error of the concentration by the variance of the concentrations and subtracting this quantity from one. The variance of the concentrations is given by $\Sigma\delta^2/(N - 1)$, where δ is the difference between the concentration of a standard and the mean concentration of all of the standards.

The program next causes the printing of a request for the user to select a degree of polynomial and then pauses in execution. The user enters his choice by typing a number at the terminal. This choice is based on the output provided up to this point as well as any other information available, such as previous experience with the analytical method and the group of standards. The program then provides as output the coefficients of the equation and a table of results for the standards. The table includes, for each standard, the input instrument reading, the concentration given, the concentration found, and the absolute and percentage differences between the concentrations given and found. The operator may intervene again at this time and stop the processing of the program, if he decides to edit the data to reject the values for one or more of the standards for which the selected equation does not provide an acceptable fit.

If the program has not been stopped, the computer processes the instrument readings for the samples. Each reading

(5) F. W. Anderson and J. H. Moser, ANAL. CHEM., **30**, 879 (1958).
(6) R. H. Wendt, *At. Absorp. Newsletter*, **7**, 28 (1968).
(7) G. Thompson, F. T. Manheim, and K. Paine, Woods Hole Oceanographic Institution Technical Report 68-26, Woods Hole, Mass., 1968; *Appl. Spectrosc.*, **23**, 264 (1969).
(8) J. G. Kemeny and T. E. Kurtz, *Science*, **162**, 223 (1968).
(9) S. D. Rasberry, M. Margoshes, and B. F. Scribner, Natl. Bur. Stds. Technical Note 407, Washington, D. C., 1968.
(10) J. G. Kemeny and T. E. Kurtz, "A Manual for BASIC," Dartmouth College Bookstore, Hanover, N. H., 1965.

is converted to a corresponding concentration through the calibration function. The output includes, for each sample, the average reading and concentration, the number of runs, and, if there were three or more runs for the unknown, the standard deviation of the concentration.

As an example, Table I lists the output of one run of the program. There was no significant improvement in the statistical parameters for polynomials above the first degree, except, of course, for the fifth degree where a perfect fit was obtained since the number of coefficients is equal to the number of standards. Although this function will give a perfect fit to the data for the standards, its use would be dangerous because the function might give unreasonable concentrations for instrument readings interpolated between the standards. The operator chose to use a first-degree polynomial, and entered this information into the computer terminal.

The program then provided the second part of the data, and examination of this list showed that the equation gave a much poorer fit to the data for the second standard than for the others. (Because of the large range of concentrations covered, a fit was made to Equation 2, so that the square of the relative deviations was minimized. Accordingly, the decision to reject the data for the second standard was based on the percentage differences between the concentrations given and found, rather than the absolute differences.) The operator stopped the processing of the program at this point, edited the data from the terminal, and restarted processing of the program.

The second part of Table I shows the output after editing of the input data. The first-degree equation was chosen again, and this time the user decided that the fit to the standards was satisfactory and permitted the computer to process the data for the samples. The output for only two of the samples is shown.

Selection of Polynominal and Rejection of Standards. Only some general rules can be given for the choice of polynomial, since much depends on the user's judgment, which may be influenced by past experience. In most cases, it is preferable that the number of standards be large compared to the number of coefficients in the equation; a useful rule-of-thumb is to have at least two standards per coefficient. As the fit of the equation to the data improves, the index of determination will approach a value near 1, and the standard error of the concentration will approach a fixed value, the limiting precision of the measurements. (Both of the parameters become indeterminate when the number of coefficients equals the number of standards; for convenience, the program prints the numbers 1 and 0 in this case.) If, as in Table I neither statistic changes very much for polynomials higher than the first degree, the first-degree polynomial should be selected. In some cases, the two statistical parameters will show a sudden improvement on going from degree N to degree $N + 1$ and only a small change for higher degrees. The degree to be chosen is then $N + 1$. Table II lists an example of output of this type, and shows a case for which the second-degree polynomial would be chosen.

The absolute and percentage differences in the concentrations given and found for the standards provide information on which the decision is based to reject standards or not. When Equation 1 is fitted, the sum of the squares of the absolute differences is minimized, and the absolute difference is the basis for judgment. When Equation 2 is fitted, the percentage difference should be the criterion for judgment. A standard which shows a much larger absolute or relative difference than the others should probably be rejected. It

Table I. Output from Program for Procedure I[a]

Degree	Index of determination	Standard error of concentration
1	0.858	0.235
2	0.818	0.266
3	0.826	0.260
4	0.898	0.199
5	1	0

What degree do you want to use? 1

Term	Coefficient
0	−4.13739
1	1.54412

Instrum. reading	Concn. given	Concn. found	Diff.	% Diff.
409.5	1.110	0.787	−0.323	−29.0
633.0	0.634	1.543	0.909	143.3
254.5	0.387	0.378	−9.18E-3[b]	−2.4
154.0	0.178	0.174	−4.06E-3	−2.3
134.5	0.140	0.141	1.13E-3	0.8
923.5	4.590	2.764	−1.82	−39.8

Degree	Index of determination	Standard error of concentration
1	0.997	3.53E-2
2	0.996	4.18E-2
3	0.998	2.99E-2
4	1	0

What degree do you want to use? 1

Term	Coefficient
0	−4.75698
1	1.82740

Instrum. reading	Concn. given	Concn. found	Diff.	% Diff.
409.5	1.110	1.039	−7.09E-2	−6.4
254.5	0.387	0.436	4.87E-2	12.6
154	0.178	0.174	−4.02E-3	−2.3
134.5	0.140	0.136	−4.16E-3	−3.0
923.5	4.590	4.593	2.51E-3	5.5E-2

RESULTS FOR SAMPLES

Sample No.	Av. Reading	Av. Concn.	Std. Dev. of Concn.	No. of Runs
9182	146.6	0.159	4.5E-3	5
9183	782.2	3.390	2.4E-2	4

[a] The number of significant figures has been truncated; program provides six.
[b] Exponential notation employed; 5E-2 = 0.05

Table II. Partial Output from Program for Procedure I

Degree	Index of determination	Standard error of concentration
1	0.9951	0.4012
2	0.9999	1.756E-2
3	0.9999	1.607E-2
4	1	0

may also be rejected if the sign of its difference is different from all of the others. The level of the absolute or relative difference that will be considered significant will depend on the accuracy expected for the analysis.
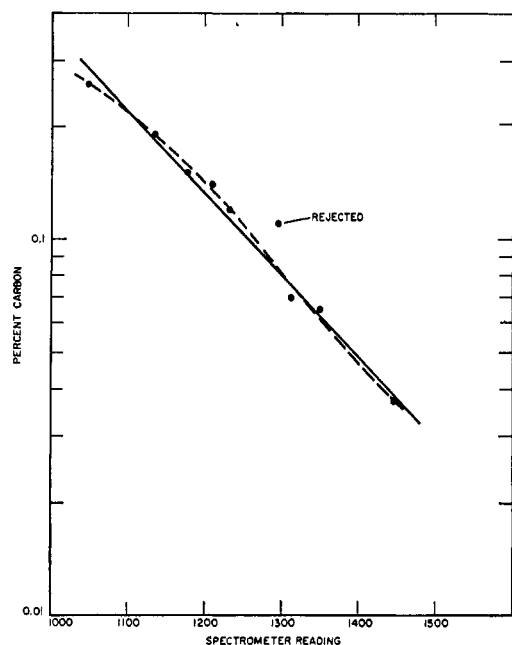
Figure 1. Functions fitted by Procedure I to emission spectrometer calibration data for carbon in steels

— First-degree polynomial; - - - third-degree polynomial

Figure 2. Per cent differences between concentrations given and found vs. spectrometer reading, for data and functions shown in Figure 1

Results. The results obtained by Procedure I will depend, as can be expected, on the experience and judgment of the user. Given the same calibration data, different operators may choose different analytical functions, much as they might construct different curves graphically. For example, Figure 1 shows the curves selected by two individuals; the data are from a paper by LeRoy (11) and represent calibration data for the determination of carbon on a vacuum spectrometer. The instrument used gave a reading inversely proportional to the logarithm of the intensity ratio, so Equation 3 was fit. The inverse proportionality accounts for the negative slopes of the curves.

In this instance, both users chose to reject the data for the standard containing 0.11% carbon, but one user selected a first-degree polynomial while the other chose a third-degree equation. Neither curve appears to be an unreasonable representation of the data, and thus both equations would be acceptable. Figure 2 is a plot of the percentage differences between the concentrations given and found; it provides a
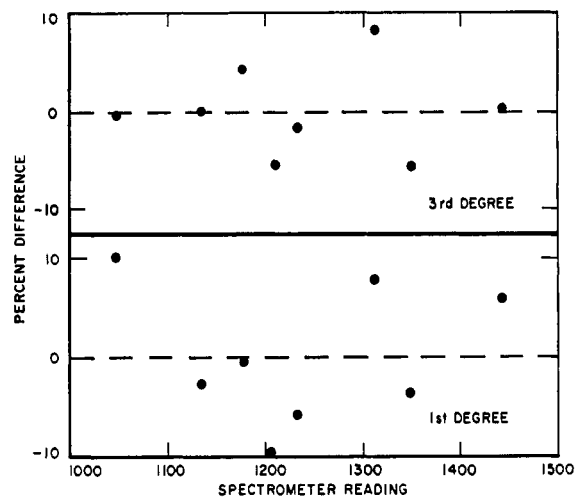
more sensitive comparison of the equations. The points for the first-degree polynomial exhibit somewhat more scatter than those for the equation containing two more terms, and the former also show a tendency to be negative in the middle range of instrument readings and positive at either end. These observations indicate that the fit obtained with the higher polynomial might be preferable, but the choice is not clear-cut. There are statistical tests to determine whether or not the extra terms in the polynomial are justified, but these are of doubtful reliability with a small number of data points.

Table III lists the first part of the output from the program for data on silicon, also taken from the paper by LeRoy (11). In this instance, the two operators selected third- and fourth-degree polynomials, which are shown graphically in Figure 3. Neither operator detected that the standard of lowest concentration did not fit well on a curve with the other standards. LeRoy used a first-degree polynomial to represent the analytical function, but without rejection of any standards. The results obtained with these data by Procedure II will be described later.

Table III. Partial Output from Program for Procedure I with Data for Silicon from Paper by LeRoy (11)

Degree	Index of determination	Standard error of concentration
1	0.9920	0.1183
2	0.9972	6.96E-2
3	0.9991	3.99E-2
4	0.9994	3.20E-2
5	0.9993	3.48E-2

(11) J. R. LeRoy, Ann. ISA (Instrum. Soc. Amer.) Conf. Proc., 21 (3), 6.1--1-66, 8 pp (1966).
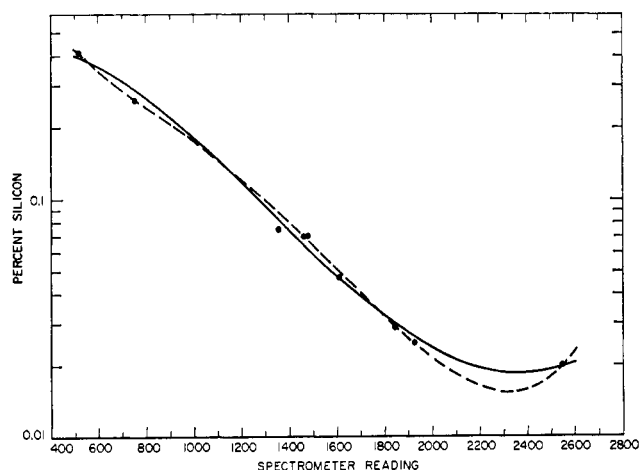
Figure 3. Functions fitted by Procedure I to emission spectrometer calibration data for silicon in steels

— third-degree polynomial; - - - fourth-degree polynomial.

This example, and a few others which were similar, showed a weakness in Procedure I, which was related to the limited data output rate of the teletypewriter terminal for the time-sharing computer system. It was not feasible to program the plotting of even a rough representation of each curve, since too much time would have been taken. To circumvent this problem, the program has been revised to test for a zero slope or a change in the sign of the slope. If either of these conditions is found, a brief statement is printed after the other data for that equation. Also, a coarse plot of the selected equation can be printed, at the option of the user.

When the data of Figure 3 were processed by the revised program, the user was alerted to the unsatisfactory nature of the polynomial equations above the second degree. The listing of the data for the standards, for either the first or second degree, showed only that the general fit was unsatisfactory, but did not point to any one standard for rejection. A manual plot of the data is called for under these circumstances.

An output device with a higher writing rate would also be useful. If a fast line printer is available, much more information can be printed by the computer in a limited time than is possible with a teletypewriter. A cathode-ray tube display would be even better, since it would allow the user to see the calibration data and the curves fit to it in a familiar graphical form. Neither a high-speed line printer nor a cathode-ray display has been available to us with a time-shared computer.

PROCEDURE II

Description of Procedure. The input data to the program for this Procedure are similar to the data for Procedure I.

The program first fits a series of first-degree polynominals, one polynomial through all of the n data points, and n equations, each through $n - 1$ data points. This set of computations permits a comparison to be made to determine whether a significantly better fit to the data can be obtained by rejecting any one of the standards.

The decision whether a fit without one of the standards is better than the fit to all of the points is based on the parameter

$$P_{i,j} = [\Sigma d^2/(N - 1)]^{1/2} \qquad (4)$$

where the subscripts i and j refer to the degree of polynomial and to the standard being considered for rejection, respectively ($j = 0$ when all standards are included), N is the number of data points to which the equation is fit, and d is given by

$$d = (C_g - C_f)/C_g \qquad (5)$$

where C_g and C_f are the concentrations given and found for each standard. $P_{i,j}$ is quite different from the standard error of the concentrations, and the latter parameter has been found to be a less reliable test datum for this purpose.

Each time an equation is fit, $P_{i,j}$ is determined and, after a series of polynomials of degree i has been fitted, this list is scanned to find the lowest value. If $P_{1,0}$ has the lowest value, no standard is rejected. Otherwise, standard k is rejected if $P_{1,k}$ has the lowest value and it is less than one third of $P_{1,0}$.

If no standard is rejected from the result of the first-degree fits, a similar set of computations is made for a second-degree polynomial, provided that the number of standards is large enough. (In Procedure II, the absolute requirement is made that the number of standards must always be at least twice the number of coefficients in the equation.) This set of computations gives a list of $P_{2,j}$ values. $P_{1,0}$ is now compared with $P_{2,0}$. If the latter is less than three fourths of the former, it becomes the basis for later comparisons;

otherwise, the first-degree equation is assumed to be the best so far. Then the list of other $P_{2,j}$ values is scanned to select the lowest values. This is compared with either $P_{1,0}$ or $P_{2,0}$, depending on the results of the previous test, to determine whether a standard should be rejected at this point. The basis for rejection is again that at least a factor-of-three improvement be obtained in $P_{i,j}$ by this rejection. If a rejection is made, a new trial of first-degree polynomials is made. Otherwise, the next higher polynomial is tried, provided that there are enough standards. No attempt is made to fit polynomials above the fourth degree.

The computations involve some arbitrary decision levels and requirements. These include: 1) the restriction that the number of standards always be at least twice the number of coefficients in the equation, and 2) the decrease in $P_{i,j}$ required to reject a standard or to include additional terms in the equation. Statistical decision levels are available for these choices, but we have not employed them because the number of data points is normally too small to make the usual statistical procedures reliable. The decision levels and requirements which were applied were selected on the basis of a number of trials with analytical data. It is not expected that they will always give acceptable decisions, but the level of reliability does seem to be high, as will be discussed below.

An illustration of the computation procedure will now be given. Table IV lists calibration data for the determination of silicon in steels by photoelectric emission spectrometry. Table V lists the values of $P_{i,j}$ which were obtained when Equation 2 was fitted to the data. On the first set of computations, the lowest $P_{1,j}$ was not less than one third of $P_{1,0}$ and no points were rejected. $P_{2,0}$ was much smaller than three fourths of $P_{1,0}$, so the other $P_{2,j}$ values were compared to $P_{2,0}$. The smallest $P_{2,j}$ was $P_{2,1}$, and this was less than one third of $P_{2,0}$, so the data for the first standard were rejected.

At this point, the program returned to a trial of first-degree polynomials, with seven standards, giving the results
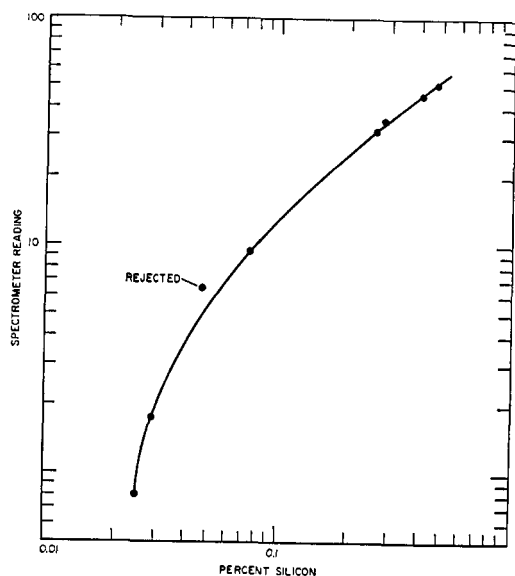
Table IV. Calibration Data

Standard No.	Silicon concn., % by wt.	Chart reading
1	0.047	6.45
2	0.28	35.3
3	0.41	46.4
4	0.48	51.9
5	0.029	1.75
6	0.025	0.8
7	0.075	9.45
8	0.26	31.85

Table V. Values of $P_{i,j}$ Found

	Eight standards			Seven standards	
j	$i = 1$	$i = 2$	j	$i = 1$	$i = 2$
0	0.337	0.0532	0	0.256	0.0160
1	0.256	0.0160	—	—	—
2	0.364	0.0569	1	0.282	0.0058
3	0.346	0.0574	2	0.265	0.0175
4	0.335	0.0573	3	0.251	0.0164
5	0.368	0.0527	4	0.284	0.0174
6	0.232	0.0555	5	0.184	0.0174
7	0.347	0.0426	6	0.169	0.0174
8	0.364	0.0546	7	0.282	0.0149

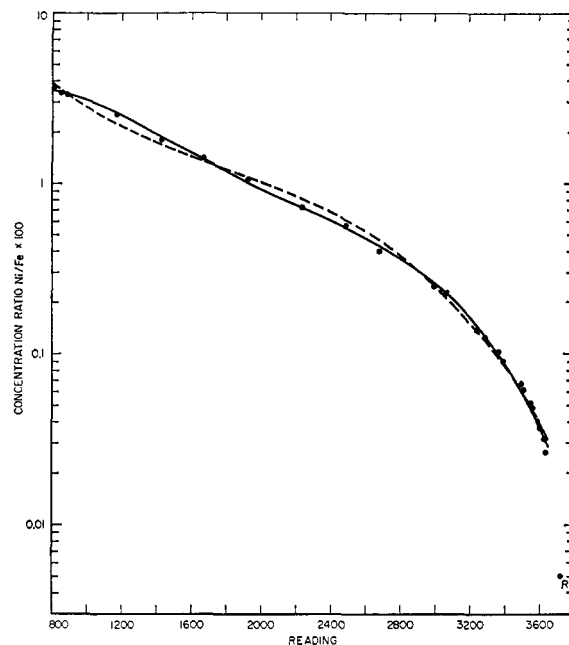Figure 4. Function fitted by Procedure II to emission spectrometer calibration data for silicon in steels

Figure 5. Functions for calibration of an emission spectrometer for determination of nickel in irons and steels

Point marked R was rejected. - - - third-degree polynomial; —— fourth-degree polynomial.

shown on the right of Table V. No significant improvement in $P_{1,j}$ was found compared to $P_{1,0}$, so no standards were rejected. $P_{2,0}$ was less than three fourths of $P_{1,0}$ and no $P_{2,j}$ values were less than one third of $P_{2,0}$. The seven standards remaining did not give enough data points to fit a third-degree polynomial, and the final decision was therefore to use the second-degree polynomial fitted to seven data points. A plot of this curve is shown in Figure 4.

Results. The function which was fitted to the data in Table IV and is shown in Figure 4 is not the only function which might be acceptable. Instead of rejecting standard number 1, it would have been acceptable if the program had rejected the two standards of lowest silicon content and fitted a straight line to the remaining data points. This illustrates one of the difficulties in appraising any program for fitting analytical calibration functions, which is the result of an inadequate theory and sometimes of inadequate information. For example, the shape of the curve shown in Figure 4 can be explained as being due to self-absorption, but proving this explanation would require considerably more data than is normally obtained in calibrating a spectrometer. The concentration at which self-absorption will cause curvature of the analytical function can be predicted for flame emission data (12), but not as yet for arc and spark excitation.

The data listed in Table III and shown graphically in Figure 3 were also processed through Procedure II. In this case, the result was rejection of the data for the standard of lowest silicon content and fitting of a first-degree polynomial (straight line) through the remaining data points. This is in good agreement with what most chemists would have done in plotting an analytical curve with these data.

Figure 5 shows data for the determination of nickel in ductile irons and cast steels by spark excitation in an argon atmosphere (13). The concentraion of iron was not constant in this series of standards, so the concentration ratio, 100 × C_{Ni}/C_{Fe}, was plotted vs. the instrument readings. In the

instrument used, the output signal is inversely proportional to the logarithm of the intensity ratio, so Equation 3 was fit to the data. There were 24 standards.

Figure 5 shows two curves, representing a third-degree polynomial (dashed line), and a fourth-degree equation (solid line). In both cases, the point for the standard of lowest nickel content has been rejected. Figure 6 is a plot of the percentage deviations for both curves, showing that a fourth-degree polynomial gave a significantly better representation of the calibration data, as evidenced by the systematic variations of the deviations. The fourth-degree equation was selected by the program. We have not found any set of data which required a polynomial equation above the fourth degree to give an adequate representation, although the possibility cannot be ruled out.

Figure 7 shows a fit of equation 1 to calibration data for the determination of vanadium in ductile irons and cast steels by X-ray fluorescence spectrometry. The data were not corrected for the background radiation reaching the detector, but the only effect of such a correction would have been to change one term in the polynomial equation and there would have been no effect on the fit of the equation to the data or on the degree of polynomial chosen. This would not have been true for a fit of Equation 2 to the data, as will be shown in the Discussion.

Procedure II may produce an unacceptable function under some circumstances, as when the data are poorly correlated. Figure 8 shows input data and the curve fitted by Procedure II for one such case. Curves like this could be detected by checking for a change in the sign of the slope, as was done in Procedure I. However, Procedure II is designed for use with a batch-loaded computer equipped with a high-speed line printer, and a rough plot of the curve can be obtained at little cost or delay. Accordingly, the program incorporates a subroutine to produce a line-printer plot showing the input data and the curve fitted, and this may be inspected by the user.

(12) J. D. Winefordner, T. J. Vickers, and L. Remington, ANAL. CHEM., **37**, 1216 (1965).
(13) H. C. Dilworth, *Amer. Soc. Testing Mater., Spec. Tech. Publ.*, **376**, 23 (1965).
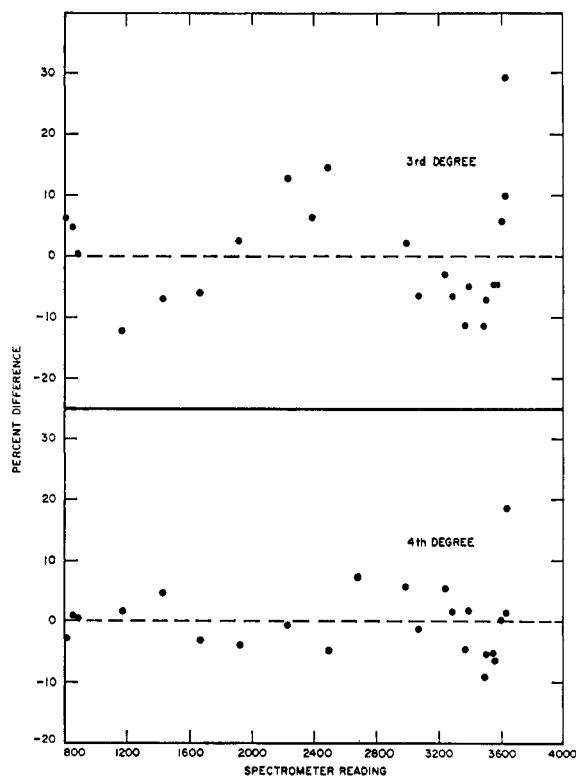
Figure 6. Per cent differences between concentrations given and found *vs.* the spectrometer reading, for data and functions shown in Figure 5

The data point marked R in Figure 5 is not shown here

Procedure II has only rarely produced equations which exhibited a change in the sign of the slope within the region of calibration. In our experience, this has occurred mainly when the data were so poorly correlated that they were not really suitable for calibration.

DISCUSSION

The Polynomial Equations. The functions fitted, as shown in Equations 1, 2, and 3, have the reading as the independent variable and the concentration as the dependent variable. There is also the possibility of reversing the equations, and treating the reading as the dependent variable. This procedure would seem to be logical, since in a physical sense the reading is a consequence of the concentration and the latter variable is the one that is controlled through the choice of standards.

The choice of the forms of equation shown is a practical one, since it avoids the problem that would occur with polynomials of the second degree or higher with the forms $R = f(C)$ or $\log(R) = f[\log(C)]$. For equations of this form, when the readings for the samples are to be converted to concentrations, second-degree polynomials will have two roots, third-degree equations will have three roots, etc. Thompson, Mannheim, and Paine (7) and Wendt (6) have employed polynomial equations with the concentration as the independent variable for fitting analytical functions. For selecting one of the roots they have devised procedures which are said to work well for their applications. However, the generality of the procedures is not known, and we have chosen to avoid the problem by treating the concentration as the dependent variable.

The choice between fitting Equations 1 or 2 can be compared to plotting data on linear or logarithmic coordinates.
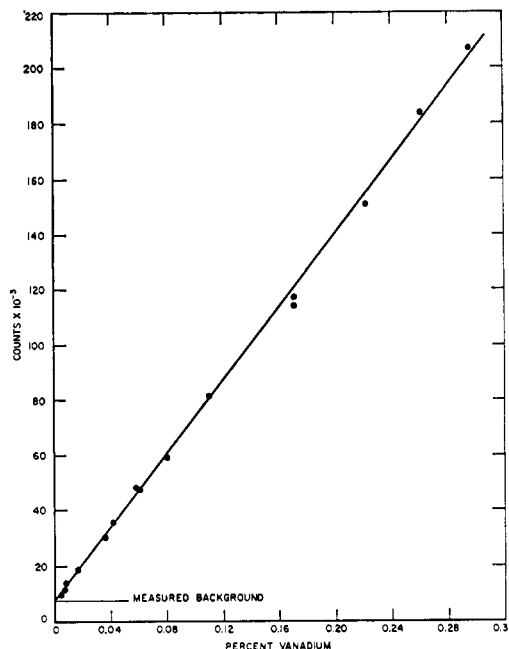

Figure 7. Function fitted by Procedure II to X-ray fluoresence spectrometer calibration data for vanadium in irons and steels

Equation 2 clearly cannot be used when there is a blank, or a standard with a nominal concentration of zero. Fitting of Equation 1 to the data will minimize the sum of the squares of the absolute deviations, while fitting of Equation 2 will minimize the sum of the squares of the relative deviations. This results in a different weighting of the data points. When the range of compositions of the standards is small, either equation may be used, since the practical effect of the difference in weighting will probably not be important. When there is a large range of compositions for the standards, Equation 2 is usually preferred, since then the relative error will tend to remain constant (see discussion below). In most instances, we have used Equation 1 only when the ratio of the highest to the lowest concentration is ten or less or when there is a blank standard.

In some cases, the relative errors are not expected to be constant. The X-ray fluorescence data in Figure 7 represent one such instance. Here, the relative counting errors and the relative errors in the analyses of the standards are ex-
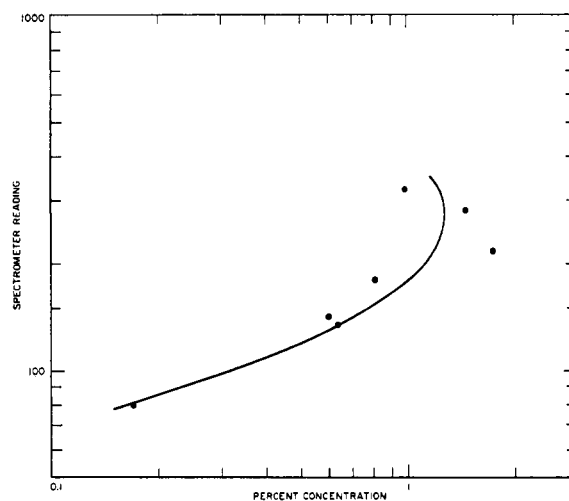

Figure 8. Function fitted by Procedure II with poorly correlated calibration data

pected to decrease with increasing element content. In such instances, Equation 1 should be employed.

The data of Fig. 7 also illustrate another reason for sometimes using Equation 1 when the range of compositions is large. Here there was a background signal, which was not corrected, and it represented a significant fraction of the total signal measured for the standards of low concentration. When Equation 1 is fitted to such data, the only effect of the background correction is to change the value of one term in the equation, as was mentioned above. However, the situation is more complex when Equation 2 is employed for such data. This is illustrated in Figure 9, where the upper curve shows the second-degree polynomial selected by the program of Procedure II for the uncorrected data, and the lower curve shows the third-degree polynomial employed for the data after correction for background. The counts subtracted for background were the signal measured at a wavelength near the analytical line with one of the standards; apparently, this was somewhat too large a correction, judging by the shape of the corrected curve at low concentrations. In this instance, fitting Equation 2 results in the use of more coefficients to aquately describe the data.

The Parameter $P_{i,j}$. The choice of $P_{i,j}$ as a decision datum is based on the assumption that, for many spectrometric methods of analysis, the relative error tends to remain constant over a wide range of concentrations. Smith and Mathews (14) have claimed that the relative error remains rather constant for many methods of analysis, and have given data from a gas chromatograph as an example. Ramírez-Muñoz (15) has computed relative errors in concentration as determined by atomic absorption spectrometry, based on the assumption that the per cent transmission or per cent absorption scale can be read with a constant absolute error. If it is assumed that the constant error in the scale reading is 0.2, which seems to be a reasonable value for most instruments of this type, the relative error in concentration is nearly independent of the per cent absorption over most of the range of these measurements, increasing only for very small or very large values of the per cent absorption. A similar conclusion has been reached by Slavin (16) in a study of the relative errors in photographic microphotometry.

Other steps in the analytical procedure will also contribute a constant relative error. Examples of this are dilutions, whether these involve mixing of sample with diluent by weight or volume, so long as the amounts of material used and the methods of measurement are not determined by the concentration of the analyte in the sample.

In some instances, the relative error cannot be expected to remain constant. The case of X-ray fluorescence analysis has been discussed already. Another example is analysis near the limit of detection, which has been defined (17) as the concentration of the element giving a ratio of signal-to-noise of three. Near the limit of detection, the relative error is determined by the magnitudes of the signal and the noise. If the noise increases slowly or not at all with the signal, the relative error at higher concentrations can be expected to be considerably smaller. In some instances, a major cause of variability (noise) may be chance contamination. Contamination will clearly introduce larger relative errors at low concentrations of the analyte.
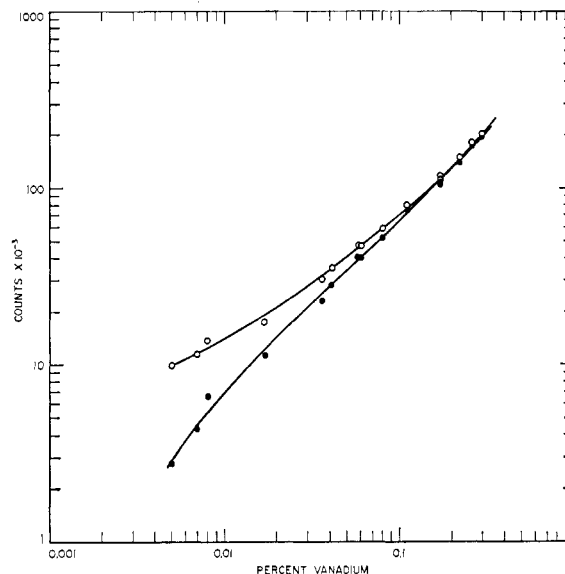

Figure 9. Function fitted to Equation 2 by Procedure II with X-ray fluorescence spectrometer calibration data for vanadium in irons andsteels

○ line-plus-background counts; ● same data with a constant background subtracted

It is interesting that, even with data that are not expected to be subject to a constant relative error, acceptable results are produced by Procedure II. This is illustrated by the X-ray fluorescence data shown in Figure 7. We have also plotted and computed many sets of data taken near the limit of detection, and these also gave acceptable results.

Recognition of Unacceptable Results. The functions fitted to the data may be unacceptable for either of two reasons: the deviations of the points from the equation may be very large; or the shape of the graph of the function may be unlike any that might be drawn by an experienced analyst. Functions which are unacceptable for the first reason can be recognized from the statistical data provided by the programs. An example is the first output shown in Table I, for the computations prior to deletion of the data for one standard. For each degree of polynomial, the index of determination is far from unity and the standard error of the concentration is rather large. If a better fit had not been obtainable by rejection of one of the standards, the data would probably not have been usable for analysis.

The parameter $P_{i,j}$ of Procedure II also gives an indication of a poor fit. For the data shown in Figure 8, $P_{i,j} = 0.211$, indicating that in about one third of a large number of analyses, the concentration found can be expected to deviate by 21% or more from the true concentration. Thus, this function would not be considered to be acceptable unless only a very approximate analysis was required.

Functions which are unacceptable for the second reason are more difficult to recognize. The procedure of checking for a point of zero slope or a change in the sign of the slope has already been mentioned, and this is useful for detecting curves such as those shown in Figures 3 and 8. Other procedures can probably be developed to recognize other cases. A residual plot, as in Figure 6, would be useful in some cases, and could be made readily on a line printer or teletypewriter. Probably the most generally satisfactory procedure is a direct plot of the curve. Both procedures incorporate plots with a resolution of one part in fifty, which will be an adequate check in most cases. Additional protection could be provided by a

(14) E. D. Smith and D. M. Mathews, J. Chem. Educ., **44**, 757 (1967).
(15) J. Ramírez-Muñoz, Microchem. J., **12**, 196 (1967).
(16) M. Slavin, Appl. Spectrosc., **19**, 28 (1965).
(17) H. Kaiser, Z. Anal. Chem., **209**, 1 (1965).

high-resolution plot on an X-Y plotter; we have not attempted this because of the extra time and cost involved. If the computer is equipped with a cathode-ray screen display, this can be employed to show a representation of the data at reasonably high resolution; such a display has not been available to us.

The procedures described here gave unacceptable functions in only a small fraction of the cases tested. However, even an occasional failure can have serious consequences, and it is a wise precaution to incorporate tests such as those described above into the program. It may also be wise to have the program determine whether or not the readings for the samples lie near those for one or more standards. Clearly, more reliance can be placed on the analytical results when the portion of the analytical function near the instrument readings for the samples is well calibrated by standards of nearly identical composition. When several elements are being determined in the same sample, it would be possible to include in the program some measure of the over-all composition of the standards and samples, to help ensure that the standards are similar to the unknowns. We have not yet attempted to do this, but it would be a desirable feature.

Type Standards. In some instances, only one or a very few standards are available which are similar to the expected compositions of the samples and can thus be employed for calibration. When a calibration curve is constructed graphically, this problem is sometimes circumvented by employing other standards to define the shape of the curve and shifting the curve through the data for whatever standards are similar to the samples, called type standards. The analog of this process can be incorporated into either of the Procedures fairly simply. Since we employ this procedure very rarely, we simply include the data for the instrument readings on the type standards with the data for the samples. The differences between the concentrations found and the nominal concentrations are then averaged (if there is more than one type standard) and this is applied as a correction to the samples. When a logarithmic function has been used for the analytical equation, the correction should be the relative deviation.

The correction could be easily incorporated into the program if this procedure were used more often. It is only necessary to change the value of one term in the function to "shift" the curve.

Accuracy of the Analysis. The precision of analysis can be determined rather easily by replicate measurements on the same sample or on different portions of a homogeneous sample. It is more difficult to evaluate the accuracy of the analyses. It is often said of spectrometric analytical methods that the accuracy is as good as the precision, but this implies that there is no significant error in the analytical curve or analytical function. In practice, imprecise knowledge of the curve or function can contribute inaccuracies which are comparable in magnitude to the random errors of the measurement, or even larger. Still, this rather obvious limitation on accuracy has been generally ignored, and it is usual for an analytical report to include only the average composition found and an error estimate based only on the replicate measurements. This is not surprising when graphical methods are employed, since it will then be difficult to evaluate possible errors caused by imprecise calibration.

Occasionally, systematic matrix effects are known and appropriate correction can be applied. If this can be done, corrections should be made to the instrument readings before the curve is plotted or the function is fitted. In this discussion, it is assumed that either no corrections can be made or that they have already been made.

When an analytical function is fitted on a computer, additional statistical treatment of the data can be done at the expense of additional programming and computer time. Well-established statistical procedures are known for computing error limits around a function. However, we know of no accepted procedure for combining the error of the function with the random errors in the instrument readings. Svehla, Páll, and Erdey (*18*) and Khalifa, Svehla, and Erdey (*19*) have described a procedure they have employed in spectrophotometry and atomic absorption spectrometry. Their procedure is based on certain assumptions that are not acceptable to us. In atomic absorption spectrometry for example, they compute the relative standard deviation in the concentration found from the linear combination of the relative errors of the absorbance measurements and of the two coefficients obtained by a least-square fit of a first-degree polynomial to data for standards. However, unless each standard has been run a very large number of times, the error of the coefficients includes the contribution of the error of the absorbance measurements. Each coefficient is also, in part, dependent on the other. In this circumstance, a linear combination of the relative errors gives too large an estimate of the error in the concentration found.

We have adopted a procedure based on the assumption that the standards behave like the samples; this assumption is basic to the use of standards for calibration of an instrument. Our usual procedure is to carry out the same number of replicate measurements on each of the standards and samples. Then the deviations of the points for the average readings on the standards should fairly accurately correspond to the deviations which would be obtained if the samples could be treated as standards. Thus, the accuracy of the analysis of the standards is determined and is employed as a measure of the accuracy of the analysis of the samples.

The errors in the analysis of the standards are determined as an integral part of the computation procedure. Just how this information is used is based on what assumptions are made. In Procedure II, the parameter $P_{i,j}$ is employed as a measure of the relative error to be expected in the analysis, and the expected absolute error is computed by multiplying the concentration found for each sample by $P_{i,j}$. Other procedures would have to be employed when the assumption of constant relative error cannot be adopted.

SUMMARY

Two procedures are described for fitting spectrometric analytical functions on digital computers. In both procedures, the only decision to be made before the computations are started is to select either a fit of concentration *vs.* instrument reading or log(C) *vs.* log(R). In either case, a series of polynomial functions are fit to the calibration data.

Procedure I, intended for use on time-sharing computer systems, yields a shorter program, but leaves to the analyst the decisions of equation selection and possible rejection of unsuitable standards. Procedure II includes the logic necessary for having the computer perform these operations. Although the program for this is more complex, it is especially useful when time-sharing is not available; furthermore, it consumes less of the analyst's time.

With both procedures, fitting of the analytical function is followed by conversion of the instrument readings for the

(18) G. Svehla, A. Páll, and L. Erdey, *Talanta*, **10**, 719 (1963).
(19) H. Khalifa, G. Svehla, and L. Erdey, *Talanta*, **12**, 703 (1965).

unknowns into element concentrations, together with appropriate statistical computations.

The use of either of these methods results in a significant saving in time for the analyst. The machine time required for processing of the data depends critically on the particular computer system. The time required for the analytical chemist depends primarily on the ease of entering data. This time can be made very small if the spectrometer is equipped with a digital readout on cards or tape for direct input to the computer or if the spectrometer and computer are directly interfaced.

There are other advantages to the use of a computer for fitting analytical functions. In Procedure II, all significant decisions are incorporated in the computer, so that operator judgment will not influence the analytical results. Similarly,

the effects of human error are minimized. An additional advantage is that the computer can be programmed to perform further computations and thus to provide in the output an indication of both the precision and accuracy of the analyses. A computation procedure is described which bases the estimate of the accuracy on the differences between the concentrations given for the standards and those found for those standards from the analytical function and the instrument readings. The validity of this method of estimating accuracy depends on assumptions which are described. It is felt that even a limited estimate of accuracy of the analyses is preferable to basing the estimated uncertainty of the analyses entirely on the precision of replicate instrument readings.

Direct Determination of Silver in Air by Atomic Absorption Spectrometry

Harry W. Edwards

Department of Mechanical Engineering, Colorado State University, Ft. Collins, Colo. 80521

Silver-containing air streams are introduced in the primary air supply to the burner of an atomic absorption spectrometer for direct silver determination. The analytical method is calibrated by a filtration technique in which continuous silver-containing air streams are produced by passing air over heated silver iodide. The amounts of silver on the filters are determined by conventional methods of extraction and AAS solution analysis. The use of an integrating digital voltmeter to measure the spectrometer signal significantly enhances the readability of the relatively noisy absorptions associated with direct measurement. The detection limit is approximately 3 $\mu g/m^3$ which is applicable to the direct determination of silver in air in the concentration range typical of certain weather modification activities.

THE QUANTITATIVE determination of silver added to the atmosphere is of considerable interest in weather modification. Rigorous theoretical and economic evaluations of the effectiveness of silver iodide as a cloud seeding agent require the determination of silver in air. Previously reported meteorological methods for the determination of silver in the atmosphere rely primarily on the ice-nucleating properties of silver iodide (1). Criticisms of cloud chamber methods for airborne silver analysis focus upon the following characteristics of such methods: nonspecificity, lack of calibration, and excessive time between sampling and readout.

Methods for the determination of elements dispersed as particulates in air are usually characterized by concentration and extraction steps prior to analysis. Direct determinations—e.g., methods which do not involve concentration and extraction steps prior to analysis—appear comparatively infrequently in the literature. The relatively severe sensitivity requirement imposed by the $\mu g/m^3$ concentration level of solid material in a typical aerosol requires manipulation of inconveniently large volumes of air for conventional analysis. Atomic absorption spectrometry therefore offers

considerable promise for the direct determination of metallic elements dispersed as particulates in air. A predecessor of AAS has long been used for the detection of mercury vapor in air (2), and AAS has recently been employed by White (3) for the measurement of lead and cadmium fumes in foundries and by Thilliez (4) for the determination of tetraethyl- and tetramethyllead in the environment of a manufacturing plant.

Reported in this paper is a method for the direct determination of silver in air by AAS. Calibration is accomplished by a filtration technique in which the absorbances associated with air streams containing varying amounts of finely divided silver iodide are measured. Application of the method to the direct determination of silver in air for meteorological studies is discussed.

EXPERIMENTAL

Apparatus. The apparatus used is shown in Figure 1. Air enters the oil-free diaphragm pump, *A*, at the rate of 3.96 l./min as measured by the rotometer, *B*. Silver iodide is dispersed in finely divided form in the air stream by the silver iodide smoke generator, *C*. By means of pinchcocks located at *D* and *F*, the air stream is either passed directly into the Perkin-Elmer 303 spectrometer burner, *H*, or diverted through the Millipore filter apparatus, *E*. The primary air supply for the burner, which is filtered by means of the standard cartridge unit furnished by the spectrometer manufacturer, enters the system at *G*. The instrumental settings were as follows: silver lamp current, 12 mA; wavelength, 3280.7 Å; slit width, 0.3 mm; primary air pressure, 28 psi. With the diaphragm pump in operation, it was found necessary to reduce the primary air flow by approximately 15% from that recommended by the spectrometer manufacturer for flame stability. The diaphragm pump was selected to provide a system which could be readily adapted to field conditions. The primary advantage of this pump is that

(1) B. J. Mason, "Clouds, Rain, and Rainmaking," University Press, Cambridge, 1962.

(2) T. T. Woodson, *Rev. Sci. Instrum.*, **10**, 308 (1939).
(3) R. A. White, *J. Sci. Instrum.*, **44**, 678 (1967).
(4) G. Thilliez, ANAL. CHEM., **39**, 427 (1967).