Use of Weighted Least-Squares Splines for Calibration in Analytical Chemistry

Iuliu Sorin Pop,[†] Valeria Pop,[‡] Simona Cobzac,[‡] and Costel Sârbu*[,‡]

Faculty of Mathematics and Computer Science and Faculty of Chemistry and Chemical Engineering, "Babeş-Bolyai" University, RO-3400 Cluj-Napoca, Romania

Received May 27, 1999

Weighted least-squares spline functions are discussed and applied for calibration processes in analytical chemistry. Different weighting techniques are also considered, and for the evaluation of the results some quality coefficients are proposed. Depending upon the structure of the data, some weighting procedures may improve the results dramatically. Considering the results obtained in the case of TLC densitometry, it seems that nonlinear weighting procedures based upon the distance to the function are the best ones, with a plus for the *y*-distance type. It is difficult to give general rules regarding the optimal parameters of the weighted calibration splines—function order (*m*), number (*N*) and distribution of knots, and weighting technique. These depend upon the structure of the data. However, higher order splines are not recommended since the result might become extremely unstable. The example used to illustrate the performances of the procedures discussed here involved only a single independent variable. The method is general and extends practically to any number of variables, thus resulting in a multivariate approach.

INTRODUCTION

Quantitative determinations involving instrumentation are often accomplished by establishing a calibration function. This should be realized before doing any measurements on a chemical system. Calibration of an instrument for chemical quantification requires a mathematical relationship between the response of the instrument and the concentration. Several techniques for expressing this relationship are known.[1] The common model is a linear regression—usually a first-order polynomial—although second order or higher order ones have been used.[2−3] The ordinary linear least-squares regression is based on the assumption of an independent and normal error distribution with uniform variance (homoscedastic).

In the spite of the fact that the normal linear least-squares method is not optimal, there is a justification for using it in the cases where the conditions are only approximately met. In particular, the Gauss−Markov theorem states that, if the errors are random and uncorrelated, the method of least squares gives the best linear unbiased estimate of the parameters.[4]

However, with the recent development of polynomial spline functions, different problems have been satisfactorily solved. Spline functions have the ideal property of being able to describe any continuous variation in the dependent variable and still be computationally facile.

Recent results suggest that, in the absence of a good analytical model, the use of the spline functions could be a good approach in the treatment of data. Applied first in statistical sciences, spline functions became extremely important in the chemical data analysis in the last few decades.[5−9] Until now, the greatest part of successful applications of spline functions in analytical chemistry has been devoted to fields such potentiometric titrations,[10−13] spectroscopy,[14−18] X-ray spectroscopy,[19−21] and signal processing.[22−24]

In the present paper least-squares splines are used instead of interpolatory ones. These appeared to be more suitable for approximating large data sets. To estimate the parameters of the calibration spline model, a weighted regression procedure is performed. Different weighting techniques are also considered, and for the evaluation of the results some quality coefficients are proposed.

THEORETICAL CONSIDERATIONS

To fit lower order polynomials to an extensive set of data, the idea of trying to force a single polynomial through the points should be abandoned. Instead, different polynomials may be used to connect ranges of points, piecing each section smoothly together. The shape described by a spline between two adiacent points—called knots—could be practically of any polynomial degree.

Considering first a set of n data pairs (x_1, y_1), ..., (x_n, y_n), N knots are defined in the interval (a, b)—containing x_1, ..., x_n— such that

$$a < X_1 < X_2 < ... < X_N < b \tag{1}$$

By convention, if no knots are considered, we take $N = 0$. These knots divide the initial interval into $N + 1$ subintervals $[X_{i-1}, X_i]$, $i = 1, ..., N + 1$, with $X_0 = a$ and $X_{N+1} = b$ for the sake of simplicity. Restricted on each subinterval, the spline function—denoted by s_m—is an mth degree polynomial. These polynomials are generally distinct, but have the same degree (m), defining the degree of the spline.

To obtain a maximal regularity of the spline, continuity conditions for s_m and its derivatives up to the degree $m - 1$ are imposed on knots

$$s_m^{(k)}(X_j - 0) = s_m^{(k)}(X_j + 0) \tag{2}$$

[†] Faculty of Mathematics and Computer Science.
[‡] Faculty of Chemistry and Chemical Engineering.

10.1021/ci990328f CCC: $19.00
Published on Web 11/18/1999

for $j = 1, ..., N$ and $k = 0, ..., m - 1$. Here $s_m^{(k)}(X_j - 0)$ means the limit from the left side of X of the kth derivative of s_m, while $s_m^{(k)}(X_j + 0)$ denotes the same for the limit to the right in x, while $s_m^{(0)}$ stands for the function itself.

Splines of degree m can be generally written in the form

$$s_m(x) = \sum_{i=0}^{m} \alpha_i x^i + \sum_{j=1}^{N} \beta_j (x - X_j)_+^m \quad (3)$$

where α_i, $i = 0, ..., m$, and β_j, $j = 1, ..., N$, are real numbers to be determined (similar to the regression parameters). Here $(y)_+$ stands for the maximum between y and 0. We have to emphasize here that the common polynomials—and particularly linear functions—are particular cases of splines, obtained for $N = 0$. Then the second sum in (3) disappears.

The usual way of determining the above parameters consists in imposing some interpolatory conditions, namely, forcing the spline to take some prescribed values in knots. But, compared to the number of degrees of freedom ($n \gg N + m + 1$), the data sets are large, so it is difficult to establish the suitable interpolatory conditions. Moreover, interpolation makes sense only in connection with extremely precise measurements. Our approach is based on regression, where the spline, s_m, is sought so that it approximates as well as possible all the data.

In this process, some of the measurements may be more important to others, so it is natural to assign different weights to them. Concretely, if, for each $k = 1, ..., n$, r_k^2 denotes the squared residuum related to the data (x_k, y_k), namely

$$r_k^2 = (y_k - s_m(x_k))^2 = (y_k - \sum_{i=0}^{m} \alpha_i x_k^i - \sum_{j=1}^{N} \beta_j (x_k - x_j)_+^m)^2 \quad (4)$$

a positive weight $\omega_k \geq 0$ can be associated with it. These weights are important not as absolute values, but rather in connection with the others, so they can be scaled to satisfy the equality

$$\sum_{k=1}^{n} \omega_k = 1 \quad (5)$$

In particular, these weights may all be equal ($\omega_k = 1/n$). Now the parameters α_i, $i = 0, ..., m$, and β_j, $j = 1, ..., N$, or the approximating spline itself are obtained by minimizing the global weighted squared residual

$$R^2 = \sum_{k=1}^{n} \omega_k r_k^2 \rightarrow \min \quad (6)$$

which is equivalent to

$$\sum_{k=1}^{n} \omega_k (y_k - \sum_{i=0}^{m} \alpha_i x_k^i - \sum_{j=1}^{N} \beta_j (x_k - x_j)_+^m)^2 \rightarrow \min \quad (7)$$

If the above weights are assumed known a priori—depending eventually on the measurements but not on the regression results themselves—a weighted, generalized, but still linear regression is applied. This is the situation of the first three weighting approaches described below. Differentiating upon the unknown parameters—α_i, $i = 0, ..., m$, and β_j, $j = 1, ..., N$—the linear system of normal equations is obtained, and for solving it there are several methods available in the literature.[25]

It is also possible to consider weights depending on the calibration function itself, or its coefficients, as done in the last two cases, i.e.

$$\omega_k = \omega_k(s_m) = \omega_k(\alpha_0, ..., \alpha_m, \beta_1, ..., \beta_N), \quad k = 1, ..., n \quad (8)$$

As a consequence in these cases the regression becomes nonlinear, and its parameters are determined by applying an iterative method, so that the nonlinear problem is reduced to a sequence of linear ones. Starting with an initial value ω^0, i.e., $\omega_1^0 = \omega_2^0 = ... = \omega_n^0 = 1/n$, and applying the weighted linear regression, a first set of parameters $(\alpha^1, \beta^1) = (\alpha_0^1, ..., \alpha_m^1, \beta_1^1, ..., \beta_N^1)$ is obtained and, consequently, a calibration function s_m^1. Using this spline, a new n-uple of weights ω^1 can be generated. The concrete procedure for obtaining the new weights depends on the weighting technique itself, i.e., the definition of the weights. Two examples are considered later in this paper, but here we stay rather at the abstract level.

The new weights are involved in determining another calibration spline s_m^2, which leads to the weights ω^2 and so on. Generally, if $\omega^{k-1} = (\omega_1^{k-1}, ..., \omega_n^{k-1})$ represents the weights computed after $k - 1$ steps ($k > 2$), a linear regression leads to the parameters $(\alpha^k, \beta^k) = (\alpha_0^k, ..., \alpha_m^k, \beta_1^k, ..., \beta_N^k)$, or the calibration spline

$$s_m^k(x) = \sum_{i=0}^{m} \alpha_i^k x^i + \sum_{j=1}^{N} \beta_j^k (x - X_j)_+^m \quad (9)$$

Briefly, the discussion above can be comprised in the following scheme:

$$\omega^0 \rightarrow (\alpha^1, \beta^1) \rightarrow \omega^1 \rightarrow (\alpha^2, \beta^2) \rightarrow ... \rightarrow \omega^k \rightarrow (\alpha^{k+1}, \beta^{k+1}) \rightarrow ...$$

The computational process is stopped at the kth iteration if the relative variation of the parameters

$$\left|\frac{\alpha_i^k - \alpha_i^{k-1}}{\alpha_i^k}\right| \leq \epsilon, \quad \left|\frac{\beta_j^k - \beta j^{k-1}}{\beta_j^k}\right| \leq \epsilon, \quad 0 \leq i \leq m, \quad 1 \leq j \leq N \quad (10)$$

where ϵ is an admissible error.

For avoiding a situation of divergence, our stopping criterion relies rather on the weights, which—recalling their positiveness and the identity in (5)—are included in a compact set. This ensures a gain in stability for the regression algorithm. Since the weighted linear spline regression problem has a unique solution,[8] if in the iterative process described before two succesive n-uples of weights (ω_k and ω_{k+1}) are equal, the parameters generated by them are also the same. Therefore, the calibration spline generated at each of the forthcoming steps will always be the same and may be assimilated to the solution, so it does not make sense to iterate anymore. From this point of view, iterations are stopped when two successive n-uples of weights are close to one another, namely

$$[\sum_{i=1}^{n}(\omega_i^{k} - \omega_i^{k-1})^2]^{1/2} \leq \epsilon \quad (11)$$

However, since the regression problem is nonlinear, it may have more than one solution.[27] Hence, each iteration could approximate one of the solutions, while the following goes in the neighborhood of another one determining a cyclical process. This is why NRIT—a maximum number of iterations—has been considered.

It is worth mentioning here the following aspect concerning the iterative process described above. The approach is quite simple and reduces the nonlinear regression problem to some linear ones. There are many other possibilities provided by the optimization theory, which may behave better than this one. However, the solution obtained may be a local minimum instead of a desired global one. To our knowledge, all the minimization procedures available in the literature cannot guarantee that a global minimum is reached, especially when there are more than one solution, which is the case here.

Before proceeding with the choice of some weights, we should emphasize that the usual regression—where a linear function is sought—is a particular case of the approach presented above. This is obtained for $m = 1$ and $N = 0$.

WEIGHTING TECHNIQUES

The choice of a particular procedure of weighting depends on the structure of data used in the calibration process. In the case of homoscedastical data and no outliers, any measurement may have the same importance for the calibration. Then the weights are all equal, namely

$$\omega_k = 1/n, \quad k = 1, ..., n \quad (12)$$

and in fact this is similar to the unweighted case, where all weights are equal to 1.

If the data are heteroscedastical, proper weighting techniques should be applied. One possibility is based on the variance of the data (signals) measured at a constant concentration x_k,[2,26]

$$s_k^{2} = \frac{1}{n_k - 1}\sum_{i=1}^{n_k}(y_i - \bar{y})^2 \quad (13)$$

where $\bar{y}$ is the mean of the n_k signals obtained at the given concentration, assuming that there are more than one. In this case, the weights are calculated as

$$\omega_k = c/s_k^{2}, \quad k = 1, ..., n \quad (14)$$

with $c = \sum_{k=1}^{n}(1/s_k^2)$ being the scaling factor.

Depending on the quality of the measurements, other weighting techniques should also be considered. When the errors increase with increasing concentration, the weights can be taken proportional to the inverse of the squared concentration[26]

$$\omega_k = c/x_k^{2}, \quad k = 1, ..., n, \quad c = \sum_{k=1}^{n}(1/x_k^{2}) \quad (15)$$

In the definitions given in (12)−(15) only the data are involved, not the calibration result itself. Therefore, the regression is still linear. In the following approaches the weights depend on the position of the data relative to the calibration function, hence on the regression parameters, as follows from the formula

$$\omega_k = \omega_k(\alpha_0, ..., \alpha_m, \beta_1, ..., \beta_N), \quad k = 1, ..., n \quad (16)$$

The regression is nonlinear, and for determining the parameters effectively the iterative procedure presented in the previous section is applied.

In both cases we start with the graphical representation of the data samples in the concentration−signal plane. For each point the distance to the graphic of the calibration function is considered. If the measurements are done for standard concentrations, there will be less uncertainty in the x-direction, i.e., error(Y) ≫ error(X). In this case, the distance of any point (x_k, y_k) to the function (9) will have the following expression:

$$d_k^{y} = |y_k - s_m(x_k)| = |y_k - \sum_{i=0}^{m}\alpha_i x_k^{i} - \sum_{j=1}^{N}\beta_j(x_k - X_j)_+^{m}| \quad (17)$$

where k ranges from 1 to n and $|y|$ is the absolute value of y.

If errors appear also in the concentration direction, the Euclidean distance can be considered

$$d_k = \mathrm{dist}((x_k, y_k), s_m) \quad (18)$$

The distance dist((x_k, y_k), s_m) results in the square root of the minimum of the expression

$$(x - z)^2 + (y - s_m(z))^2 = (x - z)^2 + (y - \sum_{i=0}^{m}\alpha_i z^{i} - \sum_{j=1}^{N}\beta_j(z - X_j)_+^{m})^2, \quad z \in (a, b) \quad (19)$$

However, the computation of this distance is not trivial if the splines are nonlinear. For approximating it, we have considered a uniform grid and selected from these points the one closest to the sample (x_k, y_k).

Having determined the above distances, the samples closer to the graphic of the calibration function are considered more relevant, as results from

$$\omega_k = c/d_k^{y}, \quad k = 1, ..., n \quad (20)$$

and

$$\omega_k = c/d_k, \quad k = 1, ..., n \quad (21)$$

The constant c in the above formulas is included for normalization ($c = 1/\sum_{k=1}^{n}\omega_k$).

In any of the two approaches above the outliers become irrelevant. From the practical point of view, to avoid division

by 0, if one distance lies below a threshold value ϵ, c/ϵ (a value large enough) is assigned to the corresponding weight.

QUALITY COEFFICIENTS

To evaluate the correctness of the calibration models generated in this paper, employing splines, we can compute some statistics or different quality coefficients (QCs) used in the analytical literature to judge the goodness of fit of a regression function.[1,26] The variance of estimate values, namely

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(y_k - s_m(x_k))^2 \tag{22}$$

offers primary information regarding the calibration results.

Another possibility is given by the values of the quality coefficient QC_1, defined as

$$QC_1 = \left[\frac{1}{n-1}\sum_{k=1}^{n}\left(\frac{y_k - s_m(x_k)}{s_m(x_k)}\right)^2\right]^{1/2} \times 100\% \tag{23}$$

Similarily, QC_2 may be used where in the above formula y_k replaces in the denominator the estimate $s_m(x_k)$, or QC_3 and QC_4, with the same modification referring to the mean signal $\bar{y}$, instead of the signal itself, respectively, to the mean of the estimated signal, $\bar{s}_m$.

All the above coefficients give a measure of the quality of the data approximation through the regression function. The smaller the values are, the better the model is QC_1 and QC_2 are related to the goodness of fit of the regression model, while QC_3 and QC_4 appear to be more useful for comparing results provided by the same algorithm. However, since the information obtained using these coefficients is sometimes contradictory, two other coefficients were proposed in refs 1 and 26

$$QC_5 = \left[\sum_{k=1}^{n}\left(\frac{r_k}{r_{\max}}\right)^2\right]^{1/2}, \quad QC_6 = \left[\sum_{k=1}^{n}\left(\frac{r_k}{\bar{r}}\right)^2\right]^{1/2} \tag{24}$$

Here r_k stands for the residual corresponding to the kth sample ($r_k = y_k - s_m(x_k)$), $r_{\max}$ refers to the maximum of absolute residuals ($r_{\max} = \{|r_k|, k = 1, ..., n\}$), while $\bar{r}$ denotes the mean of the absolute residuals ($\bar{r} = (1/n)\sum_{k=1}^{n} r_k$). Both the coefficients are proportional to the total squared residual. Moreover, the following inequalities hold true:

$$1 \le QC_5 \le n^{1/2}, \quad n^{1/2} \le QC_6 \le n \tag{25}$$

In addition, the estimates are sharp, since the equalities appear in a particular case ($r_k = r$ for all k, and $r_1 = r$ and $r_k = 0$ for $k > 1$). As a consequence the above coefficients can be normalized to [0, 1], namely

$$NQC_5 = \frac{QC_5 - 1}{n^{1/2} - 1}, \quad NQC_6 = \frac{QC_6 - n^{1/2}}{n - n^{1/2}} \tag{26}$$

Normalization is important in practice, since it allows a simpler comparison between different regression approaches. It is easy to see that QC_5—and therefore NQC_5—is minimal if outliers are present in the calibration process. On the other hand, if the residuals are of the same order of magnitude, these coefficients tend to the maximal values. In the case of QC_6 and NQC_6 the situation is reversed; i.e., lower values mean equal residuals, while outliers are suggested by some significantly larger residuals.

It is worth noticing here that all the above formulas should be modified in the case of weighted regression by considering weighting factors. The residual variance, for example, takes the form

$$s_w^2 = \sum_{k=1}^{n}\omega_k(y_k - s_m(x_k))^2 \tag{27}$$

where ω_k replaces the ratio $1/(n-1)$ in the original form. In fact, for a better compatibility with the unweighted version, $n/(n-1)$ should multiply each term in the sums above, but this is nonsignificant for large data sets. Similarly, QC_5 and QC_6 become

$$QC_5^{\omega} = \left[n\sum_{k=1}^{n}\omega_k\left(\frac{r_k}{r_{\max}}\right)^2\right]^{1/2}, \quad QC_6^{\omega} = \left[n\sum_{k=1}^{n}\omega_k\left(\frac{r_k}{\bar{r}}\right)^2\right]^{1/2} \tag{28}$$

In (28), the sums are multiplied by the number of samples—n, again, for compatibility with the unweighted versions. The range of variation for QC_5^{ω} can be easily determined

$$0 \le QC_5^{\omega} \le n^{1/2} \tag{29}$$

while for the QC_6^{ω} only a lower bound is available

$$n^{1/2} \le QC_6^{\omega} \tag{30}$$

The normalized versions of QC_5^{ω} and QC_6^{ω} become

$$NQC_5^{\omega} = \frac{QC_5^{\omega}}{n^{1/2}} = \left[\sum_{k=1}^{n}\omega_k\left(\frac{r_k}{r_{\max}}\right)^2\right]^{1/2} \tag{31}$$

and

$$NQC_6^{\omega} = \frac{QC_6^{\omega} - n^{1/2}}{QC_6^{\omega}} \tag{32}$$

The interpretation of weighting coefficients is similar to the unweighted ones. However, weights play an important role here too. For example, if the last two weighting techniques described in the previous section are considered, outliers become almost insignificant in the calibration process. Therefore, the weights assigned to them go to 0, so that the corresponding extreme values for the quality coefficients stand out.

RESULTS AND DISCUSSION

The approaches to be used are general, but they will be illustrated with reference to a specific example. Table 1 provides a set of data that was obtained in an experiment designed to determine simazine in water by TLC using a Shimadzu CS-9000 dual-wavelength flying spot densitometer, with a reflectance mode at 225 nm.

Table 1. Data for Simazine[a]

c (ng)	s_1	s_2	s_3	s_4	s_5	s_6
200	10775	11410	8303	9919	9214	10287
300	12906	13145	11715	11686	10380	11866
400	14544	13875	14157	14629	14378	15069
500	18199	20114	19070	21186	20678	17218
600	24412	25233	24452	23489	24215	24104
700	28514	30584	28887	28538	26435	24612
800	30647	31347	28786	27685	29693	28081
900	29870	29841	30809	30727	28837	28488
1000	32270	31756	30996	31906	32312	32135

[a] Concentration vs response peaks.

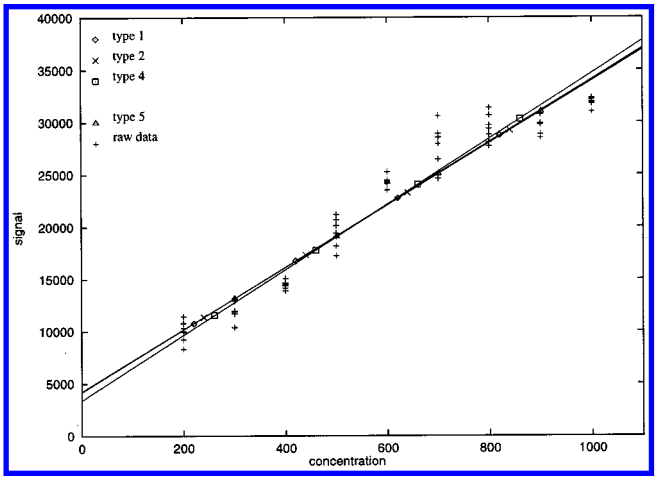

Figure 1. Linear spline calibration ($m = 1$), zero knots.

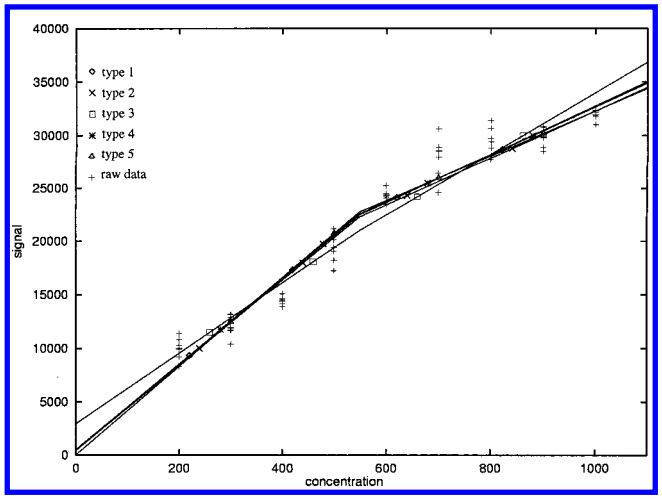

Figure 2. Linear spline calibration ($m = 1$), one knot ($N = 1$, $X_1 = 600$).

The calibration functions were obtained by scanning zones via the peak area (six readings), from 1 to 10 μL of TLC standard (containing 100−1000 ng of simazine). The program was written in standard C, and GNUPLOT has been used for the visualization of the calibration functions and calculating the residual standard deviation and quality coefficients.

Although the spline theory works for any m and N, for practical purposes linear or cubic splines are recommended, eventually parabolic ones. So the degree of splines should not be too large. The number, N, of knots depends on the number of points. Five to fifty points per knot seems to be a suitable choice, but the displacement of data should also be taken into account. A good idea would be to find the optimal number and position of knots automatically—applying, e.g., spline functions with free knots,[28] but this problem lies beyond the aim of our work.

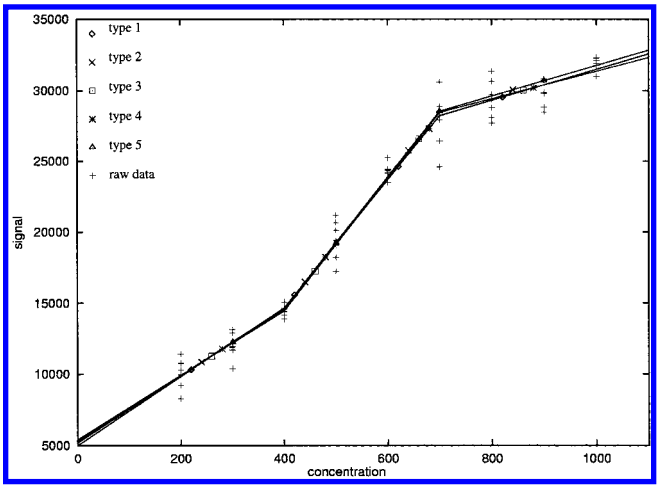

Figure 3. Linear spline calibration ($m = 1$), two knots ($N = 2$, $X_1 = 400$, $X_2 = 700$).

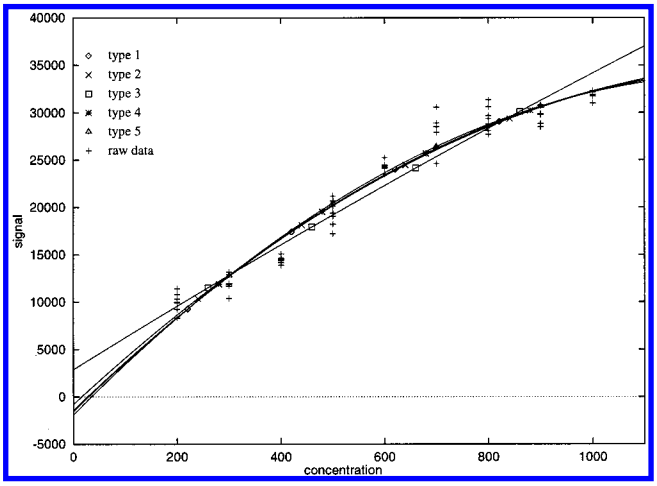

Figure 4. Parabolic spline calibration ($m = 2$), zero knots.

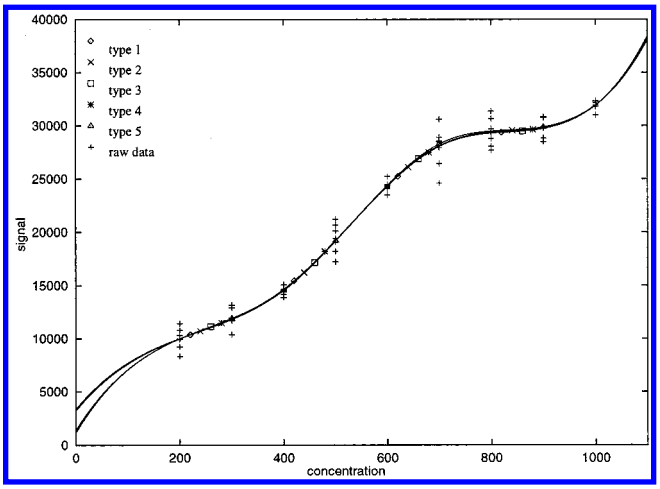

Figure 5. Cubic spline calibration ($m = 3$), two knots ($N = 2$, $X_1 = 400$, $X_2 = 700$).

In this paper linear, parabolic, or cubic splines have been considered, while N was taken as 0, 1, or 2. In the case of a single knot the corresponding value was $X_1 = 550$ ng, while for two knots the values were $X_1 = 400$ ng and $X_2 = 700$ ng. We have chosen these values since the points are

Table 2. Calibration Results, Unweighted Coefficients

type	s	QC_1	QC_2	QC_3	QC_4	QC_5	NQC_5	QC_6	NQC_6
Linear Splines ($m = 1$), Zero Knots ($N = 0$)									
1	1997	9.54	9.82	9.12	9.12	268	0.265	875	0.030
2	2029	9.32	9.30	9.25	9.26	281	0.285	863	0.028
3	2000	9.66	9.96	9.15	9.13	264	0.259	881	0.031
4	1999	9.59	9.83	9.15	9.12	264	0.259	881	0.032
5	1999	9.59	9.83	9.15	9.12	264	0.259	881	0.032
Linear Splines ($m = 1$), One Knot ($N = 1$, $X_1 = 600$)									
1	1686	10.28	9.78	7.70	7.70	265	0.261	914	0.038
2	1880	9.11	9.13	8.59	8.58	263	0.257	866	0.028
3	1687	10.36	9.86	7.70	7.70	266	0.263	921	0.040
4	1702	10.95	10.17	7.80	7.77	269	0.267	952	0.047
5	1712	10.61	9.74	7.91	7.81	253	0.242	945	0.045
Linear Splines ($m = 1$), Two Knots ($N = 2$, $X_1 = 400$, $X_2 = 700$)									
1	1145	6.09	6.32	5.23	5.23	232	0.208	934	0.043
2	1152	6.04	6.32	5.26	5.26	220	0.189	955	0.047
3	1145	6.09	6.32	5.23	5.23	233	0.211	932	0.042
4	1164	6.05	6.39	5.28	5.31	217	0.186	994	0.056
5	1164	6.06	6.42	5.27	5.31	218	0.186	991	0.055

uniformly distributed, the concentrations ranging from 200 to 1000 ng.

The results discussed below refer to the following five weighting strategies: (type 1) uniform weights ($\omega_k = c/n$); (type 2) weights depending on the concentration ($\omega_k = c/x_k^2$); (type 3) weights depending on the variance ($\omega_k = c/s_k^2$); (type 4) y-distance weighting ($\omega_k = c/|y_k - s_m(x_k)|$); (type 5) distance-based weighting ($\omega_k = c/\mathrm{dist}((x_k, y_k), s_m)$). The last two weighting techniques lead to a nonlinear regression problem. For solving it, the iterative procedure described in the Theoretical Considerations has been applied. An approximation of a solution was obtained generally in two to five steps; thus, the computational effort is not significantly higher. However, a maximal number of iterations (NRIT = 15) was admitted, to avoid a cyclical process—as described in the above-mentioned section.

The graphs are displayed in Figures 1−5. For a more rational comparison we will refer to the quality coefficients. Tables 2−7 present the values of the residual standard deviation and the quality coefficients, in both the unweighted and the weighting alternatives. The tables reflecting the weighting versions do not include the type 0 weighting, since those results were already included in the tables for the unweighted cases. Moreover, in the computation of the weighted quality coefficients, the weights involved here are the same as those appearing in the calibration procedure.

By examining the data in Tables 2−7, we can see that the residual standard deviation decreases with the increasing number of knots and the degree of splines. This trend is more pregnant for the weighted version of splines, the best values being obtained for linear and parabolic splines with two knots. Values close to 0 for the first two coefficients are obtained for binodal splines by using the distance as the weighting factor. Similar results are obtained for QC_3, QC_3^{ω}, QC_4, and QC_4^{ω} in the case of y-distance weighting. This

Table 3. Calibration Results, Weighted Coefficients

type	s^{ω}	QC_1^{ω}	QC_2^{ω}	QC_3^{ω}	QC_4^{ω}	QC_5^{ω}	NQC_5^{ω}	QC_6^{ω}	NQC_6^{ω}
Linear Splines ($m = 1$), Zero Knots ($N = 0$)									
2	1478	10.17	10.29	10.30	10.30	204	0.279	886	0.171
3	1914	9.42	10.10	8.50	8.79	253	0.344	792	0.073
4	194	1.00	1.04	1.44	1.44	25.76	0.035	6289	0.883
5	194	1.00	1.04	1.44	1.44	25.76	0.035	6289	0.883
Linear Splines ($m = 1$), One Knot ($N = 1$, $X_1 = 600$)									
2	1465	10.47	10.51	10.20	10.20	205	0.279	876	0.162
3	1560	9.58	10.24	6.92	7.17	246	0.336	871	0.157
4	205	1.32	1.25	0.71	0.71	32.55	0.044	4688	0.843
5	36	0.25	0.23	0.11	0.11	5.34	0.007	27659	0.973
Linear Splines ($m = 1$), Two Knots ($N = 2$, $X_1 = 400$, $X_2 = 700$)									
2	1033	8.22	8.65	7.20	7.20	197	0.269	928	0.208
3	771	4.24	4.38	3.56	3.54	157	0.214	925	0.206
4	1.48	0.01	0.01	0.01	0.01	0.28	0.000	426634	0.998
5	1.42	0.01	0.01	0.01	0.01	0.27	0.000	474682	0.998

Table 4. Calibration Results, Unweighted Coefficients

type	s	QC_1	QC_2	QC_3	QC_4	QC_5	NQC_5	QC_6	NQC_6
Parabolic Splines ($m = 2$), Zero Knots ($N = 0$)									
1	1653	10.73	10.10	7.55	7.55	276	0.279	919	0.040
2	1926	9.20	9.21	8.79	8.79	270	0.268	861	0.027
3	1654	10.82	10.17	7.55	7.55	277	0.280	922	0.040
4	1661	11.02	10.47	7.55	7.58	292	0.303	932	0.042
Parabolic Splines ($m = 2$), One Knot ($N = 1$, $X_1 = 600$)									
1	1262	6.49	6.61	5.76	5.76	285	0.292	887	0.033
2	1284	6.39	6.57	5.86	5.86	310	0.332	901	0.036
3	1264	6.53	6.64	5.77	5.77	277	0.280	885	0.032
4	1270	6.40	6.54	5.80	5.80	300	0.316	901	0.036
5	2250	9.75	9.12	10.61	10.27	266	0.263	1117	0.082
Parabolic Splines ($m = 2$), Two Knots ($N = 2$, $X_1 = 400$, $X_2 = 700$)									
1	1247	6.83	6.95	5.69	5.69	262	0.255	923	0.041
2	1289	6.60	6.77	5.88	5.88	273	0.274	889	0.033
3	1247	6.86	6.97	5.70	5.69	259	0.251	927	0.041
4	1256	6.97	7.11	5.73	5.73	258	0.249	954	0.047
5	1585	8.34	9.16	7.34	7.23	331	0.365	922	0.040

Table 5. Calibration Results, Weighted Coefficients

type	s^{ω}	QC_1^{ω}	QC_2^{ω}	QC_3^{ω}	QC_4^{ω}	QC_5^{ω}	NQC_5^{ω}	QC_6^{ω}	NQC_6^{ω}
Parabolic Splines ($m = 2$), Zero Knots ($N = 0$)									
2	1473	10.41	10.48	10.26	10.26	206	0.281	878	0.163
3	1606	10.10	10.79	7.15	7.38	269	0.367	904	0.188
4	245	1.59	1.52	2.10	2.10	43	0.059	3917	0.812
5	994	6.69	6.54	4.30	4.30	161	0.220	1117	0.342
Parabolic Splines ($m = 2$), One Knot ($N = 1$, $X_1 = 600$)									
2	1067	8.29	8.68	7.43	7.43	258.08	0.351	917	0.199
3	983	5.27	5.46	4.51	4.51	215.83	0.294	860	0.146
4	167	0.90	0.91	1.57	1.57	39.56	0.054	4545	0.838
5	716	3.70	3.65	3.72	3.72	84.93	0.116	1192	0.384
Parabolic Splines ($m = 2$), Two Knots ($N = 2$, $X_1 = 400$, $X_2 = 700$)									
2	1101	8.53	8.92	7.67	7.67	233	0.318	897	0.181
3	863	5.11	5.22	3.95	3.97	179	0.245	916	0.198
4	0.73	0.00	0.00	0.01	0.01	0.15	0.000	973145	0.999
5	711	4.01	4.00	3.84	3.84	148	0.202	1059	0.306

Table 6. Calibration Results, Unweighted Coefficients

type	s	QC_1	QC_2	QC_3	QC_4	QC_5	NQC_5	QC_6	NQC_6
Cubic Splines ($m = 3$), Zero Knots ($N = 0$)									
1	1360	6.98	7.04	6.21	6.21	279	0.282	864	0.028
2	1400	6.74	6.92	6.39	6.39	316	0.341	878	0.031
3	1362	7.07	7.11	6.23	6.22	271	0.270	863	0.028
4	1361	6.97	6.98	6.22	6.21	278	0.281	866	0.028
5	2402	10.49	9.75	11.23	10.96	262	0.257	1140	0.087
Cubic Splines ($m = 3$), One Knot ($N = 1$, $X_1 = 600$)									
1	1202	6.69	6.83	5.49	5.49	264	0.259	929	0.042
2	1246	6.47	6.63	5.69	5.69	270	0.268	899	0.035
3	1202	6.73	6.85	5.49	5.49	262	0.257	931	0.042
4	1206	6.74	6.89	5.50	5.51	260	0.254	946	0.045
5	1290	7.67	8.79	5.86	5.89	271	0.270	954	0.047
Cubic Splines ($m = 3$), two Knots ($N = 2$, $X_1 = 400$, $X_2 = 700$)									
1	1088	5.89	6.10	4.97	4.97	232	0.209	981	0.053
2	1088	5.88	6.10	4.97	4.97	229	0.203	983	0.053
3	1088	5.88	6.10	4.97	4.97	232	0.209	981	0.053
4	1094	5.89	6.12	4.98	4.99	219	0.189	1006	0.058
5	1090	5.90	6.10	4.97	4.98	229	0.204	994	0.056

Table 7. Calibration Results, Weighted Coefficients

type	s^{ω}	QC_1^{ω}	QC_2^{ω}	QC_3^{ω}	QC_4^{ω}	QC_5^{ω}	NQC_5^{ω}	QC_6^{ω}	NQC_6^{ω}
Cubic Splines ($m = 3$), Zero Knots ($N = 0$)									
2	1112	8.42	8.79	7.75	7.75	251	0.342	909	0.192
3	1151	6.27	6.53	5.27	5.29	229	0.312	827	0.112
4	781	4.27	4.29	3.44	3.44	159	0.217	1085	0.323
5	722	3.50	3.48	3.82	3.82	79	0.108	1293	0.432
Cubic Splines ($m = 3$), One Knot ($N = 1$, $X_1 = 600$)									
2	1083	8.49	8.88	7.55	7.55	235	0.320	901	0.185
3	799	4.84	4.93	3.65	3.67	174	0.238	937	0.225
4	30	0.18	0.18	0.30	0.30	6.68	0.009	22269	0.967
5	594	3.15	3.13	2.48	2.47	125	0.170	1243	0.409
Cubic Splines ($m = 3$), Two Knots ($N = 2$, $X_1 = X_1 = 400$, $X_2 = 700$)									
2	1009	8.14	8.57	7.03	7.03	212	0.289	944	0.222
3	671	3.94	4.08	3.07	3.08	143	0.195	1023	0.282
4	24	0.14	0.14	0.20	0.20	4.82	0.007	24469	0.970
5	564	3.18	3.24	2.63	2.63	118	0.162	1302	0.436

weighting technique appears to be more appropriate according also to the last quality coefficients.

CONCLUDING REMARKS

The discussion above justifies the fact that splines can be successfully used for calibration in analytical chemistry and other fields of science. Different weighting techniques have also been considered, and for the evaluation of the results some quality coefficients were proposed. Depending upon the structure of the data, some weighting procedures could improve the results dramatically. Considering the results obtained in the case of TLC densitometry, it seems that both nonlinear weighting procedures are the best ones, with a plus for the *y*-distance type. It is difficult to give some general rules regarding the optimal parameters of the weighted calibration spline (*m*, *N*, the distribution of knots, and the weighting factor). The optimal values depend upon the structure of the data. However, higher order splines are not recommended because the result may become extremely unstable. The example used to illustrate the performances of the

procedures discussed in this paper involved only a single independent variable. However, the method is general, and practically, any number of variables can be used to built a spline function.

REFERENCES AND NOTES

(1) Pop, F. H.; Sârbu, C. A new fuzzy regression algorithm. *Anal. Chem.* **1996**, *68*, 771−778.
(2) Miller, J. N. Basic statistical methods for analytical chemistry, Part 2. Calibration and regression methods - A review. *Analyst* **1991**, *116*, 3−14.
(3) Meloun, M.; Mlitky, J.; Forina, M. *Chemometrics for Analytical Chemistry, vol I: PC−aided statistical data analysis*; Ellis Horwood: Chichester, 1989.
(4) Klimov, G. *Probability Theory and Mathematical Statistics*; Mir Publishers: Moscow, 1986.
(5) Wold, S. Spline functions in data analysis. *Technometrics* **1974**, *16*, 1−11.
(6) Eubank, R. L. *Spline Smoothing and Nonparametric Regression*; Marcel Dekker: New York, 1988.
(7) Adams, J. M. *Chemometrics in Analytical Chemistry*; The Royal Society of Chemistry: Cambridge, 1995; pp 34−54.
(8) Pop, I. S. Weighted least-squares splines. *Res. Seminars Prepr., "Babeş-Bolyai" Univ. Cluj-Napoca* **1996**, *1*, 79−91.
(9) Pop, V. Applications of Spline Functions in Analytical Chemistry. Thesis, "*Babeş-Bolyai*" *Univ. Cluj-Napoca* 1998.
(10) Hazi, J.; Daruhazi, L.; Pinter, O.; Szakaly, Z. Critical comparison of evaluation methods of potentiometric titration data using a microcomputer-controlled automatic potentiometric titration system. *Mikrochim. Acta* **1986**, *2*, 379−388.
(11) Greschonig, H.; Glatter, O. Determination of equivalence points of sigmoidal potentiometric titration curves. *Mikrochim. Acta* **1986**, *2*, 401−409.
(12) Ren, K.; Ren-Kurc A. A numerical method of finding potentiometric end-points by use of approximative spline functions. *Talanta* **1990**, *37*, 667−671.
(13) Bozdogan, A. Determination of the end points in titration of mixtures of acids using piecewise cubic splines and comparing with the intersecting straight lines. *Turk. J. Chem.* **1992**, *16*, 112−122.
(14) Frigieri, P. F.; Rossi, F. B. Calibration of photographic emulsions by cubic spline functions and application in spectrochemical analysis. *Anal. Chem.* **1979**, *51*, 54−57.
(15) Gans, P. Improved methods for numerical differentiation of spectroscopic curves. *Anal. Proc. (London)* **1980**, *17*, 133−135.
(16) Klein, G. Numerical reconstruction of fine structure in obscured backscattering spectra. *J. Radioanal. Chem.* **1982**, *75*, 237−244.
(17) Gans, P.; Gill, J. B. Smoothing and differentiation of spectroscopic curve using spline functions. *Appl. Spectrosc.* **1984**, *38*, 370−376.
(18) Losev, A. Smoothing by spline functions: applications in electron spectroscopy. *J. Electron. Spectrosc. Relat. Phenom.* **1990**, *50*, C19−C23.
(19) Saitou, N.; Iida, A.; Gohshi, Y. Data processing in x-ray fluorescence spectroscopy. I. A smoothing method using B-splines. *Spectrochim. Acta, Part B* **1983**, *38B*, 1277−1285.
(20) Schulze, D. G. Correction of mismatches in 2.theta.scales during differential x-ray diffraction. *Clays Clay Miner.* **1986**, *34*, 681−685.
(21) Malitesta, C.; Rotunno, T. Quantitative resolution of x-ray photoelectron spectra of mixtures of chromium compounds by the Kalman filter after cubic spline background removal. *Surf. Interface Anal.* **1991**, *17*, 251−258.
(22) Lunga, G. D.; Basosi, R. A simple method for baseline correction in EPR spectroscopy. 2. The use of cubic spline functions. *J. Magn. Reson., Ser. A* **1995**, *112*, 102−105.
(23) Cataldi, T. R. I.; Rotunno, T. Kalman filter optimized cubic spline functions. *Anal. Methods Instrum.* **1995**, *2*, 27−34.
(24) Ferreira, M. M. C.; Ferreira, W. C.; Kowalski, B. R. Rank determination and analysis of non-linear processes by global linearizing transformation. *J. Chemom.* **1996**, *10*, 11−30.
(25) Stoer, J.; Bulirsch, R. *Introduction to Numerical Analysis*; Springer-Verlag: New York, 1996.
(26) Sârbu, C. A comparative study of regression concerning weighted least-squares methods. *Anal. Lett.* **1995**, *28*, 2077−2094.
(27) Björck, A. *Numerical Methods for Least Squares Problems*; SIAM: Philadelphia, 1996.
(28) Nürnberger, G. *Approximation by spline functions*; Springer-Verlag: Berlin, 1989.

CI990328F