Formation Constants of Molecular Complexes from Spectrophotometric Data. Experimental Design for Concurrent Models Discrimination and Optimal Determination of the Parameters

Gino Carta
Dipartimento di Chimica e Tecnologie Inorganiche e Metallorganiche, Università di Cagliari, Via Ospedale 72, 09124 Cagliari, Italy

A typical experimental procedure as is now established collects all the data together at the beginning of the study, before the true model is known. The resulting design (*i.e.* all the available experimental points) may not be very effective for discriminating between the tentative models initially assumed, and/or not very appropriate for the optimal determination of the parameters of the model finally chosen. Furthermore, in a non-linear situation, as in our case, the experiment cannot be optimally designed *a priori*, even if the model is known. An alternative experimental procedure is proposed, in which the observations are added sequentially, chosen on the basis of the task currently requested. The first and central part of the experiment is mainly devoted to the discrimination among the initial competing models. In the final phase, when hopefully only the correct model remains, the observations are added (still sequentially) to maximize the determinant of the information matrix |X'X|, until the parameter variances are sufficiently low to stop the experiment.

Spectrophotometric methods have been widely used for the determination of association constants of molecular complexes.[1–8] Until the late sixties the approach was mainly graphic-oriented; the problem was generally assumed to be 1:1 and was linearized by suitable mathematical simplifications and experimental limitations.[9–12] The obtained values of K (formation constant) and ε (extinction coefficient) were often not reproducible.[3] The explanation was sometimes sought either in a poor design (Person)[13] or in the inadequacy of the model.[3,14,15] With Derenleau[16] the conditions to be met to have both model check and good determination of the parameters in the 1:1 case were clearly stated.

Elsewhere, a different computer-oriented approach was taking place.[17–19] With it the mathematical and experimental limitations used till then were no longer necessary, as the experimental points (*i.e.* the design) could be chosen freely, subject only to the system constraints (solubility of reagents, constancy of activity coefficients, region experimentally attainable, *etc.*).

A typical computer-oriented procedure as established today could be summarized in the following steps.

(1) Collect the experimental points. This is probably done by keeping constant some different initial concentrations of a reactant, say A, while varying the initial concentration of the other reactant B, and/or *vice versa* (titration curves).[20] Generally, different wavelengths are used.

(2) Estimate the number (and hopefully the nature) of the species in solution. To this end either graphical and/or numerical methods (matrix rank) may be employed.[21,22,23a]

(3) Use the information from point (2) and any possible previous knowledge or chemical intuition and set up all plausible models.

(4) Choose a suitable least-squares routine and for each plausible model calculate all the parameters, their standard deviations and some statistical indices.[21,24]

(5) Compare the statistical indices obtained in the preceding point (4) and select the most probable model. If the procedure does not give an unequivocal result then it has been suggested that the choice should be made by applying the following considerations in order of priority:[24]

(i) simplicity—the simplest model should be preferred by excluding poorly-defined species;

(ii) similarity to previous results either on the same or on analogous systems;

(iii) chemical significance.

The weakness of this approach lies mainly in the fact that experimental points are collected uniquely at the beginning, before the true model (or, say, the model best representing the data) is known. As will be clarified later, these experimental points may not be effective for discriminating between the tentative, concurrent models initially proposed, and/or not very good for the optimal determination of the parameters of the model finally chosen.

In a preceding paper[25] a sequential method was described which, for a model known *a priori*, indicates how the experimental points could be chosen in order to have the best determination of the parameters. In practice the model is not known in advance and must be determined from the experiment. Hence the need for an experimental design which allows us to discriminate efficiently among rival competing models and can simultaneously enable an optimal parameter determination. An optimal determination experiment, for non-linear models, cannot be designed *a priori* (see later), but must necessarily be set up sequentially; this is another reason for the intrinsic weakness of the standard procedure outlined above and generally adopted.

An alternative procedure will be presented along with an essential, brief mathematical-statistical framework needed for a better understanding of the treatment. It will be assumed that the absorbance errors are independent, drawn from a normally-distributed population with mean $\mu = 0$ and constant variance σ^2.

Results and Discussion

Mathematical-statistical Framework.—Generally, the values of the parameters in a functional relationship are obtained by a least-squares calculation based on the Gauss–Newton method,[23b,26] *i.e.* by minimizing expression (1), where y_i and yc_i

$$\text{CHISQ} = \sum_{i=1}^{n} (y_i - yc_i)^2 w_i \qquad (1)$$

are the experimental and calculated absorbances, for unit path length, w_i is the weight given to the ith observation and n is the

number of the experimental points. A fuller expression for yc_i would be $y(a_i, b_i, P_1, P_2, \cdots, P_m)$ where a_i and b_i are the initial concentrations of the two reagents A (ligand, donor) and B (acid, metal), m is the number of parameters of the assumed model and $P_1, P_2, \cdots, P_m$ are the unknown parameters being sought (*i.e.* step formation constants and extinction coefficients).

Eqn. (1) is not linear in the parameters and therefore the ordinary least-squares method cannot be applied. This problem is overcome by approximating $y(a_i, b_i, P_1, P_2, \cdots, P_m)$ with a Taylor series truncated to the first term. The equation to be minimized becomes eqn. (2), where y_i^0 represents $y(a_i, b_i, P_1^0,$

$$\text{CHISQ} = \sum_{i=1}^{n} [y_i - y_i^0 - \sum_{j=1}^{m} (\delta y_i^0/\delta P_j \, \Delta P_j)]^2 \, w_i \quad (2)$$

$P_2^0, \cdots, P_m^0$), the superscript zero refers to the starting values of the parameters (initially guessed values) and ΔP_j is the correction to be given to P_j^0.

In any experiment of n observations a matrix (3), called a

$$\mathbf{X} = \begin{matrix} x_{11} & x_{12} \cdots & x_{1m} \\ x_{21} & x_{22} \cdots & x_{2m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} \cdots & x_{nm} \end{matrix} \quad (3)$$

design matrix, can be constructed, where $(x_{ij}) = (\delta y_i^0/\delta P_j w_i^{\frac{1}{2}})$. The matrix $(\mathrm{X'X})_{mxm} = (c_{jk})_{mxm}$ is known as an information matrix; its inverse is $(\mathrm{X'X})^{-1}{}_{mxm} = (c_{jk}^{-1})_{mxm}$.

The sample variance S_j^2 of a parameter P_j in a functional relationship is estimated by eqn. (4), where $S^2 = \text{CHIMIN}/$

$$S_j^2 = S^2 \, c_{jj}^{-1} \quad (4)$$

$(n - m)$ and CHIMIN is the minimum value of CHISQ. The lower S_j, the higher the precision and reliability of the estimated parameter P_j.

If the variables in the model are highly correlated, the determinant $|\mathrm{X'X}|$ becomes close to zero and serious round off errors may occur in the inverse $(\mathrm{X'X})^{-1}$ and hence in the calculation of the parameters and related statistics. The action generally taken to minimize this problem is a transformation of the variables and therefore a reparametrization of the model.[27a] Each variable is centred and scaled according to the standardizations scheme (5).

$$t_i = (y_i - ym)/u_y, \, z_{ij} = (x_{ij} - xm_j)/u_j \quad (5)$$

where $ym = \sum_{i=1}^{n} (y_i)/n$, $xm_j = \sum_{i=1}^{n} (x_{ij})/n$, $u_j = [\sum_{i=1}^{n} (x_{ij} - xm_j)^2]^{\frac{1}{2}}$ and $u_y = [\sum_{i=1}^{n} (y_i - ym)^2]^{\frac{1}{2}}$.

The matrix $(\mathrm{Z'Z})_{mxm} = (r_{jk})_{mxm}$ is called the correlation matrix; it is symmetric and the main diagonal consists of 1. The term r_{jk} is a partial correlation coefficient, *i.e.* a measure of the correlation between the independent variables x_j and x_k (or z_j and z_k); if all the variables were orthogonal to each other the resultant correlation matrix would be the identity matrix I_m. The relation between an original parameter P_j and the new one P_j' is $P_j = (u_y/u_j)P_j'$ and the new parameter variances are given in eqn. (6),

$$(S_j')^2 = (S')^2 r_{jj}^{-1} \quad (6)$$

where $(S')^2$ is the minimum value of CHISQ for the standardized model.[27a]

The smaller the variance S_j^2 of a parameter P_j, the greater the precision of the parameter estimate. To keep S_j as low as possible we can act on two completely independent factors, S^2 and c_{jj}^{-1}. S^2 is a measure of the total final experimental error and, in general, it can be little affected in a given practical situation, considering that both scientific instrumentation and the relative operators are fixed. Though the experimental error can be little affected, the precision of S^2 can be improved by increasing the number of observations (*i.e.* the degrees of freedom of the estimate).

The term c_{jj}^{-1} depends only on the experimental design, *i.e.* on the initial concentrations of the reagents A and B for each observation. An ideal way of keeping c_{jj}^{-1} low would be to choose points having absolute values $|x_{ik}|_{k=j} \gg 0$ and $|x_{ik}|_{k \neq j} \sim 0$ (*i.e.* an orthogonal design) and so on for the other P_js. Unfortunately, in practice, these designs are unattainable. In fact, in a non-linear functional relationship the role of the independent variables is covered by the derivatives of the function (here absorbance) with regard to the parameters $(x_{ij} = \delta y_i^0/\delta P_j \, w_i^{\frac{1}{2}})$ and these derivatives are highly correlated and cannot vary independently of one another.

An extremely important statistical indicator of the quality of the data is the multiple correlation coefficient R_j given by eqn. (7).

$$R_j = [1 - 1/(c_{jj}c_{jj}^{-1})]^{\frac{1}{2}} = (1 - 1/r_{jj}^{-1})^{\frac{1}{2}} \quad (7)$$

If each $R_j \sim 0$ then the data will have been collected according to an orthogonal design and all the parameters P_j will be independent of one another. On the other hand, if one $R_j \sim 1$ then the data are highly correlated (or multicolinear) and any error in the given P_j can be counterbalanced by suitable variations in the other parameters, thus giving just the same fitting for the minimum, CHIMIN; P_j is then largely indeterminate.

The relevance and peculiar meaning of R_j for a parameter P_j (or P_j') is apparent in eqn. (8), obtained by combining relations (6) and (7).

$$(S_j')^2 = (S')^2 r_{jj}^{-1} = (S')^2/(1 - R_j^2) \quad (8)$$

The term $r_{jj}^{-1} = 1/(1 - R_j^2)$ is called the variance inflation factor (VIF). In an ideal orthogonal design $R_j = 0$ and $\text{VIF}(P_j) = r_{jj}^{-1} = 1$ so that $(S_j')^2 = (S')^2$; the variance of the parameter P_j' depends only on the precision of the experimental data, considering that $(S')^2$ is the absorbance variance in the transformed model. If the variables are highly correlated, and this is often the case in the determination of formation constants, R_j may be close to 1 and $\text{VIF} \gg 1$ so severely inflating the parameter variance.[27b,28]

Known Model.—If the model representing the data is known, then a simple procedure can be followed in order to obtain an optimal determination of the parameters. Though in practice this case is fairly improbable, it may nevertheless be useful to consider it separately, both because of its theoretical importance and for reasons of clarity and simplicity.[29]

We can say that an experiment is optimally designed if it allows the most precise statistical inference within the constraints of the physical and instrumental apparatus employed (*e.g.* solubilities, activity coefficients constancy, available cells, *etc.*). To compare designs different criteria have been proposed; probably the most largely used is the one[29,31] that maximizes the determinant of the information matrix $|\mathrm{X'X}|$ (or of the correlation matrix $|\mathrm{Z'Z}|$). Such a criterion has a tendency to automatically reduce the correlation among the variables and therefore R_j and the variance inflation factor (VIF).

In models that are non-linear in the parameters, the experiments cannot be optimally designed *a priori* because the determinant depends on the design matrix $(x_{ij})_{nxm} = (\delta y_i^0/\delta P_j \, w_i^{\frac{1}{2}})_{nxm}$ which is based on the actual values of the unknown

parameters. This is an unusual situation in which, to design the experiment optimally, we would need to know the same parameters we are searching for. An efficient way of doing so is to adopt a sequential strategy. The parameter estimates available at a particular stage are required to choose the next observation. New improved parameter estimates are calculated using the preceding point and they are then used again to choose the subsequent observation, and so on.

The scheme that will be shown below has been treated in more detail and includes a worked example.[25] To implement this procedure a grid in the (log a, log b) plane should be made using the parameter estimates of the first least-squares calculation based on the initial design (see later) and then possibly updated as the parameter estimates improve. The points on this grid, having a measurable absorbance with our apparatus, constitute the experimental region and are called 'candidates'. The procedure can be detailed as follows.

(1) Collect some points on the basis of some sensible criterion (starting design).

(2) Use the points so far collected, in a non-linear least-squares routine and estimate the parameters and their variances. If the variances are low enough, stop the experiment.

(3) Find the point of the experimentally attainable region that, when added to the points already collected, maximizes the determinant |X′X|. In other words, each candidate must in turn be added to the current design matrix $(x_{ij})_{nxm}$ and the one resulting in a design matrix $(x_{ij})_{(n+1)xm}$ for which the determinant |X′X| is a maximum will be chosen.

(4) Measure the absorbance of the point chosen in (3) and go to (2).

General Procedure: Model Unknown.—In practice both the model and its parameters must be determined from the experiment; the approach previously proposed can only be applied once a single model has been picked out from all the plausible ones initially chosen. Therefore the first and central part of the experiment has to be devoted principally to the discrimination among competing models.

Suppose there are two rival models (Y1 and Y2) to represent the data, one of which is correct. Let Y1c(a, b) and Y2c(a, b) be the calculated values at the point (a, b) for the two models, whose respective parameter estimates have been obtained with least-squares calculations based on the n observations already collected. Then the next observation $n + 1$ will be taken at the point (a, b) where the two models diverge most, *i.e.* where $[\text{Y1c}(a, b) - \text{Y2c}(a, b)]^2$ = maximum, referring to the whole region of operability.[30]

Once the new observation has been measured the least-squares parameter estimates are updated for both models and, as before, the subsequent observation will be found. This sequential procedure will continue until the variance for a model (the incorrect one) is high enough to put it aside. This will occur if, in the region experimentally explored, the two models differ significantly by more than the experimental error. In fact, while the variance of the correct model depends only on the experimental error, that of the incorrect one is due both to this error and to the bias for lack of fit. Furthermore, information will be obtained as to the unexplored region and where to direct our attention, if necessary.

A possible line of action for the general case is suggested.

(1) Collect some points on the basis of some sensible criterion. Ideally this starting design should be both model robust (*i.e.* able to give good parameter estimation for any model) and model sensitive (*i.e.* suitable for highlighting possible inadequacies in the models initially assumed).

(2) As in the standard procedure, estimate the number and possibly the nature of the species in solution, by using either numerical and/or graphical methods.[23a]

(3) Use the information derived from point (2) and any previous knowledge of the system or chemical intuition to set up all the plausible models.

(4) Use a non-linear least-squares routine (here Gauss–Newton) and all the available experimental points in order to calculate, for each of the q models still eligible, the parameters, their variances and the general variance $V(g)$ = CHIMIN-$(g)/[n - m(g)]$, where $g = 1, 2, \cdots, q$ is a model index. If only one plausible model is left from the ones initially entertained, go to (9).

(5) Select the two models giving the best agreement with the experimental data (for instance, the lowest variances), say the models Y1 and Y2.

(6) Calculate the F statistics $F^0 = V(1)/V(2)$ where $V(1)$ and $V(2)$ are the variances [with $V(1) \geqslant V(2)$] of the currently best-fitting models Y1 and Y2, based on all the available experimental points and estimated, respectively, with $N1$ and $N2$ degrees of freedom.[24] Let $F_c = F(1 - \alpha, N1, N2)$ be the critical value of the F distribution for the given degrees of freedom and at a significance level α (often 0.05 or 0.01 *i.e.* 5% or 1%); this means that the probability of $F^0 = V(1)/V(2)$ exceeding $F_c = F(1 - \alpha, N1, N2)$ by chance is α. If $F^0 \geqslant F_c$ remove the model Y2 and go to (4); if $F^0 < F_c$ go to (7).

(7) Working as described earlier and considering all candidates, find the point for which the two models are furthest apart, *i.e.* $[\text{Y1c}(a, b) - \text{Y2c}(a, b)]^2$ = maximum. Add another point in the same way and measure them both.

(8) Find a point each, to maximize |X′X| both for models Y1 and Y2. Measure them and go to (4).

(9) At this stage hopefully only one model is left, because all the others have been eliminated in the process of discrimination. Working according to the principles of the maximum determinant (see last section 'Known Model') add the observations sequentially until the parameter variances are sufficiently low, then stop the experiment.

Worked Example.—The scheme examined above will be illustrated through a simple simulated worked example; the aim is to clarify the proposed handling of the problem. Two species A and B react in a non-absorbing solvent according to eqns. (i) and (ii) to give the only absorbing species AB and A_2B whose extinction coefficients are ε_{21} and ε_{22}, respectively.

$$A + B \underset{}{\overset{K_{21}}{\rightleftharpoons}} AB \qquad (i)$$

$$AB + A \underset{}{\overset{K_{22}}{\rightleftharpoons}} A_2B \qquad (ii)$$

The activity coefficients are assumed constant throughout the experiment and Beer's law is valid.

The values of K_{21} and K_{22} used to generate the data are 10.00 and 0.100 $dm^3\ mol^{-1}$; those of ε_{21} and ε_{22} are 10 000.0 and 21 000.0 $dm^3\ mol^{-1}\ cm^{-1}$, respectively. Errors, normally distributed with mean $\mu = 0$ and $\sigma^2 = 0.0010^2$, have been added to the calculated absorbances through a FORTRAN program based on random number generation. The observations have been chosen in a (log a, log b) plane from a 21 × 21 grid ($-4.0 \leqslant \log a, \log b \leqslant -1.0$; $\Delta\log a = \Delta\log b = 0.15$); the points in the grid having an absorbance between 0.1 and 4.0 units (cells of unit path length) have been supposed experimentally attainable (*i.e.* candidates).

The starting design consists of nine points, deliberately collected to cover the experimental region principally at the borderline (see later). The initial plausible models have been assumed to be those of 2:1 and 1:1 stoichiometry that in the present case can hardly be distinguished, considering that $K_{22} \ll K_{21}$. Tests have been made with different initial guess estimates for the parameters and generally there has been a convergence about the same final values. The calculations have

Table 1 The step-by-step output of the program DISCRI for the worked example[a]

n	log a	log b	Absorbance	k_{21}	k_{22}	ε_{21}	ε_{22}	k_1	ε_1	Det2	Det1	CHIMIN2	CHIMIN1
1 (0)	−4.000	−1.000	0.500	100.0	11.0	200×10^2	520×10^2	100.0	200×10^2	—	—	316	155
2 (0)	−3.550	−1.000	1.408										
3 (0)	−3.100	−1.000	3.965										
4 (0)	−3.400	−2.500	0.122										
5 (0)	−2.650	−2.500	0.672										
6 (0)	−1.900	−2.500	3.458										
7 (0)	−1.900	−4.000	0.113										
8 (0)	−1.450	−4.000	0.264										
9 (0)	−1.000	−4.000	0.507	10.0	0.0798	100×10^2	242×10^2	10.0	100×10^2	—	0.32×10^{-8}	0.573×10^{-6}	0.920×10^{-5}
10 (3)	−1.000	−3.250	2.853	10.0	0.0935	100×10^2	221×10^2	9.92	101×10^2	0.21×10^{-24}	0.45×10^{-8}	0.594×10^{-6}	0.178×10^{-3}
11 (3)	−1.150	−3.100	3.321	10.0	0.0896	100×10^2	227×10^2	9.91	101×10^2	0.62×10^{-23}	0.59×10^{-8}	0.754×10^{-6}	0.203×10^{-3}
12 (1)	−2.200	−2.200	3.552	10.0	0.0874	100×10^2	230×10^2	9.91	101×10^2	—	—	—	—
13 (2)	−1.300	−2.950	3.760	10.0	0.0829	100×10^2	241×10^2	9.88	101×10^2	0.43×10^{-22}	0.15×10^{-7}	0.635×10^{-6}	0.182×10^{-3}
14 (3)	−3.100	−1.150	3.283	10.0	0.0694	100×10^2	281×10^2	9.91	101×10^2	0.45×10^{-22}	0.18×10^{-7}	0.791×10^{-6}	0.203×10^{-3}
15 (3)	−1.000	−3.400	2.021	10.0	0.0648	100×10^2	299×10^2	9.88	101×10^2	0.56×10^{-22}	0.19×10^{-7}	0.722×10^{-6}	0.214×10^{-3}
16 (1)	−2.350	−2.050	3.525	10.0	0.0647	100×10^2	300×10^2	9.88	101×10^2	—	—	—	—
17 (2)	−1.450	−2.950	2.941	10.0	0.0712	100×10^2	275×10^2	9.87	101×10^2	0.12×10^{-21}	0.31×10^{-7}	0.900×10^{-6}	0.192×10^{-3}
18 (3)	−3.250	−1.000	2.807	10.0	0.0854	100×10^2	239×10^2	9.90	101×10^2	0.25×10^{-21}	0.36×10^{-7}	0.788×10^{-6}	0.204×10^{-3}
19 (3)	−2.950	−1.300	3.728	10.0	0.0871	100×10^2	235×10^2	9.90	101×10^2	0.34×10^{-21}	0.40×10^{-7}	0.738×10^{-6}	0.207×10^{-3}
20 (2)	−2.500	−1.900	3.451	10.0	0.0950	100×10^2	220×10^2	9.89	101×10^2	0.53×10^{-21}	0.50×10^{-7}	0.759×10^{-6}	0.198×10^{-3}
21 (2)	−1.600	−2.800	3.166	10.0	0.101	100×10^2	210×10^2	9.90	101×10^2	0.73×10^{-21}	0.55×10^{-7}	0.736×10^{-6}	0.189×10^{-3}
22 (2)	−2.050	−2.350	3.528	10.0	0.106	100×10^2	202×10^2	9.90	101×10^2	0.10×10^{-20}	0.67×10^{-7}	0.731×10^{-6}	0.180×10^{-3}
23 (2)	−1.300	−3.100	2.665	10.0	0.0969	100×10^2	216×10^2	9.89	101×10^2	0.11×10^{-20}	0.70×10^{-7}	0.798×10^{-6}	0.177×10^{-3}
24 (2)	−2.650	−1.750	3.326	10.0	0.0999	100×10^2	211×10^2	9.89	101×10^2	0.13×10^{-20}	0.78×10^{-7}	0.844×10^{-6}	0.172×10^{-3}
25 (2)	−1.750	−2.650	3.339	10.0	0.107	100×10^2	201×10^2	9.89	101×10^2	0.17×10^{-20}	0.86×10^{-7}	0.885×10^{-6}	0.165×10^{-3}
26 (2)	−1.000	−3.550	1.430	10.0	0.108	100×10^2	198×10^2	9.88	101×10^2	0.20×10^{-20}	0.88×10^{-7}	0.845×10^{-6}	0.166×10^{-3}
27 (2)	−2.800	−1.600	3.150	10.0	0.110	100×10^2	196×10^2	9.88	101×10^2	0.23×10^{-20}	0.94×10^{-7}	0.811×10^{-6}	0.162×10^{-3}
28 (2)	−1.150	−3.250	2.354	10.0	0.110	100×10^2	196×10^2	9.87	101×10^2	0.26×10^{-20}	0.98×10^{-7}	0.780×10^{-6}	0.164×10^{-3}
29 (2)	−3.400	−1.000	1.989	10.0	0.111	100×10^2	195×10^2	9.88	101×10^2	0.29×10^{-20}	0.10×10^{-6}	0.782×10^{-6}	0.164×10^{-3}
30 (2)	−2.500	−2.050	2.523	10.0	0.110	100×10^2	195×10^2	9.88	101×10^2	0.32×10^{-20}	0.11×10^{-6}	0.763×10^{-6}	0.158×10^{-3}
31 (2)	−2.350	−2.200	2.552	10.0	0.111	100×10^2	194×10^2	9.87	101×10^2	0.34×10^{-20}	0.12×10^{-6}	0.760×10^{-6}	0.153×10^{-3}
32 (2)	−1.450	−3.100	2.085	10.0	0.111	100×10^2	194×10^2	9.87	101×10^2	0.38×10^{-20}	0.12×10^{-6}	0.733×10^{-6}	0.149×10^{-3}
33 (2)	−3.250	−1.150	2.327	10.0	0.111	100×10^2	194×10^2	9.88	101×10^2	0.42×10^{-20}	0.13×10^{-6}	0.730×10^{-6}	0.149×10^{-3}
34 (2)	−2.200	−2.350	2.552	10.0	0.112	100×10^2	193×10^2	9.88	101×10^2	0.44×10^{-20}	0.14×10^{-6}	0.734×10^{-6}	0.144×10^{-3}
35 (2)	−1.600	−2.950	2.248	10.0	0.115	100×10^2	189×10^2	9.88	101×10^2	0.48×10^{-20}	0.14×10^{-6}	0.741×10^{-6}	0.141×10^{-3}
36 (2)	−2.650	−1.900	2.460	10.0	0.117	100×10^2	186×10^2	9.88	101×10^2	0.56×10^{-20}	0.15×10^{-6}	0.718×10^{-6}	0.137×10^{-3}
37 (2)	−1.300	−3.250	1.889	10.0	0.115	100×10^2	188×10^2	9.88	101×10^2	0.60×10^{-20}	0.15×10^{-6}	0.714×10^{-6}	0.135×10^{-3}
100 (2)	−2.050	−2.950	0.911	10.0	0.116	100×10^2	188×10^2	9.88	101×10^2	0.21×10^{-19}	0.31×10^{-6}	0.978×10^{-6}	0.625×10^{-4}
207 (2)	−2.050	−3.850	0.115	10.0	0.124	100×10^2	179×10^2	9.88	101×10^2	0.26×10^{-19}	0.34×10^{-6}	0.894×10^{-6}	0.307×10^{-4}

[a] For each step of the simulated experiment of the worked example and for both assumed models [(Y1 = $y(a, b, K_1, \varepsilon_1)$ and Y2 = $y(a, b, K_{21}, K_{22}, \varepsilon_{21}, \varepsilon_{22})$] some pertinent quantities are shown. Beside n, in round brackets, is a code, 0 means that the observation belongs to the starting design, 1 that the observation has been chosen because it maximizes the determinant $|(X'X)_{2\times2}|$ for model Y1, 2 because it maximizes $|(X'X)_{4\times4}|$ for model Y2, and 3 because it gives the best discrimination between Y1 and Y2. The first points added appear to be either very influential on the values of the determinants or very effective in discriminating between Y1 and Y2. K_{22} and ε_{22} are strongly correlated [$r^2(K_{22}, \varepsilon_{22}) \sim 1$] and consequently their values show a remarkable drift in the course of the experiment. The experiment is terminated ($n = 37$) when det2 appears to increase moderately (see text).

Table 2 Some other statistically significant results from the worked example[a]

n	c_{11}^{-1}/K_{21}^2	c_{22}^{-1}/K_{22}^2	$c_{33}^{-1}/\varepsilon_{21}^2$	$c_{44}^{-1}/\varepsilon_{22}^2$	VIF(K_{21})	VIF(K_{22})	VIF(ε_{21})	VIF(ε_{22})	VIF(K_1)	c_{11}^{-1}/K_1^2	$c_{22}^{-1}/\varepsilon_1^2$
1 (0)											
2 (0)											
3 (0)											
4 (0)											
5 (0)											
6 (0)											
7 (0)											
8 (0)											
9 (0)									13.6	0.970	0.442
10 (3)	13.2	0.104×10^8	3.66	8.19×10^7	211.8	25 270	140.9	25 040	13.8	0.857	0.356
11 (3)	1.7	0.692×10^6	0.614	0.399×10^6	33.7	2 338	30.2	2 307	16.9	0.848	0.340
12 (1)	—	—	—	—	—	—	—	—	—	—	—
13 (2)	0.754	0.274×10^6	0.374	0.167×10^6	27.2	1 192	28.3	1 170	18.5	0.508	0.241
14 (3)	0.724	0.381×10^6	0.330	0.226×10^6	28.8	1 613	28.6	1 593	19.7	0.491	0.225
15 (3)	0.715	0.383×10^6	0.327	0.241×10^6	29.2	2 030	29.6	2 010	19.3	0.470	0.211
16 (1)	—	—	—	—	—	—	—	—	—	—	—
17 (2)	0.533	0.224×10^6	0.278	0.231×10^6	29.4	1 630	31.0	1 612	20.2	0.363	0.179
18 (3)	0.470	0.156×10^6	0.228	0.119×10^6	26.9	1 025	27.2	1 010	19.1	0.332	0.159
19 (3)	0.470	0.145×10^6	0.224	0.102×10^6	29.8	916	29.9	906	21.2	0.332	0.158
20 (2)	0.400	0.114×10^6	0.205	0.804×10^5	28.9	734	29.7	727	21.2	0.291	0.144
21 (2)	0.400	0.895×10^5	0.205	0.649×10^5	31.4	613	31.7	604	22.4	0.282	0.143
22 (2)	0.364	0.807×10^5	0.193	0.507×10^5	32.2	524	32.3	515	22.4	0.252	0.133
23 (2)	0.357	0.865×10^5	0.191	0.481×10^5	32.7	554	33.3	545	23.1	0.251	0.131
24 (2)	0.338	0.776×10^5	0.187	0.516×10^5	33.6	550	34.6	542	24.0	0.240	0.128
25 (2)	0.335	0.657×10^5	0.186	0.430×10^5	35.9	478	36.4	469	25.0	0.231	0.126
26 (2)	0.334	0.608×10^5	0.185	0.376×10^5	35.9	464	36.7	456	24.5	0.227	0.122
27 (2)	0.329	0.575×10^5	0.185	0.355×10^5	37.5	440	38.5	433	25.7	0.224	0.122
28 (2)	0.327	0.561×10^5	0.184	0.341×10^5	37.9	470	39.4	463	25.5	0.219	0.118
29 (2)	0.307	0.557×10^5	0.170	0.338×10^5	35.9	466	36.9	459	24.8	0.210	0.113
30 (2)	0.293	0.551×10^5	0.165	0.320×10^5	35.7	449	36.9	442	24.8	0.202	0.110
31 (2)	0.280	0.539×10^5	0.161	0.333×10^5	35.6	454	37.0	447	24.8	0.193	0.107
32 (2)	0.279	0.504×10^5	0.160	0.300×10^5	36.1	426	37.6	419	25.2	0.193	0.107
33 (2)	0.270	0.508×10^5	0.152	0.301×10^5	35.5	426	36.6	420	25.1	0.189	0.104
34 (2)	0.260	0.499×10^5	0.149	0.305×10^5	35.6	426	36.7	419	25.1	0.182	0.101
35 (2)	0.260	0.453×10^5	0.149	0.295×10^5	36.4	407	37.5	400	25.6	0.181	0.101
36 (2)	0.253	0.428×10^5	0.147	0.258×10^5	36.6	372	37.9	366	25.8	0.177	0.0993
37 (2)	0.252	0.426×10^5	0.147	0.239×10^5	36.9	366	38.3	360	26.0	0.177	0.0988
100 (2)	0.186	0.317×10^5	0.114	0.190×10^5	44.0	345	45.3	340	30.4	0.128	0.0757
207 (2)	0.178	0.271×10^5	0.110	0.158×10^5	44.3	297	45.6	293	30.7	0.123	0.0734

[a] Some other relevant results for both models (Y1 and Y2) of the worked example are shown. c_{jj}^{-1}/P_j^2 is the component of the relative variance ($S^2 c_{jj}^{-1}/P_j^2$) of P_j, that can be ascribed to the design. The first points added have a dramatic impact on this quantity; the effect is decidedly more pronounced and persistent for c_{22}^{-1}/K_{22}^2 and $c_{44}^{-1}/\varepsilon_{22}^2$, as K_{22} and ε_{22} are strongly correlated. A similar behaviour is noticeable in the VIFs (Variance Inflation Factor) of the various parameters. The closer R_j^2 (square multiple correlation coefficient) to 1 the higher the value of VIF(P_j). For an ideal orthogonal design $R_j^2 = 0$ and consequently VIF(P_j) = 1.

been made with a FORTRAN program DISCRI based on the Gauss–Newton method as modified by Marquardt[26] and supplemented with subroutines to perform the required tasks. The first index or number of entries in the headings of both Tables 1 and 2 refer to the model: 1 to Y1 = $y(a, b, K_1, \varepsilon_1)$ and 2 to Y2 = $y(a, b, K_{21}, K_{22}, \varepsilon_{21}, \varepsilon_{22})$ (except c_{jj}^{-1}).

In Table 1 some relevant results of the program DISCRI are shown as each experimental point is added to the design. At the side of n, in parentheses, is a code, 0 means that the point in question belongs to the starting design, 1 that the point has been chosen because it maximizes the determinant $|(X'X)_{2\times 2}|$ for model Y1 at the current stage, 2 because it maximizes $|(X'X)_{4\times 4}|$ for model Y2, and 3 because it gives the best discrimination between Y1 and Y2. At the significance level of $\alpha = 0.0010$ and $n = 9$, $F^0 = 16.1$ and $F_c(7, 5) = 28.2$ so that model Y1 cannot be rejected. For $n = 10$, $F^0 = 300$ and $F_c(8, 6) = 19.0$ therefore Y1 could be discarded. However, the process of discrimination is not terminated until $n = 19$ [for which $F^0 = 280$ and $F_c(17, 15) \sim 5.4$], both to avoid relying too much on so few points and also to give a better demonstration of how the method functions.

As can be seen, whatever the basis for collecting the experimental points, they demonstrate a tendency to concentrate on the borderline of the region being explored. K_{22} and ε_{22} display a remarkable drift and this can be ascribed to their very strong correlation [$r^2(K_{22}, \varepsilon_{22}) \sim 1.0$]; though $r^2(K_{22}, \varepsilon_{22})$ decreases during the experiment it remains very high, and consequently the estimates of K_{22} and ε_{22} are strongly responsive to the error fluctuations of the absorbance. The experiment is terminated when the determinant value of Y2 (det2) appears to increase moderately ($n = 37$). In fact, since the determinant embodies the variance of all the parameters, its monitoring has been preferred to that of a single-parameter variance.

For each entry, in Tables 1 and 2, are shown the values for the experiment as if it had been continued till $n = 100$ and 207 (all candidates). It is manifest that the information gathered after $n = 37$ is not highly significant.

In Table 2 the component of the relative variance ($S^2 c_{jj}^{-1}/P_j^2$) of P_j, that can be ascribed to the design (c_{jj}^{-1}/P_j^2), is shown; the remaining part (S^2) depends only on the experimental error which can vary unpredictably from point to point. The impact of the first four points (10, 11, 12, 13) for each c_{jj}^{-1}/P_j^2 of Y2 is dramatic as can globally be seen from the increase of det2 in Table 1. The effect is still strong until $n = 22$ when it diminishes and becomes insignificant after $n = 37$. The effect is considerably more pronounced for c_{22}^{-1}/K_{22}^2 and $c_{44}^{-1}/\varepsilon_{22}^2$, as K_{22} and ε_{22} are the most correlated and hence the most indeterminate parameters. The VIFs [VIF(P_j) = $1/(1 - R_j^2)$] for K_{21} and ε_{21} present an initial noticeable

decrease and then a slow gradual increase. The VIFs for K_{22} and ε_{22}, however, show a constant decrease that is much more marked for the first points. In fact, the selected points are very effective in gradually removing the strongest correlations (here between K_{22} and ε_{22}) that result in very small values of the determinant (if $R_j = 1$ then det = 0), while their effect on the less correlated parameters is smaller.

The standard deviation for a parameter and a given n can be easily computed from the Tables. For $n = 37$, $S(K_{22}) = [V(K_{22})]^{\frac{1}{2}} = [(c_{22}^{-1}/K_{22}^2)\ K_{22}^2\ \mathrm{CHIMIN2}]^{\frac{1}{2}} = (0.426 \times 10^5 \times 0.115^2 \times 0.714 \times 10^{-6})^{\frac{1}{2}} = 0.020$; $S(\varepsilon_{22}) = (0.239 \times 10^5 \times 18\,800^2 \times 0.714 \times 10^{-6})^{\frac{1}{2}} = 2456$; $S(K_{21}) = 0.0042$ and $S(\varepsilon_{21}) = 3.2$. It is evident that the relative standard deviations of K_{22} and ε_{22} (0.17 and 0.13, respectively) are very high, revealing that their correlation has been removed only to a small extent; $\mathrm{VIF}(K_{22}) = 366$ and this is due almost exclusively to $r^2(K_{22}, \varepsilon_{22})$.

Calculations have been performed using data with different sample variances S^2. The results may vary significantly, particularly for the ones associated with the correlated parameters. They become significantly better if $S^2 < 0.0010$, worse if $S^2 > 0.0010$. So for data with high correlations the experimental error becomes very important.

Conclusions

A set of experimental points can be good to discriminate amongst some models but not amongst others and, at the same time, they can either be good or useless for the determination of a certain model. The quality of a design depends only on the particular model considered, which is initially unknown. A procedure that collects all the data together at the start is therefore not recommended.

The method described allows, in a sequential manner, both the discrimination between whatever concurrent models are used, and the optimal determination of the final model. One way to proceed is *via* the one illustrated, where both the type and the number of points of the starting design could be a problem. In this case, further research is probably needed in order to obtain maximum efficiency. Another open question is the number of points to allocate both to the discrimination and to the optimal parameter determination before the final model is chosen. During the early steps, perhaps, more emphasis should be given to the discrimination, though the points here have been equally distributed between the two different tasks.

The principles expounded can also be applied to repair designs of the standard type, *i.e.* when the data are collected together before they are processed. In such a case experimental points might be added to obtain either a better discrimination among certain particular models or a better determination of the final chosen model[31] or even both. Directions for a possible enlargement of the experimental region may also be obtained. Perhaps this is the route to follow to become accustomed to the method and, at the same time, possibly saving on the experimental time.

In general, though less so today, in experimental research there is a propensity to ignore the design and concentrate on the statistical analysis of the data. No clever analysis can extract much information from data where there is none. The design is much more important than the statistical analysis, though in practice the reverse is true. The suggested approach should give us all the information that can be drawn out of the system.

An essential practice is a close-mesh grid in the (log a, log b) plane, based on the most plausible model obtained from the starting design. This grid should be updated as the model or its parameters change. These guidelines can be applied to complexes of whatever stoichiometry and to other types of problems and, if properly used, their limit should be mainly in the system constraints and in the region experimentally attainable.

References

1 L. Newman and D. N. Hume, *J. Am. Chem. Soc.*, 1957, **17**, 4571.
2 R. A. La Budde and M. Tamres, *J. Phys. Chem.*, 1970, **74**, 4009.
3 B. Dodson, R. Foster and A. A. S. Bright, *J. Chem. Soc. B*, 1971, 1283.
4 T. O. Maier and R. S. Drago, *Inorg. Chem.*, 1972, **11**, 1861.
5 G. Carta, G. Crisponi and V. Nurchi, *Tetrahedron*, 1981, **37**, 2115.
6 G. Carta and G. Crisponi, *J. Chem. Soc., Perkin Trans. 2*, 1982, 53.
7 B. K. Seal, H. Sil, M. Banerjee and D. C. Mukherjee, *Bull. Chem. Soc. Jpn.*, 1982, **55**, 1620.
8 S. Natansohn, J. I. Krugler, J. E. Lester, M. S. Chagnon and R. S. Finocchiaro, *J. Phys. Chem.*, 1980, **84**, 2972.
9 H. Benesi and J. H. Hildebrand, *J. Am. Chem. Soc.*, 1949, **71**, 2703.
10 R. L. Scott, *Recl. Trav. Chim. Pay-Bas*, 1956, **75**, 787.
11 N. J. Rose and R. S. Drago, *J. Am. Chem. Soc.*, 1959, **81**, 6138.
12 P. R. Hammond, *J. Chem. Soc.*, 1964, 479.
13 W. B. Person, *J. Am. Chem. Soc.*, 1965, **87**, 167.
14 S. D. Ross and M. M. Labes, *J. Am. Chem. Soc.*, 1957, **76**, 79.
15 M. Tamres, *J. Phys. Chem.*, 1960, **65**, 654.
16 D. A. Derenleau, *J. Am. Chem. Soc.*, 1969, **91**, 4044; 1969, **91**, 4050.
17 K. Conrow, D. G. Johnson and R. E. Bowen, *J. Am. Chem. Soc.*, 1964, **86**, 1025.
18 W. E. Wentworth, W. Hirsch and E. Chen, *J. Phys. Chem.*, 1967, **71**, 218.
19 D. R. Rosseinsky and H. Kellawi, *J. Chem. Soc. A*, 1969, 1207.
20 R. E. Lenkinski, G. A. Elgavish and J. Reuben, *J. Magn. Reson.*, 1978, **32**, 367.
21 R. M. Alcock, F. R. Hartley, D. E. Rogers and J. L. Wagner, *J. Chem. Soc., Dalton Trans.*, 1975, 2189.
22 J. S. Coleman, L. P. Varga and S. H. Mastin, *Inorg. Chem.*, 1970, **9**, 1015.
23 F. R. Hartley, C. Burgess and R. Alcock, *Solution Equilibria*, Ellis Horwood, Chichester, 1980; (*a*) pp. 33–51; (*b*) pp. 87–98.
24 L. Bologni, A. Sabatini and A. Vacca, *Inorg. Chim. Acta*, 1983, **69**, 71.
25 G. Carta and G. Crisponi, *J. Chem. Soc., Perkin Trans. 2*, 1986, 371.
26 F. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1969, pp. 232–242.
27 J. Neter, W. Wasserman and M. H. Kutner, *Applied Linear Regression Models*, R. D. Irwin, Homewood, 2nd edn.; (*a*) pp. 288–295; (*b*) pp. 408–411.
28 R. H. Myers, *Classical and Modern Regression with Applications*, Duxbury Press, Boston, 1986, pp. 77–78.
29 G. E. P. Box and H. L. Lucas, *Biometrica*, 1959, **46**, 77.
30 W. G. Hunter and A. M. Reiner, *Tecnometrics*, 1965, **7**, 307.
31 O. Dykstra, *Tecnometrics*, 1971, **13**, 682.

Paper 2/00751G
Received 12*th February* 1992
Accepted 10*th March* 1992