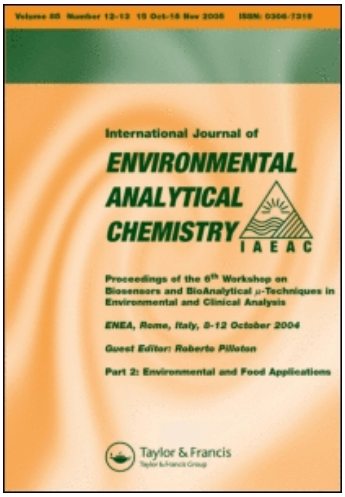
This article was downloaded by:
On: *18 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713640455

Chemometric Methods for the Study of Toxic Metals on the Growth of Plants: Use of Experimental Design and Response Surface Methodology

Mahmoud A. Allus[a]; Richard G. Brereton[a]
[a] School of Chemistry, University of Bristol, Bristol, UK

To cite this Article Allus, Mahmoud A. and Brereton, Richard G.(1990) 'Chemometric Methods for the Study of Toxic Metals on the Growth of Plants: Use of Experimental Design and Response Surface Methodology', International Journal of Environmental Analytical Chemistry, 38: 2, 279 — 304
To link to this Article: DOI: 10.1080/03067319008026933
URL: http://dx.doi.org/10.1080/03067319008026933

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Intern. J. Environ. Anal. Chem., Vol. 38, pp. 279–304
Reprints available directly from the publisher
Photocopying permitted by license only

© 1990 Gordon and Breach, Science Publishers, Inc.
Printed in Great Britain

CHEMOMETRIC METHODS FOR THE STUDY OF TOXIC METALS ON THE GROWTH OF PLANTS: USE OF EXPERIMENTAL DESIGN AND RESPONSE SURFACE METHODOLOGY

MAHMOUD A. ALLUS and RICHARD G. BRERETON*

School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

Downloaded At: 18:34 18 January 2011

Chemometrics is a recent discipline concerned, principally, with the application of mathematics and statistics to laboratory systems. One way in which the chemometrician can aid the environmental analytical chemist is via planned experimental designs. In this paper the importance of experimental design is illustrated and the main considerations prior to experimentation, namely, degrees of freedom, analytical errors, coding and modelling, are outlined. This is exemplified by a study of the influence of potentially toxic heavy metals on the growth of barley seedlings. Undesigned univariate experiments suggest that Tl is probably more toxic than Cd. A three factor central composite design is reported, to study the relative toxicities of Tl, Cd and Pb and also of Tl, Fe and Zn. The paper exemplifies how much information can be obtained from the resultant experimental response data. Multilinear regression can be employed to produce a quadratic model: this can be interpreted graphically by reconstructed univariate response curves and 3-dimensional response surfaces. Analysis of variance is a statistical method for computing how well the model has been fitted, taking into account analytical errors. With the aid of modern graphical computing, a variety of confidence intervals can be displayed for both univariate and bivariate responses. The usefulness of the design can be visualised by displaying leverage over and outside the experimental region. Finally future trends in multivariate response methodology are discussed.

1. INTRODUCTION

Heavy metals are important environmental pollutants,[1,2] often byproducts of industrial processes. Concentrations of these metals can be monitored by studying plant and animal tissues, and samples of water, sediments and soil. The influence of these metals on plant and animal growth can be observed in the field. However, many investigators also employ laboratory experiments. Clearly it is important to relate the results of these experiments to field studies. In order to do this, carefully

*Author to whom correspondence should be directed.

designed experiments are required. It is possible to obtain large quantities of laboratory data which cannot properly be related to field studies. Unless the experimental procedures are carefully planned in advance it is often hard to quantitate how valid the laboratory based results are.

In this paper we demonstrate the importance of experimental design in analytical environmental chemistry, illustrated by the influence of heavy metals on the dry weight of barley seedlings after 4 weeks' growth.

Chemometrics is a well established discipline. The interested reader is referred to established textbooks,[3,4] journals,[5,6] reviews,[7–12] series of monographs,[13] and conference proceedings.[14–16]

Analytical environmental chemistry is particularly suited to the chemometric approach since it is possible to obtain a large number of measurements from both natural and laboratory systems. However, because these processes are often influenced by several factors, many of which are hard to quantify, exact relationships (often possible to predict and establish in pure physical sciences) are not expected: instead multivariate statistical approaches are required to disentangle the various influences on these systems.

Downloaded At: 18:34 18 January 2011

2. IMPORTANCE OF EXPERIMENTAL DESIGN

2.1 Degrees of Freedom

Environmental chemists are frequently interested in how a *response* (e.g. the dry weight of a plant) is influenced by a *factor* (e.g. the concentration of Tl).

Many experiments are involved in observing and computing *response curves.* The interpretation of these curves depends on *modelling* the process: below we discuss the importance of choosing correct models in more detail, but at present we assume a model is a proposed equation that describes experimental data: much of the aim of experimentation is to discover how well such an equation actually does describe the observed trends. How well the process can be modelled depends on how the experimental sampling strategy is designed.

Consider the simple example of determining whether a linear relationship between a response (y) and a factor (x) exists. If only *two* responses are measured then the data will always exactly fit a straight line relationship, with a correlation coefficient of ± 1: thus this particular experiment is *poorly designed* as there is no information as to whether x and y are linearly related (Figure 1a). It is possible to envisage more complex systems in which, in practice, there is no information about whether a response obeys a particular model, but standard statistical indicators such as correlation coefficients, if misinterpreted, can be cited, out of context, as evidence for relationships between variables.

In the above experiment we say that there are no *degrees of freedom* to assess the validity of the model. A better experiment might be to observe three responses (Figure 1b). In such case there is some information as to whether the two parameters are linearly related: there is *one* degree of freedom. Obviously if more experiments are performed, there will, consequently, be increasingly more degrees

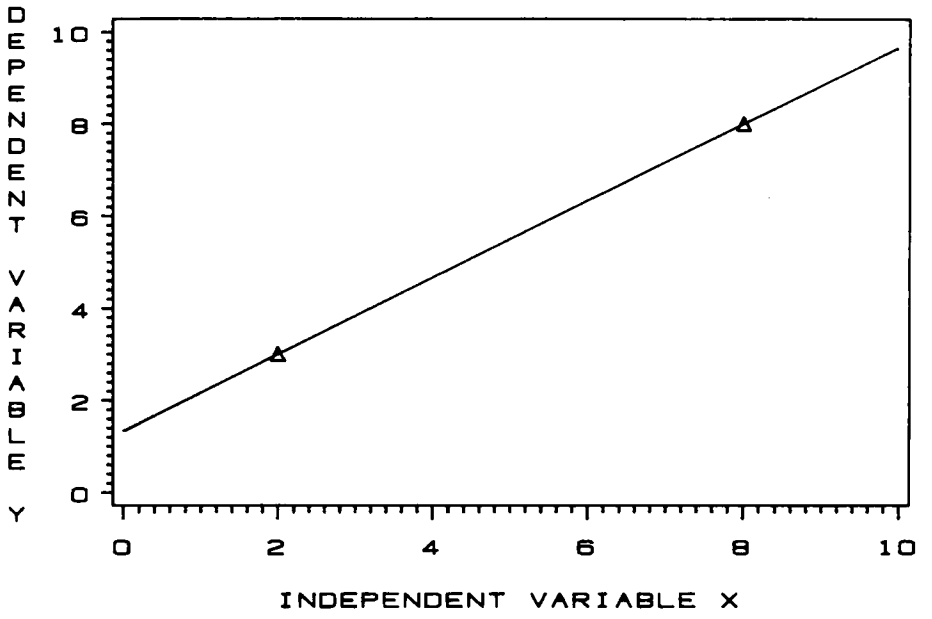


Figure 1a Example of a perfect apparent linear fit to 2 datapoints, with correlation coefficient of ± 1. This experiment tells us nothing about whether the data actually obeys a linear model or not.

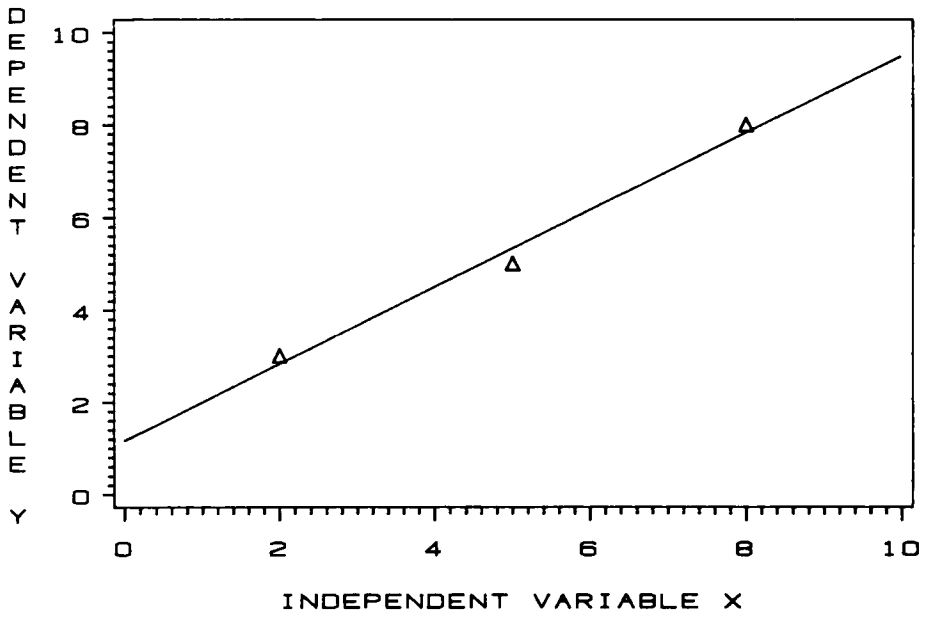


Figure 1b Example of a better designed experiment, consisting of three samples. There is now some information as to whether the data obeys a linear model, although the correlation coefficients must be treated with caution.

Downloaded At: 18:34 18 January 2011

of freedom so the experimenter can be more confident in his model. If, however, the experimenter wants to change the model to a more sophisticated one, then even more observations are required. The experiment in Figure 1b will give a perfect fit to a quadratic model, so there will be no degrees of freedom to assess how well the experimental data obey the model.

In general the number of degrees of freedom to assess the fit to a model is given by

$$v = N - P \tag{1}$$

where v is the degrees of freedom, N is the number of *non-replicated observations* (see below for a discussion of replicates) and P is the number of *parameters in the model.* Thus if, as illustrated in Figure 1b, there are 3 non-replicated sampling points and a linear model of the form

$$y = b_0 + b_1 x \tag{2}$$

(i.e. a 2 parameter model), $v = 3 - 2 = 1$.

Correlation coefficients must always be interpreted taking the number of degrees of freedom into account. If there are very few degrees of freedom, then high apparent correlations often lack significance.

Downloaded At: 18:34 18 January 2011

2.2 Analytical Error and Replicates Analysis

Obviously the more experiments performed, the greater the number of degrees of freedom and so more information is available about the model. When do we stop experimenting?

A great deal depends on the accuracy and reproducibility of the analytical measurement process. Consider a highly reproducible analytical procedure. As illustrated in Figure 1c, it is relatively easy to assess whether y is linearly related to x using only 3 points. If, however, the measurement process is less reproducible then even if y and x are linearly related, because of the analytical error, it is hard to obtain this information from only 3 experiments: this is illustrated in Figure 1d.

Thus a good experiment provides information about the analytical error as well as the model itself. This is normally performed *via* replicates analysis, i.e. repeating certain sampling points. Naturally the more replicates, the greater the degrees of freedom and so the more the information available about the analytical process. Often in the rush to obtain as many measurements as possible, investigators do not have the patience to obtain replicates information, or else average replicated measurements. In environmental applications, where measurement processes are not always highly reproducible, some information about analytical errors is *essential* prior to analysis of quantitative data.

In statistical terms the *variance* of the least squares fit to the model and the least squares estimate of each parameter is compared to that of the experimental error. This approach is called *ANOVA* (analysis of variance). There are a large number of statistical and chemometric texts on experimental design,[17–21] many of which

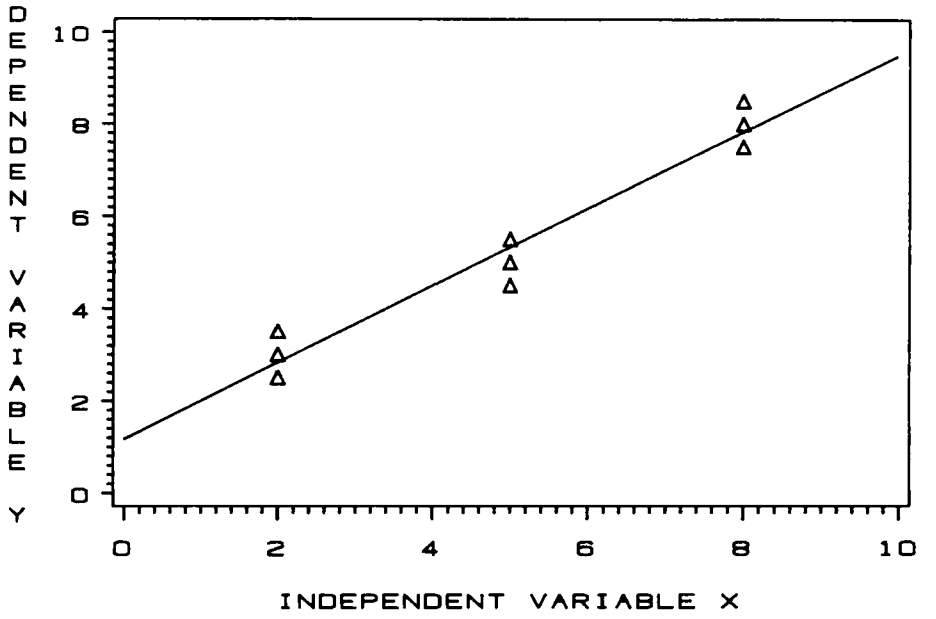


Figure 1c Influence of high analytical reproducibility on model pictured in Figure 1b. Because reproducibility is high, almost any non-replicated sampling scheme will yield the same results.

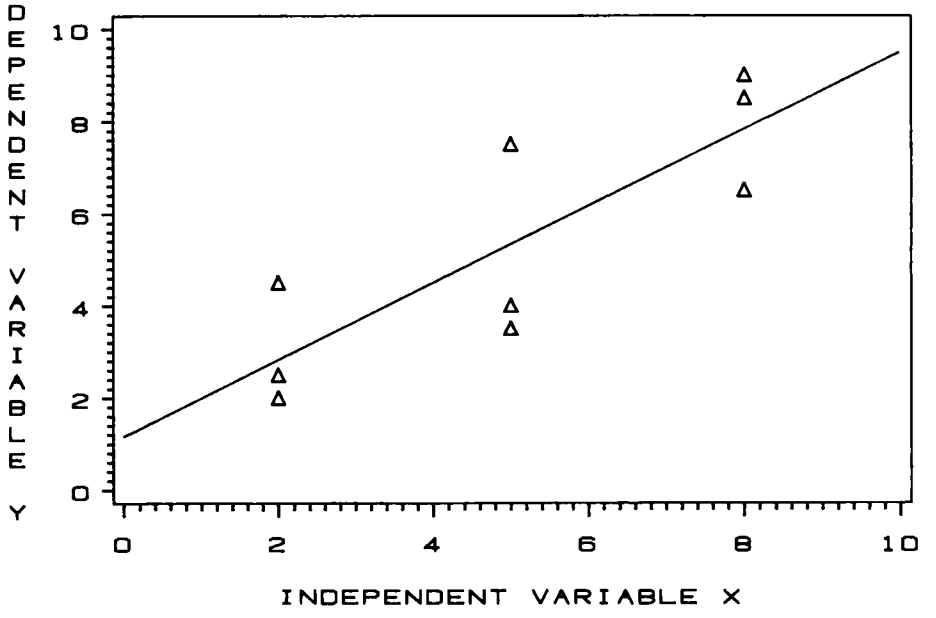


Figure 1d Influence of low analytical reproducibility on model pictured in Figure 1b. In this case, there is high variability in the data, so 3 sampling points may not give a good picture of the model so apparently low correlation coefficients may be the result of the analytical process rather than a poor fit to the actual model.

Downloaded At: 18:34 18 January 2011

are based around designing experiments to provide sufficient information to perform ANOVA on the resultant response.

As we will discuss in greater detail below, it is possible to visualise how well experimental data are described by a model by computing confidence intervals. The broader the confidence bands the less useful the experiment. It is only possible to compute these bands if there are sufficient degrees of freedom, so we cannot compute confidence intervals for the data in Figure 1a as there is no information as to how well the model is fitted. We can, however, compute confidence intervals for the data in Figures 1b to 1d: in Figures 1e to 1g we display the 95% confidence intervals for the fit of a linear model to the raw data. The effect of replication is to reduce the confidence intervals, and so produce a more certain answer. Without replication the confidence bands are extremely wide, suggesting that there is little evidence for either the absence or presence of linearity despite a high correlation between the two variables. In Figure 1f the analytical error is quite small so the confidence bands are very tight. As the analytical error increases (Figure 1g) the confidence bands are broader and almost as wide as in Figure 1e. This demonstrates the importance of measuring analytical error via replicates analysis as well as the need to properly design experiments allowing sufficient degees of freedom to assess how well the model is fit.

Downloaded At: 18:34 18 January 2011

2.3 Coding

Another frequently overlooked, but crucially important, aspect of experimental design involves converting the raw response to a form that can be analysed directly by standard statistical approaches.

Environmentalists might want to determine whether acidity is linked to the presence of a pollutant. How should acidity be measured? Should it be expressed in the form $[H^+]$ or as pH? The latter scale is logarithmically related to the first scale. The answer is that we choose whichever method of scaling or *coding* according to what is most convenient for subsequent dataprocessing. Negative values of logarithms are physically meaningful, whereas negative values of measurements such as concentrations or plant root lengths do not have a physical meaning. Also many natural processes can be modelled by exponential decays or asymptotic behaviour: it is hard to fit exponential models, but when converted to a logarithmic scale these models reduce to linear models which are relatively easy to fit by standard regression techniques. So, very frequently, the factor variable is logarithmically coded.

In addition to taking logarithms it is often important to establish the region over which an environmentally significant response is expected. For example, there is no point studying the response of Tl in molar concentrations, as plants are almost certain to be dead at these concentrations, and the physiological processes that are studied at low concentrations often break down. In many types of experiment, a concentration or factor at the "centre" of the experiment is established: this concentration is then coded to correspond to a value 0; lowest and highest "interesting" concentrations are also established, normally equidistant

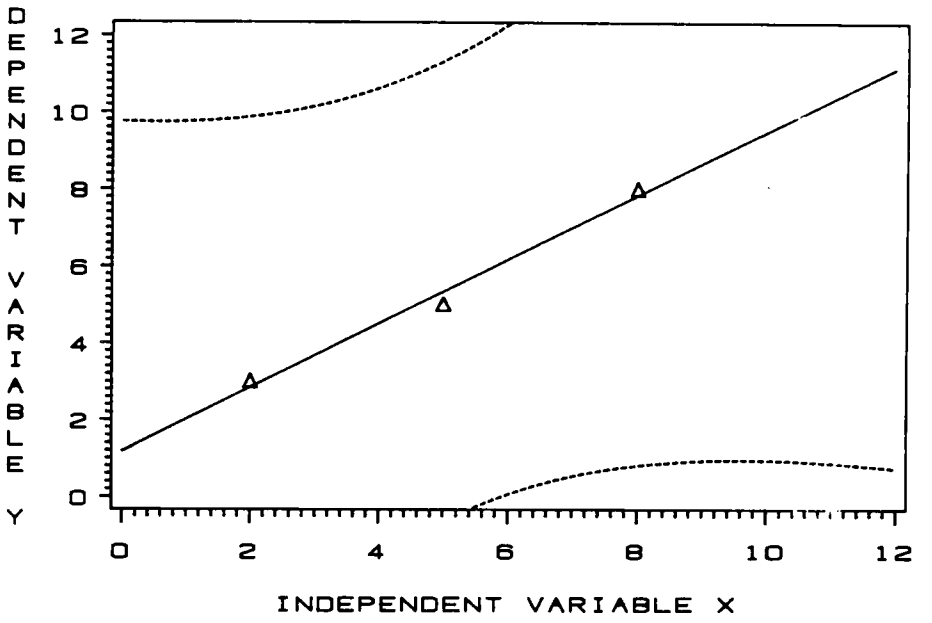


Figure 1e 95% confidence bands for data in Figure 1b.

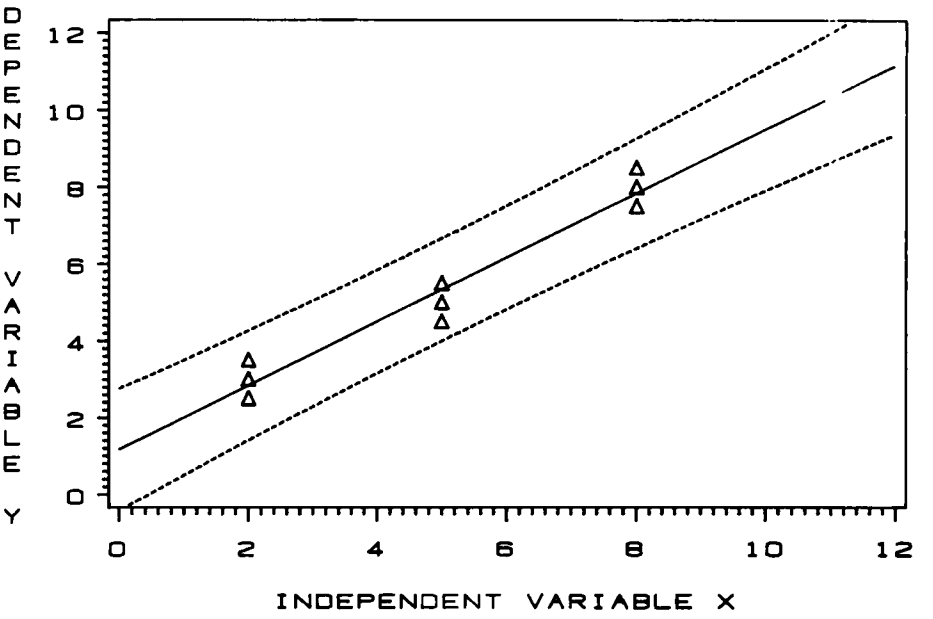


Figure 1f 95% confidence bands for data in Figure 1c.

Downloaded At: 18:34 18 January 2011

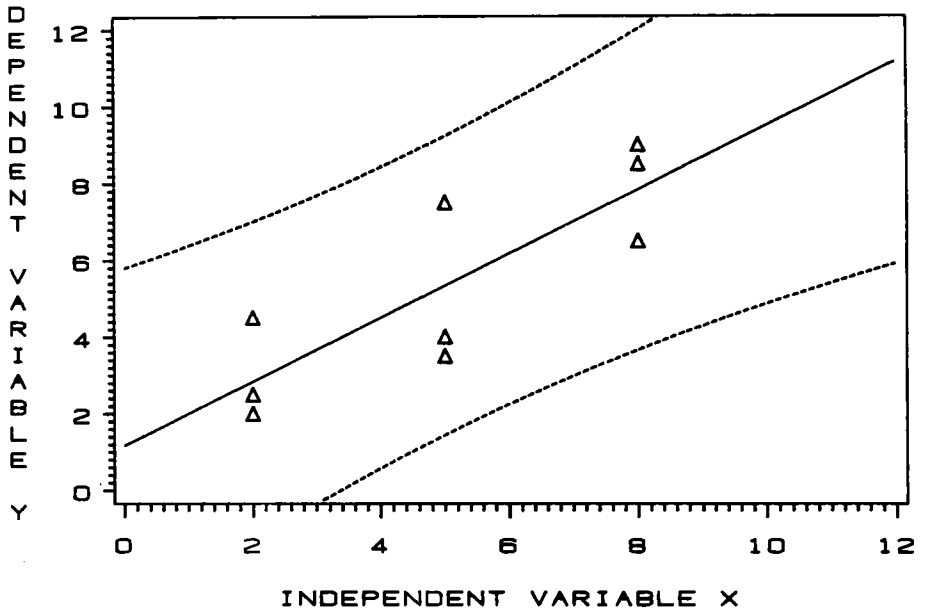


Figure 1g 95% confidence bands for data in Figure 1d.

Downloaded At: 18:34 18 January 2011

(in coded scale) from the centre point. These are normally set to correspond to points at ± 1 in the design. Obviously the choice of the method of coding requires some intuitive estimates of the process being studied.

As well as coding the "factors" it is often useful to scale the response.

2.4 Models

Another critical consideration of any experiment is how to model the process being studied. It is important to establish what model is to be tested, prior to performing the experiments. Above we considered a simple linear model: is y linearly related to x? In order to have any information to answer this question, at least 3 non-replicated experiments are required. This is because the model is a 2 parameter model.

In more complex situations it is usual to employ *multilinear* models. For a single factor experiment

$$y' = b_0 + b_1 x' + b_{11} x'^2 + b_{111} x'^3 + \cdots \tag{3}$$

where y' is the coded response and x' is the coded factor. Clearly for an nth order model at least $n+1$ measurements are required.

However, in most situations we employ *multifactor* experiments. In such cases the influence of several factors on an observed response is of interest. These types of experiments are closer to reality. For example, we might grow plants in the

presence of varying concentrations of 3 different metals. Extending the quadratic form of Eq. (3) to a multifactor response we have

$$y' = b_0 + \sum_{i=1}^{i=3} b_i x_i' + \sum_{i=1}^{i=3} b_{ii} x_i'^2 + \sum_{i=1}^{i=2} \sum_{j=1}^{j=3} b_{ij} x_i' x_j' \quad (4)$$

where x_i' is the coded value of factor i. As well as the normal linear and quadratic terms, cross product terms of the form b_{ij} are introduced to the model. These *interaction* terms can play a crucial role in relating the laboratory experiments to reality. Single factor (univariate) experiments cannot provide such information and often lead to completely misleading conclusions. Responses are rarely additive or independent. For example if the dry weight of a plant in the presence of x_1 μg/g Tl is y_1 g, and in the presence of x_2 μg/g Cd is y_2 g, is the dry weight in the presence of x_1 μg/g Tl and x_2 μg/g Cd $y_1 + y_2$ g? This is unlikely, so it is important to discover how different factors "cancel" each other out. The lack of information about interactions often causes serious misunderstandings about interpretation of laboratory based studies in the field. For example, consider a laboratory based experiment that is used to conclude that the abundance of a given series of chemicals in algal cells changes with temperature. Can these results be used as indicators of temperature in the field? Other factors are also likely to influence the abundance of these chemicals such as light intensity and wavelength, salt concentration, phase in lifecycle etc. Unless information about the magnitude of interactions is obtained, laboratory based experiments may be irrelevant when used to predict field behaviour.

The interactions and the form of the univariate response curves are predicted by the model. The model, in turn, determines the minimum size of experiment. For example, a 3 factor experiment with quadratic models in the form of Eq. (4), requires 10 parameters. Therefore, there must be at least 11 non-replicated points in the design, preferably a few more in order to provide a few degrees of freedom to assess the goodness of fit to the model.

Naturally if cubic and other higher order terms are suspected in order to provide a better fit to the model, more experiments will be required. So the nature of the questions to be asked and the model must be determined prior to experimentation.

3. AN EXPERIMENT: THE INFLUENCE OF HEAVY METALS ON THE GROWTH OF PLANTS

3.1 The Conventional Approach

The conventional approach to studying the influence of heavy metals on the growth of plants is to grow plants in varying concentrations of each metal and monitor some parameter such as the dry weight, wet weight, root length, shoot length and so on.

Figure 2 is a graph of the dry weight of barley seedlings after 4 weeks' growth in varying concentrations of Tl and Cd. This experiment suggests that plant growth

Downloaded At: 18:34 18 January 2011

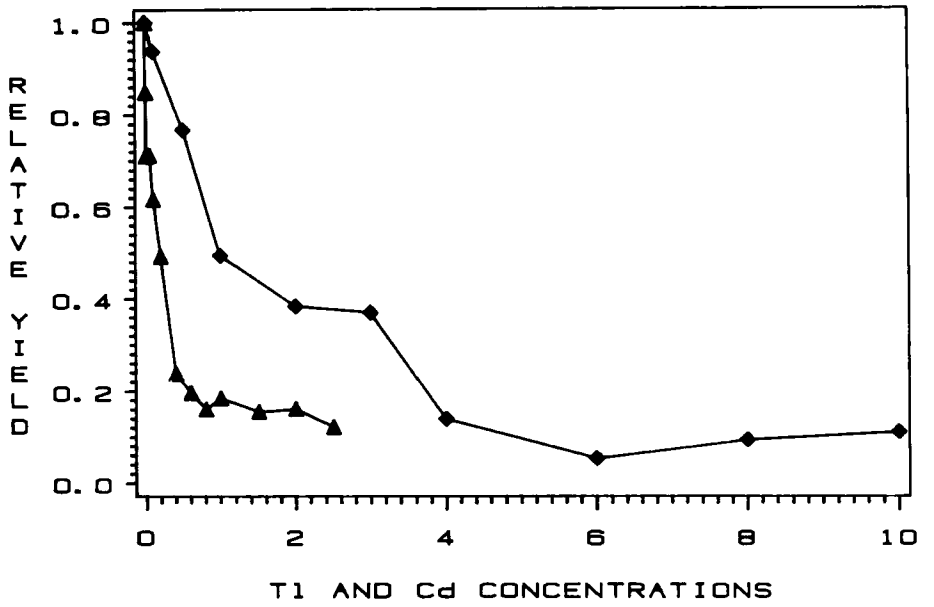


Figure 2 Relative univariate response curves for adding Tl and Cd separately in $\mu g/cm^3$, scaled so that the dry weight at 0 concentration of each metal is 1.0 (the points at higher concentrations for a given metal are relative to the yield at 0 concentration, and each set of growth experiments for a given metal was performed simultaneously). ▲ Tl and ◆ Cd.

Downloaded At: 18:34 18 January 2011

is suppressed at high concentrations of both Tl and Cd, so both metals are toxic. It also suggests that Tl is more toxic than Cd. Elsewhere we report similar conclusions for the root and shoot lengths.[22] However, very little further information can be gained from these experiments. There is no information about the interaction effect (i.e. what happens when both Tl and Cd are in high concentrations). There is no information as to how well, quantitatively, the model is obeyed or as to how certain and reproducible our conclusions are. There is no information about the experimental error, therefore the lack of fit cannot be tested.

3.2 Three Factor Central Composite Design

In order to improve the experiment, we employ experimental design methodology. Full details of growth conditions, coding and the theory of the design, and the computational methods have been reported by us elsewhere,[22–26] so we limit this discussion to essential information.

A *central composite design* is chosen. *Three factors*, corresponding to 3 metals are chosen: for Experiment A these factors are Tl, Cd and Pb and for Experiment B, these are Tl, Fe and Zn; hence Experiment A is used to look at the relative toxicities of Tl, Cd and Pb and their interactions.

Table 1 Coded concentrations for each experiment

Experiment A	*Tl*	*Cd*	*Pb*
c_{lo}	0.001	0.001	0.001
c_0	0.10	0.10	0.10
c_{hi}	10.00	10.00	10.00
Experiment B	*Tl*	*Fe*	*Zn*
c_{lo}	0.001	0.0025	0.005
c_0	0.10	0.25	0.5
c_{hi}	10.00	25.00	50.00

The metal concentrations are coded as follows. If x_i' is the coded concentration corresponding to the actual concentration x_i, then

$$x_i' = 1.68 \cdot (\log(x_i/x_{i0}))/\log(x_{ihi}/x_{ilo}) \tag{5}$$

where

$$\log(x_{i0}) = [\log(x_{ilo} \cdot x_{ihi})]/2 \tag{6}$$

The values of x_{ilo} and x_{ihi} are chosen so that they are at the low and high concentration ranges of interest in the experiment. x_{i0} is the geometric mean of these two parameters. From the above analysis it can be shown that the *coded* value of x_{ilo} is -1.68, that of x_{i0} is 0 and that of x_{ihi} is $+1.68$. The investigator chooses the values of x_{ilo} and x_{ihi} according to reasonable observations, so that the experimental region is one of high variability and interest. The values employed in this study are given in Table 1. It is important to note that these values differ according to metal. This is because some metals are expected to be more toxic than others: the least toxic metals will show significant effects at higher concentrations.

Once the factors are coded, standard experimental designs are employed which tell the experimenter which combination of factors should be used in the growth (or similar) experiments. For the three factor, central composite, design, the combination of factors given in Table 2a are employed. These can then be interpreted in terms of experimental conditions; for example, condition 8 involves setting x_1, x_2, and x_3 at $+1$: for Experiment A this involves growing barley seedlings in 1.55 $\mu g/cm^3$ of added Tl, Cd and Pb.

Several features of the design should be noted. First, heaviest experimentation is in the middle. Second, there are 5 replicates in the centre (coded 0, 0, 0) which are used to assess analytical errors. Third, there are 15 non-replicated experiments, allowing 14 degrees of freedom to assess goodness-of-fit to the model. If a quadratic model of the form of Eq. (4) is employed, there are 10 terms to estimate, so there are sufficient degrees of freedom to assess such a model. It would not be

Downloaded At: 18:34 18 January 2011

Table 2a Design used in the experiment

Condition	x_1	x_2	x_3
1	−1	−1	−1
2	−1	−1	1
3	−1	1	−1
4	−1	1	1
5	1	−1	−1
6	1	−1	1
7	1	1	−1
8	1	1	1
9	1.68	0	0
10	−1.68	0	0
11	0	1.68	0
12	0	−1.68	0
13	0	0	1.68
14	0	0	−1.68
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	0

Note: The values correspond to coded concentrations of the three metals as indicated in Table 1.

Downloaded At: 18:34 18 January 2011

possible to fit a complete cubic model which would require 10 extra terms, although it is possible to test for individual cubic terms if required.

The experiment can also be represented by a cube (Figure 3). Each experiment is represented by a point on this cube.

It must be emphasized that there are a huge number of possible experimental designs, and we do not discuss the merits of choice of design in this paper. The interested reader should, however, carefully consider the nature of the experiment and model building prior to choosing a design.

3.3 Result of Experimentation

The results of experimentation is a measured response for each of the growth conditions given in Table 2b. The 20 dry weights (or yields in g) for experiments A and B are listed in Table 3. These can be interpreted in a variety of ways and help us answer several questions about the natural processes.

4. INTERPRETATION OF THE EXPERIMENT

4.1 Regression and Curve Fitting

The first step is normally to perform multilinear regression analysis on the

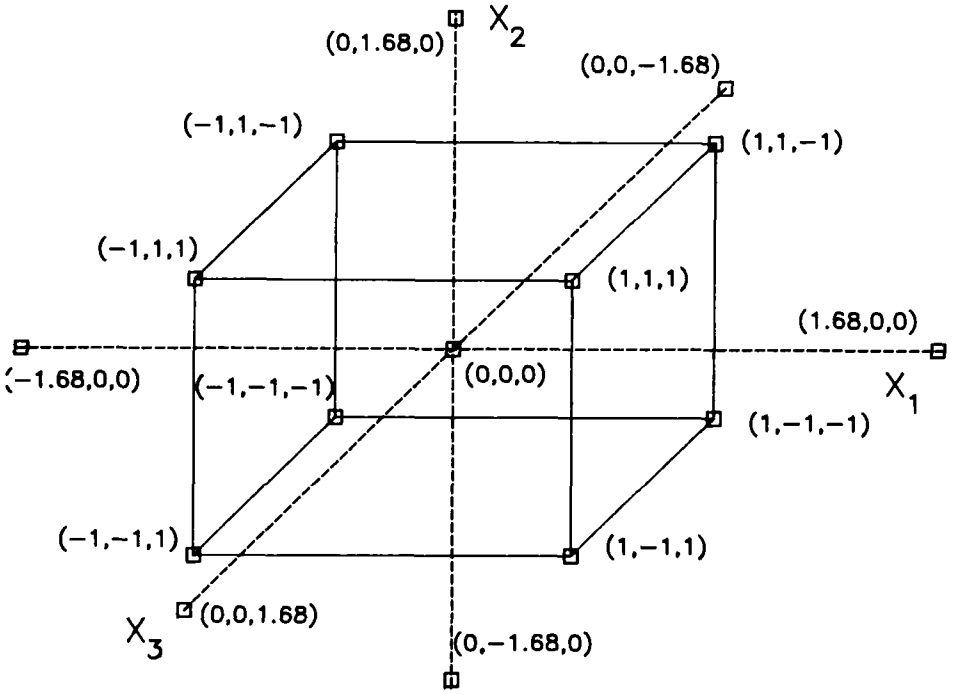


Figure 3 Cubic representation of the experimental layout (Table 2). Each point corresponds to one or more condition in the experimental design. The axes correspond to coded metal concentrations (x_1, x_2 and x_3) as discussed above.

Downloaded At: 18:34 18 January 2011

data.[22,23] In our experiments 10 coefficients can be obtained from each experiment A and B, corresponding to the 10 parameters. There are a variety of ways of scaling the responses: in this study we take the logarithm (to the base 10) of the dry weight: it is, of course, possible to fit similar models to linear responses. The resultant parameters are tabulated in Table 4; remember that the factors (x_i) are also scaled and that they are not all on the same scale (see Table 1). This will be discussed in greater detail below, but is best performed by graphical representations.

However, these parameters are probably best interpreted graphically. The parameters b_i and b_{ii} can be used to reconstruct *univariate response curves*, similar in concept to those of Figure 1a, but using reconstructed rather than directly observed data. The univariate response curves are illustrated in Figure 4: these curves are reconstructed from the regression model using $x_{j'}$ and $x_{k'}$ equal to zero, i.e. the response curve for Tl (Expt. A) is reconstructed setting the *coded* concentration of Cd and Pb at zero: it is important to realise that the coded 0 is not the same as the absence of the other two metals. Note that the concentrations are converted back to a linear scale; this means that the toxicity curves are now on the same absolute scale and the distortion of coding on the estimated coefficients (which is for experimental convenience) is now eliminated. It is, of course, possible to set the concentrations of the other metals at any value, and

Table 2b Asymmetric design as used in Figure 8

Condition	x_1	x_2	x_3
1	−3	0	0
2	−3	0	0
3	−2	0	0
4	−2	0	0
5	−1	−1	−1
6	−1	−1	1
7	−1	1	−1
8	−1	1	1
9	0	0	−2
10	0	0	0
11	0	0	0
12	0	0	2
13	0	2	0
14	0	3	0
15	1	−1	−1
16	1	−1	1
17	1	1	−1
18	1	1	1
19	2	0	0
20	2	0	0

Table 3 Plant dry weight (g) after 4 weeks' growth

Condition	*Experiment A*	*Experiment B*
1	0.93	1.61
2	0.80	0.79
3	0.61	0.95
4	0.61	1.31
5	0.28	0.21
6	0.24	0.19
7	0.25	0.29
8	0.18	0.20
9	0.21	0.17
10	0.88	1.67
11	0.31	1.33
12	0.83	1.05
13	0.67	0.49
14	0.88	1.36
15	0.85	1.24
16	0.78	1.17
17	0.78	1.29
18	0.81	1.15
19	0.75	1.29
20	0.70	0.95

Downloaded At: 18:34 18 January 2011

Table 4 Least square parameter estimates

Parameter	*Experiment A*	*Experiment B*
b_0	−0.243	0.177
b_1	−0.510	−0.761
b_{11}	−0.265	−0.377
b_2	−0.203	−0.056
b_{22}	−0.211	−0.096
b_3	−0.077	−0.187
b_{33}	−0.062	−0.227
b_{12}	0.039	0.052
b_{13}	−0.037	−0.010
b_{23}	−0.002	0.094

[a]Each metal concentration is scaled according to Table 1.
[b]Metals for Experiment A are 1-Tl, 2-Cd, 3-Pb and for Experiment B are 1-Tl, 2-Fe, 3-Zn.
[c]Dry weight is logarithmically scaled as discussed in the text.

Downloaded At: 18:34 18 January 2011

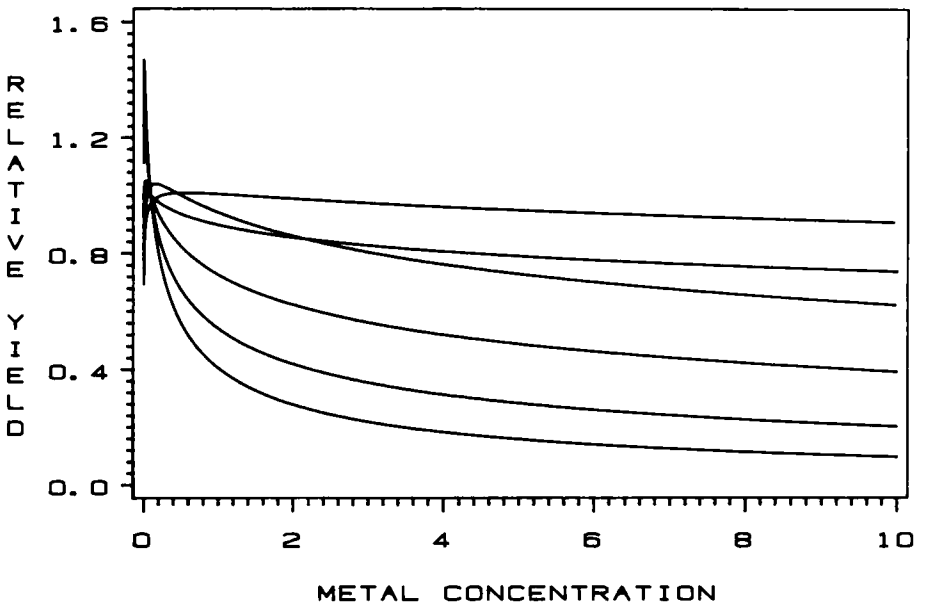


Figure 4 Reconstructed univariate response curves for Experiment A and B using relative yield (linear scale) and converting metal concentrations to $\mu g/cm^3$, from right hand top to bottom, Fe, Pb, Zn, Cd, Tl (Experiment A), Tl (Experiment B).

these univariate response curves should be regarded as a 2-dimensional projection of the 4 dimensional hyperspace in which the estimated response in the presence of varying concentrations of all 3 metals is displayed. The Tl response curves for both experiments are very similar, giving us confidence in the process; Cd is clearly less toxic than Tl. Zn and Pb have fairly similar toxicities and Fe is the least toxic.

Comparing Figure 2 and Figure 4 we clearly see that using our experimental

design smoother curves are obtained. Also only 40 points were needed to produce the data in Figure 4, compared to 23 points for Figure 2. Yet Figure 2 only provides information on the influence of two different metals on the observed response, and there is no quantitative information about confidence intervals, degrees of freedom and so on. There is absolutely no information about interaction effects. This demonstrates the increased efficiency of the design.

Another way of illustrating relative toxicities is by plotting *response surfaces.* Figure 5 is the response surfaces for Tl and Cd, at a coded value of Pb of 0. These 3 dimensional reconstructions are the estimated response for any combination of metal concentrations. There will be 3 response surfaces for any 3 combinations of metals. The surface pictured in Figures 4 and 5 shows clearly the greater toxicity of Tl as opposed to Cd: the "downward" slope is much greater along the x_1 axis as opposed to the x_2 axis, further reinforcing the evidence from the univariate response curves: in fact these reconstructions are very sensitive indicators of relative toxicities.[22] As with the univariate response curves, these surfaces are merely projections from 4 dimensional hyperspace, and the shape of the surfaces will appear different according to the value of the third metal chosen. Normally a coded value of 0 is used, as in this study.

Another point to note is the size of the interaction effects given by the terms b_{ij}. The only terms that are large relative to the univariate terms are those for Tl and Cd: this might suggest that these two metals compete for similar sites in the cell, and so are not entirely additive.

Downloaded At: 18:34 18 January 2011

4.2 How Well have the Parameters been Predicted?

One advantage of designed experiments is that there is information as to how well the parameters have been predicted.

The conventional approach is *via* ANOVA (Analysis of Variance). Errors in predicted models are normally estimated using least squares criteria: for example the greater the size of the residuals least square errors, the poorer the fit to the model. It is normal to compare this error to the overall analytical (replicate) error. The ratio of these variances is then computed. A statistic such as the F-test is then used to indicate what percentage confidence there is in each parameter or model:[23,24] this test assumes that errors are normally distributed. Usually 95% significance is indicated by *, 99% by ** and 99.9% by ***. The results are usually presented in tabular form, as in Table 5. The mean least squares estimates of all 10 parameters and of the overall model are tested against the overall mean analytical (replicate) error.

We discuss ANOVA in detail elsewhere,[23,24] but it is important to recognise that chemometricians often have to interpret ANOVA results with some caution. In Table 5 we see that many of the parameters are highly significant relative of the analytical (replicate) error (that is we are highly confident that these results are correct within the limits imposed upon the model by the analytical errors) but also that there is a significant lack-of-fit. In chemometrics applications experiments are often far more reproducible than in, for example, biology, geology or psychology.

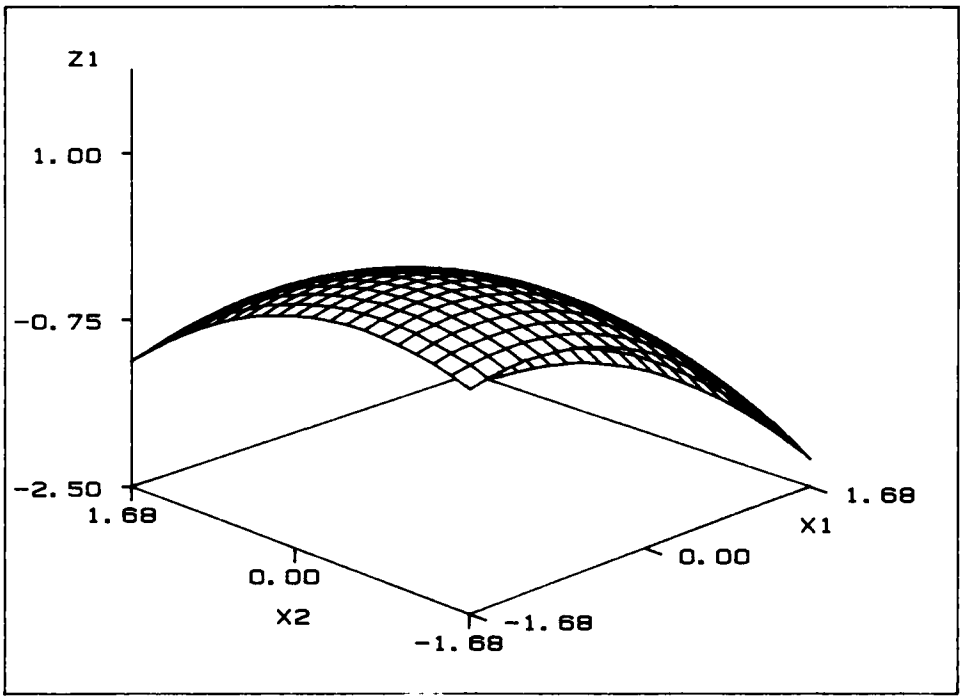


Figure 5a Response surface for Tl (x_1) and Cd (x_2) (Experiment A) using coded metal concentrations and logarithmic response (dry weight).

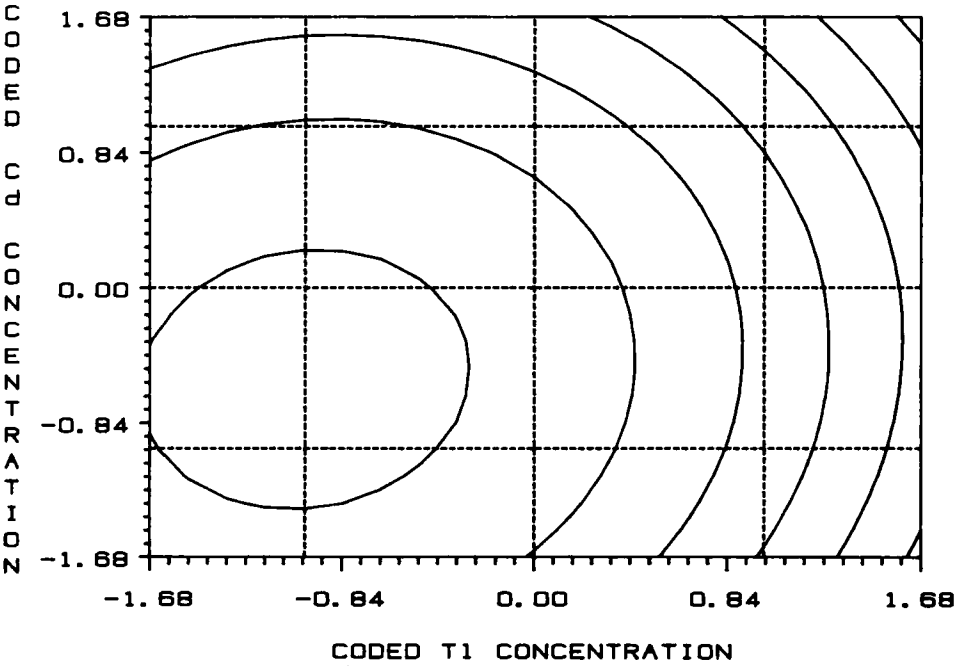


Figure 5b Contour representation of Figure 2. Levels from centre are −0.1, −0.5, −0.9, −1.3, −1.7, −2.1 and −2.5.

Downloaded At: 18:34 18 January 2011

Table 5 Analysis of variance for log of dry weight (Experiment A)

Source	*Sum of squares*	*d.f.*	*Mean square*	*Mean square ratio*
x_1	3.53	1	3.53	778.1***
x_2	0.56	1	0.56	124.2***
x_3	0.081	1	0.081	17.8*
$x_1{}^*x_1$	0.85	1	0.85	186.4***
$x_1{}^*x_2$	0.012	1	0.012	2.68
$x_2{}^*x_2$	0.62	1	0.62	134.5***
$x_1{}^*x_3$	0.011	1	0.011	2.45
$x_2{}^*x_3$	0	1	0	0
$x_3{}^*x_3$	0.056	1	0.056	12.2*
Total reg.	5.716	9	0.635	
Lack-of-fit	0.405	5	0.081	17.8**
Treatments	6.121	14	0.437	
Anal. error	0.0227	5	0.00454	
Total	6.1431	19		

Analysis of variance for log of dry weight (Experiment B)

Source	*Sum of squares*	*d.f.*	*Mean square*	*Mean square ratio*
x_1	7.91	1	7.91	578.7***
x_2	0.043	1	0.043	3.12
x_3	0.48	1	0.48	35.1**
$x_1{}^*x_1$	1.78	1	1.78	1129.9***
$x_1{}^*x_2$	0.021	1	0.021	1.56
$x_2{}^*x_2$	0.079	1	0.079	5.81*
$x_1{}^*x_3$	0	1	0	0
$x_2{}^*x_3$	0.071	1	0.071	5.53*
$x_3{}^*x_3$	0.74	1	0.74	54.2***
Total reg.	11.121	9	1.236	
Lack-of-fit	1.102	5	0.221	16.1**
Treatments	12.223	14	0.873	
Anal. error	0.0683	5	0.0137	
Total	12.2916	19		

Notes: [a]Significance levels are *95%, **99% and ***99.9%.
[b]For detailed interpretation see standard texts and references in paper.
[c]d.f. = degrees of freedom.

For example a biometrician might wish to study the reflexes of individuals of different ages: there will clearly be much less reproducibility among individuals. Although, in our case, the plants, themselves may be subject to variability, the analytical process (e.g. measuring a root length) is fairly reproducible and the concentration of chemicals metals can be determined with a high degree of accuracy. Thus analytical errors are often relatively small; normal ANOVA methods compare the experimental errors to the analytical error: clearly if this latter error is small the entire analysis is influenced. However, provided output from computers is treated with some caution ANOVA can be a valuable tool.

Downloaded At: 18:34 18 January 2011

Complementary to numerical analysis is the possibility of graphical representation of confidence intervals.[25] It is usual to compute 95% confidence intervals. There are various criteria for confidence limits, and, as is usual in chemometrics, it is important to work out in advance what is required from the experiment. Three common criteria are as follows: (1) means—this is the confidence limit for the mean readings at each point—i.e. the averaged reading, after replicates have been averaged; (2) individual—this is the confidence limit for the predicted response without averaging replicates—it will always be wider than the mean confidence interval as it takes into account analytical error; (3) Working—Hotelling—this is the confidence limits for the "surface" as a whole, and is related to the design of the experiment.

These confidence limits for the response surface of Tl and Cd are illustrated in Figure 6a. It is often clearer to consider only one pair of confidence intervals: Figure 6b illustrates the 95% confidence intervals for the mean alone. A key message of chemometrics is that it is important to choose which criterion is required.

It is, of course, also possible to compute the univariate confidence limits as are illustrated in Figure 6c for the case of Tl. Note that the scale is still coded (i.e. logarithmically related to concentration). Obviously the advantage of using modern graphical methods (such as in SAS—Statistical Analysis System) is that it is easy to display relevant information in a large variety of ways: we have only selected a few of the possibilities in this paper, for the sake of brevity.

Downloaded At: 18:34 18 January 2011

4.3 How Useful is the Design?

Most conventional statistical texts provide detailed mathematical tests as to how well a given design recovers parameters. For example, if most experimentation is in the centre of the design, then there is higher *confidence* in the recovered parameters in the centre of the design. If there is heavy experimentation on the outside of the experimental region, then the *uncertainty* in the recovered parameters is more even throughout the experimental region.

This can be visualised by plotting the *leverage* over and outside the experimental region;[24] this is illustrated for the central composite design in Figures 7a and 7b for a two factor response surface: note that the region chosen is asymmetrical and broader than the experimental region. Leverage is dependent entirely on the design, but the confidence intervals are related to leverage. The higher the leverage, the lower the confidence in the fitted results; the shape of the confidence bands is entirely dependent on leverage but the absolute size of the confidence bands depends on the mean square error from the observed experiment. Clearly the leverage for the experiment in question is least in the middle, is fairly smooth and is symmetrical. The effect of changing the number of replicates, position of outliers and the symmetry of the design can best be seen *via* visualising the leverage, and is a valuable aid to the experimenter prior to obtaining data. The leverage for two factors in an asymmetric experiment (Table 2b) is illustrated in Figure 8. It is clear that the leverage, in turn, is also asymmetrical. In addition to examining the influence of symmetry, we might want to look at the influence of outlying

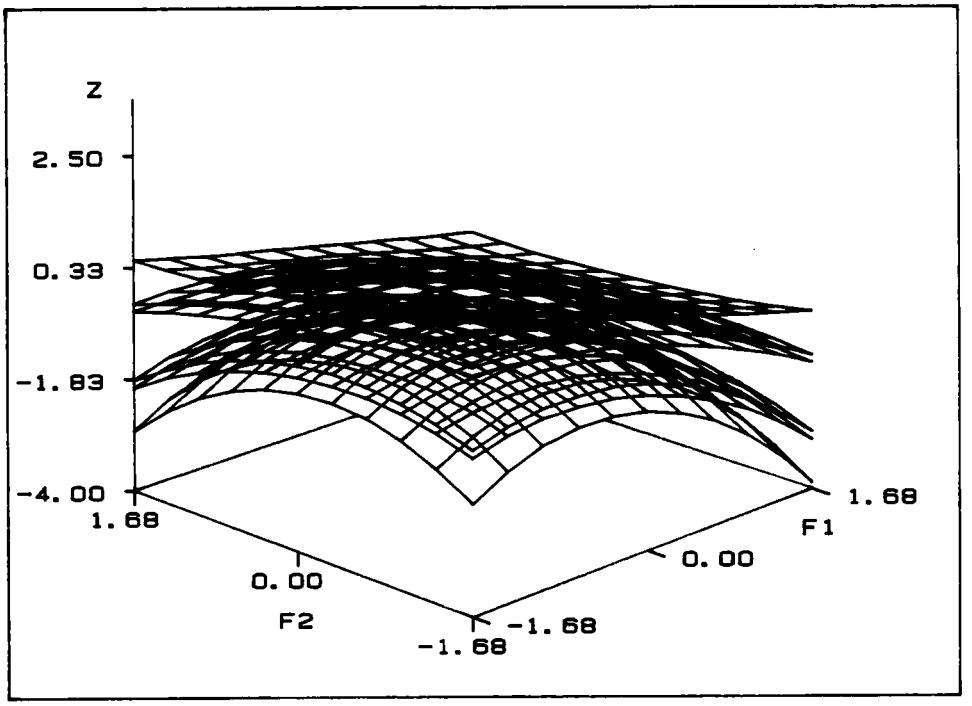


Figure 6a Confidence bands for Tl and Cd response surface (Figure 5). From left hand top to bottom 95% upper limit for (1) Individual responses (2) Working–Hotelling (3) Means responses; lower 95% limit for (4) Mean responses (5) Working–Hotelling (6) Individual responses.

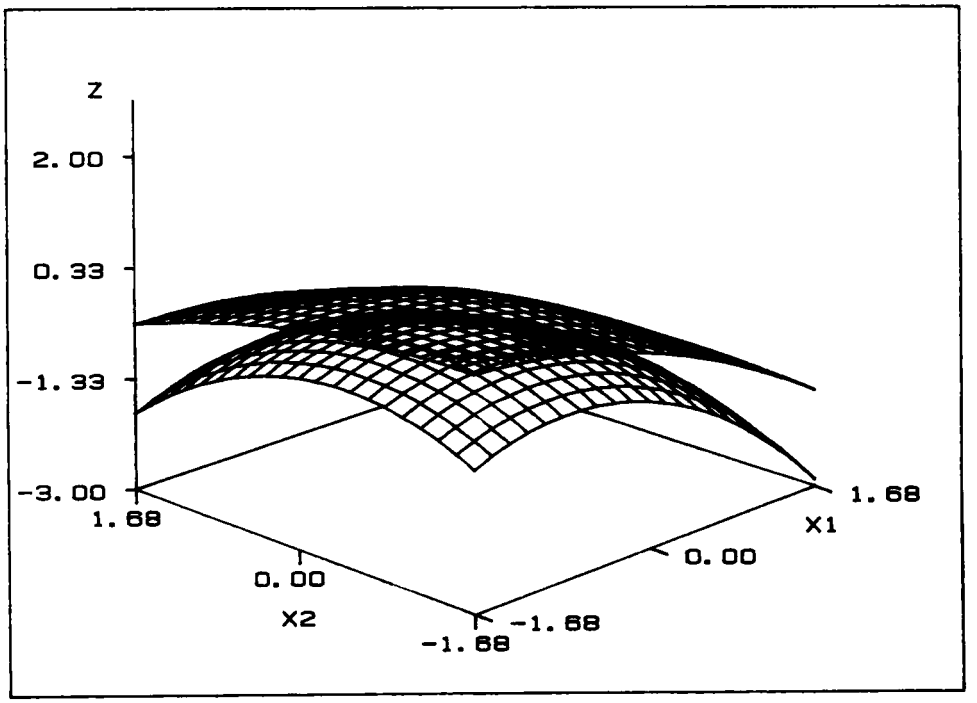


Figure 6b Upper and lower 95% confidence intervals for the mean response alone, corresponding to Figure 6a.

Downloaded At: 18:34 18 January 2011

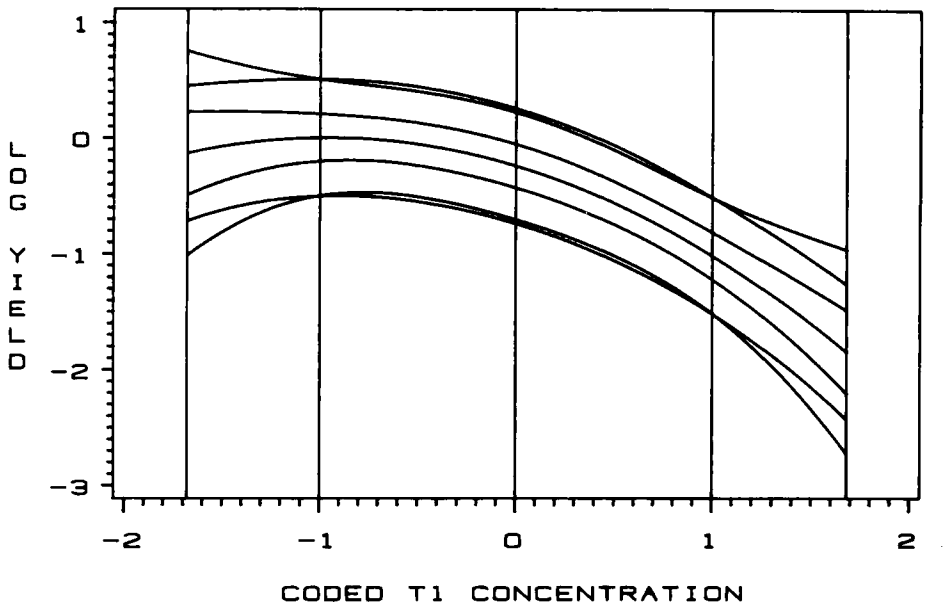


Figure 6c Univariate confidence bands for Tl (Experiment A) (logarithmically coded concentrations and response as discussed in text). From left hand top to left hand bottom; 95% upper limit for (1) Working–Hotelling (2) Individual responses (3) Mean responses; (4) the predicted univariate response; 95% lower limit for (5) Mean responses (6) Individual responses (7) Working–Hotelling.

Downloaded At: 18:34 18 January 2011

experiments. In the central composite design described above, outliers were set at ± 1.68; another design called a *face centred cube* sets these outliers at ± 1.00 instead (so that conditions 9 to 14 in Table 2 differ). The leverage for this design is illustrated in Figure 9. Clearly the leverage differs considerably. It is also, of course, possible to examine the effect to changing the number and distribution of replicates. This is discussed in more detail elsewhere.[24,25,26]

5. CONCLUSION

In this paper we have shown the advantages of using experimental design and so employing a chemometric approach to obtaining environmental analytical data.

It is likely that experimental design will be employed far more widely in chemometrics. Of especial interest is the development of truly *multivariate* response surface methodology: in the example cited above the response is a univariate one (a plant dry weight). In many real situations, the response might be multivariate, e.g. the concentration of several compounds. The response can then be modelled by a multivariate parameter such as a principal component or PLS component and so multivariate method for dimensionality reduction can be combined with methods for experimental design.[26] Analogously methods for time series analysis

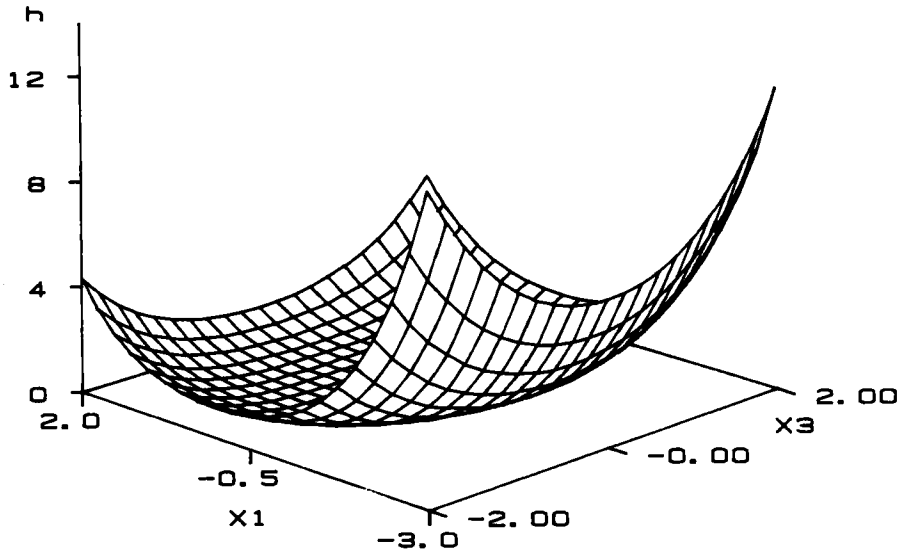


Figure 7a Three-dimensional representation of leverage for central composite design (Table 2a) as discussed in this paper.

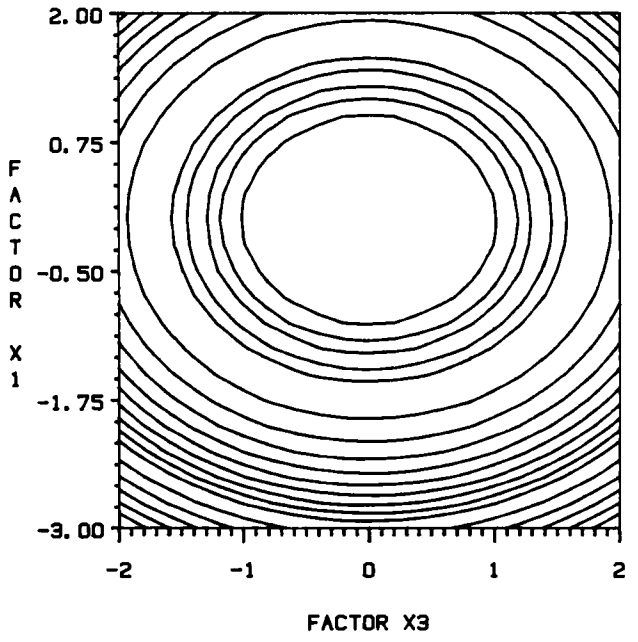


Figure 7b Contour representation of leverage in Figure 7a.

Downloaded At: 18:34 18 January 2011

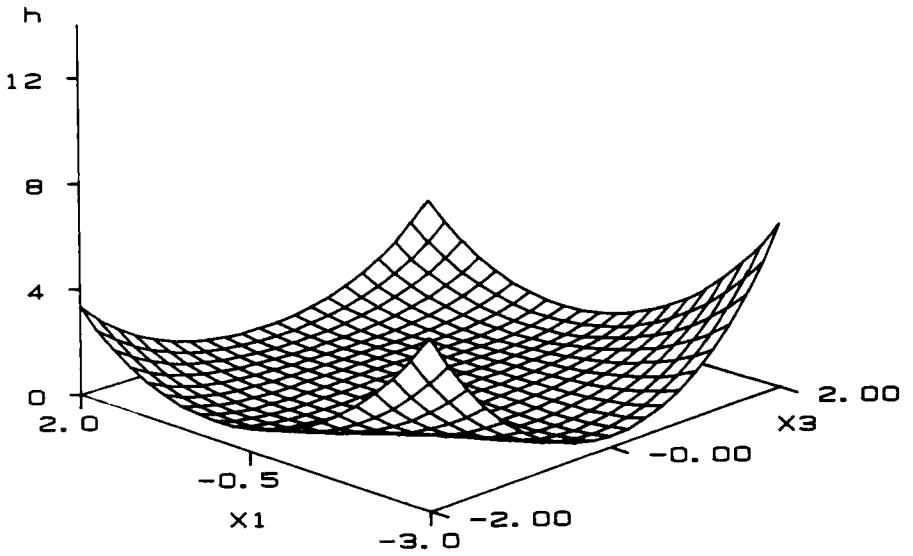


Figure 8a Three dimensional representation of leverage for the asymmetric design in Table 2b (x_1 versus x_3).

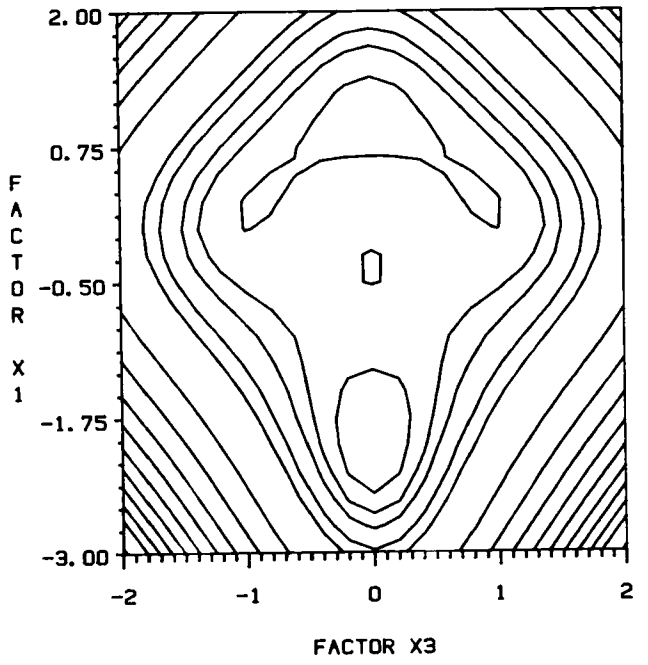


Figure 8b Contour representation of leverage in Figure 8a.

Downloaded At: 18:34 18 January 2011

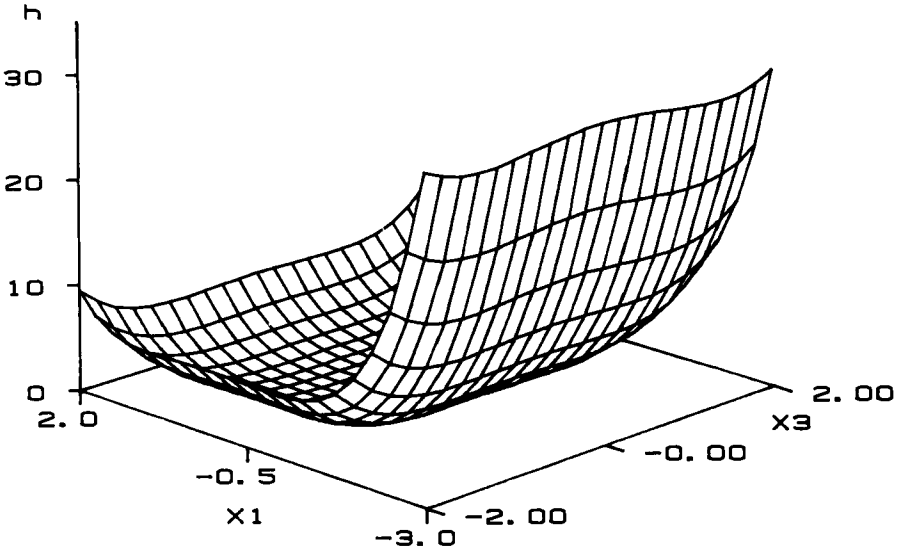


Figure 9a Three-dimensional representation of leverage for face centred cube design as discussed in text.

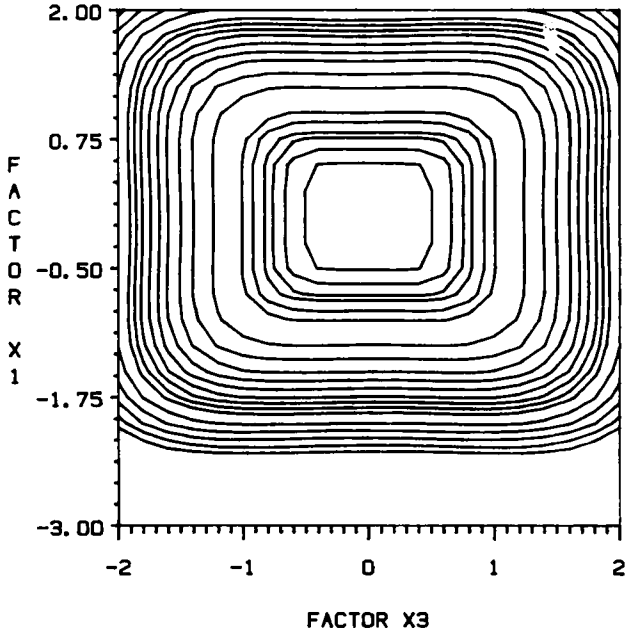


Figure 9b Contour representation of leverage in Figure 9a.

Downloaded At: 18:34 18 January 2011

have been combined with multivariate methods for the study of geochemical cyclicity.[27,28]

There remain many exciting developments in the area of experimental design. Within the field of chemometrics this is particularly important since, with the advent of modern analytical instrumentation and rapid facile on-line computing power, large quantities of data can be generated, so methods for meaningful design of sampling strategies are vital. Special statistical approaches need to be developed to cope with multivariate responses and relatively high analytical reproducibility.

The environmental analytical chemist must be aware of methods for experimental design: it is important to realise that meaningless results can be obtained if experiments are designed with an inappropriate number of degrees of freedom, for example. Often considerable time and effort can be spent on acquiring and interpreting experimental data: this time can be wasted if insufficient thought is given to the nature of the design; considerations such as the model, the distribution of points, replicates analysis, coding, size of relative errors all need to be taken into account.

In conclusion, chemometrics is likely to become an indispensible tool of the environmental analytical chemist.

Downloaded At: 18:34 18 January 2011

Acknowledgements

We thank Dr G. Nickless for helpful advice and encouragement, and the Cultural Department of the Ministry of Education and Scientific Research Institute (Tripoli, Libya) and The Department of Chemistry, Faculty of Science, Garyounis University (Benghazi, Libya) for finance for a research studentship (M.A.A.). Funds from NERC and Bristol University Computer Service are gratefully acknowledged to run SAS on the Department of Chemistry VAX 11/750 computer. Professor J. Albaiges and Dr J. Grimalt are thanked for an invitation (R.G.B.) to present these results at the 18th International Symposium on Environmental Analytical Chemistry held in Barcelona, 5–8 September, 1988.

References

1. H. J. H. Bowen, *Environmental Chemistry of the Elements* (Academic Press, London, 1979).
2. V. Zitko, *Science of the Total Env.* **4**, 185 (1975).
3. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte and L. Kaufman, *Chemometrics: A Textbook* (Elsevier, Amsterdam, 1988).
4. M. A. Sharaf, D. Illman and B. R. Kowalski, *Chemometrics* (Wiley, Chichester, 1986).
5. D. L. Massart (Editor-in-Chief), P. K. Hopke, C. H. Spiegelman, W. Wegscheider (Editors), R. G. Brereton and R. E. Dessy (Associate Editors), *Chemometrics and Intelligent Laboratory Systems* (Elsevier, Amsterdam).
6. B. R. Kowalski (Editor-in-Chief), S. D. Brown and B. G. M. Vandeginste (Associate Editors), *J. Chemometrics* (Wiley, Chichester).
7. R. G. Brereton, *The Analyst* **112**, 1635 (1987).
8. B. R. Kowalski, *Anal. Chem.* **52**, 112R (1980).
9. I. E. Frank and B. R. Kowalski, *Anal. Chem.* **54**, 232R (1982).
10. M. F. Delaney, *Anal. Chem.* **56**, 261R (1984).
11. L. S. Ramos, K. R. Beebe, W. P. Carey, E. Sanchez, B. C. Erickson, B. R. Wilson, L. E. Wangen and B. R. Kowalski, *Anal. Chem.* **58**, 294R (1986).

12. S. D. Brown, T. Q. Barker, R. J. Larivee, S. L. Monfre and H. R. Wilk, *Anal. Chem.* **60,** 252R (1988).
13. D. Bawden (Editor), "Chemometrics Series", Research Studies Press, Chichester.
14. B. R. Kowalski (Editor), Chemometrics: Theory and Applications, *ACS Symposium Series No. 52* (American Chemical Society, Washingon, DC, 1977).
15. B. R. Kowalski (Editor), *Chemometrics: Maths and Statistics in Chemistry* (Reidel, Dordrecht, 1984).
16. C. H. Spiegelman, R. L. Waters and J. Sacks (Editors), *J. Res. Natl. Bureau Stand., Special Issue* **90** (6), 391 (1985) (1988).
17. O. L. Davies, *The Design and Analysis of Industrial Experiments* (Oliver and Boyd, London, 1954).
18. R. H. Myers, *Response Surface Methodology* (Allyn and Bacon, Boston, MA, 1976).
19. D. M. Steinberg and W. G. Hunter, *Technometrics* **26,** 71 (1984).
20. S. N. Deming and S. L. Morgan, *Experimental Design: A Chemometric Approach* (Elsevier, Amsterdam, 1987).
21. N. L. Johnson and F. C. Leone, *Statistics and Experimental Design in Engineering and the Physical Sciences* (Wiley, New York, 1964).
22. M. A. Allus, R. G. Brereton and G. Nickless, *Chemometrics and Intelligent Laboratory Systems* **3,** 215 (1988).
23. M. A. Allus, R. G. Brereton and G. Nickless, *Environmental Pollution* **52,** 169.
24. M. A. Allus, R. G. Brereton and G. Nickless, *Chemometrics and Intelligent Laboratory Systems* **6,** 65 (1989).
25. M. A. Allus and R. G. Brereton, *SAS Users' Britain and Ireland Proceedings 1988,* 69 (1989).
26. M. A. Allus and R. G. Brereton, unpublished work.
27. R. G. Brereton, *Chemometrics and Intelligent Laboratory Systems* **2,** 177 (1987).
28. S. C. Brassell, R. G. Brereton, G. Eglinton, J. Grimalt, G. Liebezeit, U. Pflaumann and M. Sarnthein, *Org. Geochem* **10,** 649 (1987).

Downloaded At: 18:34 18 January 2011