4

Screening Methods

Francesca Campolongo

European Commission, Joint Research Centre, Ispra, Italy

Jack Kleijnen

Tilburg University, The Netherlands

Terry Andres

AECL Whiteshell Labs, Pinawa, Manitoba, Canada

4.1 INTRODUCTION

The central question of screening in the context of modelling and computer simulation is: Which factors—among the many potentially important factors—are really important? One of the aims in modelling is to come up with a *short list* of important factors (this is sometimes called the principle of parsimony or Occam's razor), and to do this, the choice of a well-designed experiment is essential.

This chapter gives a survey of statistical designs and analysis for screening, focusing on computer experiments. It is often assumed as a working hypothesis that the number of important factors is small compared with the total number of factors in a model (see e.g. Morris, 1987). This assumption is based on the idea that the influence of factors in models is distributed as income in nations, i.e. it follows Pareto's law, with a few, very influential factors and a majority of non-influential ones (see also Saltelli *et al.*, 1999a).

Screening designs are organized to deal with models containing hundreds of input factors. For this reason, they should be economical. As a drawback, these economical methods tend to provide qualitative sensitivity measures, i.e. they rank the input factors in order of importance, but do not quantify how much a given factor is more important than another. There is clearly a trade-off between computational cost and information.

Several approaches to the problem of screening have been proposed in the literature (see for instance, references given in Section 4.2 about screening techniques developed in the

Sensitivity Analysis. Edited by A. Saltelli *et al.*

context of physical experimentation). In this chapter, we shall describe a few of them through numerical examples and case studies. The methods described here include the one-factor-at-a-time (OAT) experiment proposed by Morris (1991), the design of Cotter (1979), the iterated fractional factorial designs (IFFDs) introduced by Andres (Andres and Hajas, 1993), and sequential bifurcation proposed by Bettonvil (Bettonvil, 1990; Bettonvil and Kleijnen, 1997). The examples used in this chapter demonstrate that these techniques can be simple, efficient, and effective.

The main conclusions of this chapter are as follows:

(i) Screening may involve several techniques, which are quite simple and yet efficient and effective.
(ii) These screening techniques have already been applied to several practical simulation studies, in different domains.

The remainder of this chapter is organized as follows. Section 4.2 gives notation, definitions, and key assumptions. Section 4.3 evaluates OAT designs. Section 4.4 describes the OAT design proposed by Morris. Section 4.5 describes the design type proposed by Cotter. Section 4.6 describes Andres' IFFD. Section 4.7 describes Bettonvil's sequential bifurcation. A summary and conclusions are given in Section 4.8.

4.2 DEFINITIONS

We use the term *factor* to denote any input included in the sensitivity study. Thus, a factor is any quantity that can be changed in the model prior to its running. This quantity can be a parameter, an input variable, or a module of the model. *Input variables* are directly observable in the corresponding real system, whereas *parameters* are not (they may be estimated). For example, parameters in simulation models of queuing problems in supermarkets and telecommunication systems are customer arrival and service rates; an input variable may be the number of parallel servers; a module may be the submodel for the priority rules (first-in-first-out or FIFO, shortest-processing-time or SPT, and so on).

By definition, factors are not kept constant during the whole experiment: a factor takes at least two *levels* or 'values'. A factor may also be *qualitative*. A brief discussion on quantitative and qualitative factors can be found in Chapter 3.

The term *metamodel* is used to denote the model of the underlying simulation model, i.e. the approximation of the simulation program's input/output (I/O) transformation.

Since the 1980s, screening design has become a common term. Some authors (e.g. Klejnen, 1998) restrict the term *screening designs* only to *supersaturated* designs, i.e. designs with fewer runs than factors. On the other hand, some others speak of screening designs when the number of runs is larger than the number of factors. For example, Myers and Montgomery (1995) refer to fractional factorials as being used as screening designs, with no hint of using supersaturated designs. Similarly, Rahni *et al.* (1997) refer to parameter screening designs as being equivalent to two-level experimental designs (but go on to discuss more efficient group screening methods). In this chapter, we shall use the term *screening design* to indicate any preliminary activity that, independently of the number of experimental runs it uses, aims to discover which of the input factors involved in a model are *important*, i.e. control most of the output variability.

Note that such a general definition does not necessarily imply the existence of a computer code, and can be applied in the case of a physical experiment. In the context of physical experimentation, a screening design indicates an experimental design that aims to discover

which of a collection
an observable respor
error have been prop
oldest used in physi
(1959), who used sup
models. Box and Me
few important factor
(1961), Srivastava (19
Watson (1961), propo
negligible factors in
Watson's idea, are th
(1984). However, in th
instead on screenin
screening designs in

Also excluded fror
Such a design in fa
not only at identifyir
to determine the w
other words, Welch *e*
strongly, and propos
accurate predictor.

A highly desirable
computational cost o
(computer runs) requ
in the analysis and of
requires a high numl
evaluation of the mo
method is to provide
while keeping the cor

4.3 ONE-AT-A-

The simplest class of
these designs, the im
1973). This approach i

The standard OAT d
is taken from the liter
'control' scenario. Tw
values for each factor
extremes. The magnit
the 'control' are then c

Although most con
implementing an OAT
into five categories:

- standard OAT desig
- strict OAT designs
  experimental run;

which of a collection of experimentally controlled factors have an important influence upon an observable response. Screening methods aimed at physical experiments with random error have been proposed by Plackett and Burman (1946), whose designs are among the oldest used in physical experimentation for screening purposes, and by Satterthwaite (1959), who used supersaturated random-balance designs for screening the factors in linear models. Box and Meyer (1986) suggested a Bayesian analysis based on the assumption of few important factors. Other designs are the supersaturated designs proposed by Watson (1961), Srivastava (1975), and Morris (1987), who, following the approach introduced by Watson (1961), proposed a generalization of the group screening technique for finding non-negligible factors in a first-order model. Similar works, which are still developments of Watson's idea, are those of Mauro and Smith (1982), Patel (1962), and Patel and Ottieno (1984). However, in the present chapter we shall not illustrate these methods. We shall focus instead on screening methods developed in the context of computer experiments. For screening designs in the context of physical experimentation, see also Chapter 3.

Also excluded from the present chapter is the design proposed by Welch *et al.* (1992). Such a design in fact focuses on an objective different from that treated here: it aims not only at identifying the subset of most important factors in the model, but also wants to determine the way in which those factors jointly affect the model's predictions. In other words, Welch *et al.* (1992) do not separate the screening and prediction objectives so strongly, and propose a design whose results can be used to both screen and to build an accurate predictor.

A highly desirable property of a screening method is its low computational cost. The *computational cost* of the experiment is defined as the number of model evaluations (computer runs) required. This cost is usually a function of the number of factors involved in the analysis and of the complexity of the input/output behavior. A screening exercise that requires a high number of model evaluations would be inappropriate, especially when the evaluation of the model requires much time. Thus, the main goal in developing a screening method is to provide adequate information about the sensitivity of the model to its inputs, while keeping the computational cost of the whole experiment low.

4.3 ONE-AT-A-TIME (OAT) DESIGNS

The simplest class of screening designs is that of the one-at-a-time (OAT) experiments. In these designs, the impact of changing the values of each factor is evaluated in turn (Daniel, 1973). This approach is also known sometimes as *ceteris paribus*.

The standard OAT designs use the 'nominal' or 'standard' value per factor; often this value is taken from the literature. The combination of nominal values for the k factors is called the 'control' scenario. Two extreme values are usually proposed to represent the range of likely values for each factor; normally the 'standard' value of a factor is 'midway' between the two extremes. The magnitudes of the differences between the outputs for the extreme inputs and the 'control' are then compared to find those factors that significantly affect the model.

Although most commonly used, the standard strategy is not the only one followed when implementing an OAT experiment. According to Daniel (1973), OAT designs can be classified into five categories:

- standard OAT designs, which vary one factor from a standard condition;
- strict OAT designs, which vary one factor from the condition of the last preceding experimental run;

- paired OAT designs, which produce two observations and hence one simple comparison at a time;
- free OAT designs, which make each new run under new conditions;
- curved OAT designs, which produce a subset of results by varying only one easy-to-vary factor.

In physical experiments, the practice of an OAT experiment is acceptable if the random error is small compared with expected main effects. Nevertheless, such a design may give biased estimates, the biases originating from the effect of interactions. Daniel (1973) proposed methods for refinement of estimates provided by an OAT design; the improvement comes in by removing two-factor interaction biases from main effect estimates. Daniel's methods have been shown to work best for strict OAT; they do not work so well for standard designs that have the limitation of their conservatism (Daniel, 1973).

In general, the number of model evaluations required of an OAT design is of the order of k (often, $2k+1$), k being the number of factors examined. The low computational cost is one of the main advantages of the OAT design. However, Kleijnen (1998, p. 195) argues that a resolution-3 design (see Chapter 3), requires only roughly $k+1$ runs, and provides more accurate estimators *of the main effects*.

One limitation of OAT designs is that they do not enable estimation of interactions among factors. (Likewise, resolution-3 designs allow only the estimation of the main effects.)

Furthermore, many OAT experiments in the literature are *local* experiments; that is, factors are changed over small intervals around their nominal values. These nominal values represent a specific point of the input space (the 'control' scenario). Results of such a local experiment are thus dependent on the choice of this point, and the model behavior is identified only locally in the input space (namely, around the selected point). This is acceptable only if the input–output relationship can be adequately approximated through a first-order polynomial. If the model shows strong nonlinearity then a change in the selected nominal values provides totally different sensitivity results. The limitation of such local experiments (sometimes also called 'elementary OAT', or EOAT) is highlighted in an application example illustrated in Chapter 18.

An OAT design that is not dependent on the choice of the specific point in the input space is that proposed by Morris (1991).

4.4 MORRIS'S (1991) OAT DESIGNS

We call the OAT design proposed by Morris (1991) a *global* sensitivity experiment, because his experiment covers the entire space over which the factors may vary (in a local experiment, the factors vary only around their nominal values, and the results depend on the choice of these values). Morris estimates the main effect of a factor by computing a number (say) r of local measures, at different points $\mathbf{x}_1, \ldots, \mathbf{x}_r$ in the input space, and then taking their average (this reduces the dependence on the specific point that a local experiment has). These r values are selected such that each factor is varied over its interval of experimentation.

Morris assumes an expensive (in computer power/time) computational model, or a model with a large number of factors; the number of computer runs needed by his design is proportional to k (number of factors). His design does not need simplifying assumptions about the input/output behavior. Morris wishes to determine which factors have (a) negligible effects, (b) linear and additive effects, or (c) nonlinear or interaction effects. His design is composed of individual randomized OAT designs, in which the impact of changing the value of each of

the chosen factors is evaluated in turn. In the terminology proposed by Daniel (1973), the Morris design is a strict OAT.

The k-dimensional factor vector $\mathbf{x}$ for the simulation model has components x_i that have p values in the set $\{0, 1/(p-1), 2/(p-1), \ldots, 1\}$. The region of experimentation Ω is then a k-dimensional p-level *grid*. In practical applications, the values sampled in Ω are subsequently rescaled to generate the actual (non-standardized) values of the simulation factors. Let Δ be a predetermined multiple of $1/(p-1)$. Then Morris defines the *elementary effect* of the ith factor at a given point $\mathbf{x}$ as

$$d_i(\mathbf{x}) = \frac{[y(x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_k) - y(\mathbf{x})]}{\Delta}, \tag{4.1}$$

where $\mathbf{x}$ is any value in Ω selected such that the perturbed point $\mathbf{x} + \Delta$ is still in Ω. A finite distribution (say) F_i of elementary effects for the ith input factor is obtained by sampling $\mathbf{x}$ from Ω. The number of elements of each F_i is $p^{k-1}[p - \Delta(p-1)]$.

The characterization of the distribution F_i through its mean μ and standard deviation σ gives useful information about the influence of the ith factor on the output. A high mean indicates a factor with an important overall influence on the output; a high standard deviation indicates either a factor interacting with other factors or a factor whose effect is nonlinear.

In its simplest form, the total computational effort required for a random sample of r values from each distribution F_i is $n = 2rk$ runs (k is the number of factors): each elementary effect requires the evaluation of y twice. Morris defines the *economy* of a design as the number of elementary effects estimated by the design, divided by the number of runs. The larger the value of the *economy* for a particular design, the better it is in terms of providing information for sensitivity and uncertainty analysis. The simplest form of Morris' design has an *economy* of $rk/2rk = 1/2$.

Morris proposed a more economical design than the simple design discussed so far. This design is based on the construction of a matrix $\mathbf{B}^*$ with rows that represent input vectors $\mathbf{x}$, for which the corresponding experiment provides k elementary effects (one for each input factor) from $k+1$ runs. This increases the economy of the design to $k/(k+1)$. In the development of such a design, it is convenient to assume that p is even and $\Delta = p/[2(p-1)]$. With such assumptions, each of the $p^{k-1}[p - \Delta(p-1)] = p^k/2$ elementary effects for the ith input factor has equal probability of being selected (see Morris, 1991). The key idea of the Morris design is the following:

1. A 'base' value $\mathbf{x}^*$ is randomly chosen for the vector $\mathbf{x}$, each component x_i being sampled from the set $\{0, 1/(p-1), \ldots, 1 - \Delta\}$.
2. One or more of the k components of $\mathbf{x}^*$ are increased by Δ such that a vector (say) $\mathbf{x}^{(1)}$ results that is still in Ω.
3. The estimated elementary effect of the ith component of $\mathbf{x}^{(1)}$ (if the ith component of $\mathbf{x}^{(1)}$ has been changed by Δ) is (see Equation (4.1))

$$d_i(\mathbf{x}^{(1)}) = \frac{y(x_1^{(1)}, \ldots, x_{i-1}^{(1)}, x_i^{(1)} + \Delta, x_{i+1}^{(1)}, \ldots, x_k^{(1)}) - y(\mathbf{x}^{(1)})}{\Delta} \tag{4.2a}$$

if $\mathbf{x}^{(1)}$ has been increased by Δ, or

$$d_i(\mathbf{x}^{(1)}) = \frac{y(\mathbf{x}^{(1)}) - y(x_1^{(1)}, \ldots, x_{i-1}^{(1)}, x_i^{(1)} - \Delta, x_{i+1}^{(1)}, \ldots, x_k^{(1)})}{\Delta} \tag{4.2b}$$

if $\mathbf{x}^{(1)}$ has been decreased by Δ.

4. Let $\mathbf{x}^{(2)}$ be the new vector $(x_1^{(1)}, \ldots, x_{i-1}^{(1)}, x_i^{(1)} \pm \Delta, x_{i+1}^{(1)}, \ldots, x_k^{(1)})$ defined in the above step. Select a third vector $\mathbf{x}^{(3)}$ such that $\mathbf{x}^{(3)}$ differs from $\mathbf{x}^{(2)}$ for *only one component j*: either $x_j^{(3)} = x_j^{(2)} + \Delta$ or $x_j^{(3)} = x_j^{(2)} - \Delta$, with $j \neq i$. The estimated elementary effect of factor j is then

$$d_j(\mathbf{x}^{(2)}) = \frac{y(\mathbf{x}^{(3)}) - y(\mathbf{x}^{(2)})}{\Delta} \tag{4.3a}$$

if $\Delta > 0$, or

$$d_j(\mathbf{x}^{(2)}) = \frac{y(\mathbf{x}^{2}) - y(\mathbf{x}^{(3)})}{\Delta} \tag{4.3b}$$

otherwise.

Step 4 is repeated such that a succession of $k+1$ input vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k+1)}$ is produced with two consecutive vectors differing in only one component. Furthermore, any component i of the 'base vector' $\mathbf{x}^*$ is selected at least once to be increased by Δ to estimate one elementary effect for each factor. The successive vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k+1)}$ define a *trajectory* in the parameter space; an example is given in Figure 4.1 for $k = 3$ and $p = 4$.

Each component x_i of the 'base' vector $\mathbf{x}^*$ can only be increased (not decreased) by Δ (see the example below). Thus, any point of the trajectory in the parameter space will have Euclidean distance from the origin (the k-dimensional $\mathbf{0}$ vector) greater than the distance of the 'base' vector. This does not imply that each point has a greater distance than the previous one: a component x_i of $\mathbf{x}^*$ that was increased at a certain stage can be decreased in a successive step (maintaining a greater value than the 'base' one).

The rows of a $\mathbf{B}^*$ matrix are the vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k+1)}$ described above. This matrix is called the *orientation* matrix. It corresponds to one trajectory of k steps in the parameter space, with starting point $\mathbf{x}^{(1)}$. This provides a single elementary effect per factor.

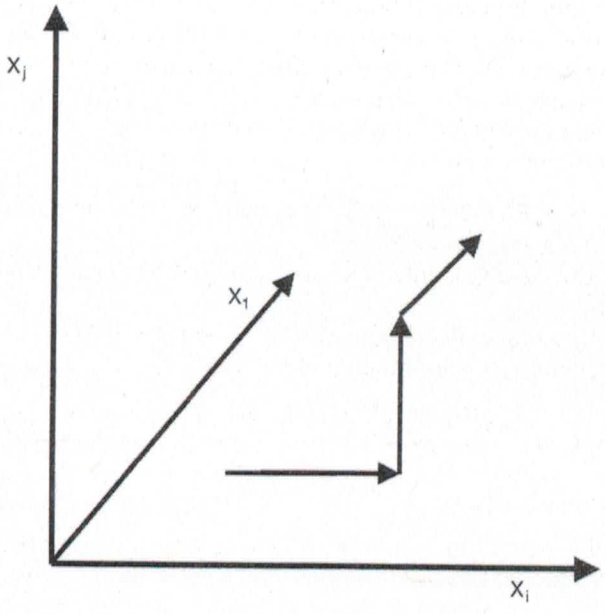

Figure 4.1 An example trajectory in the input factor space for $k = 3$ and $p = 4$.

To build a $\mathbf{B}^*$, the first step is the selection of a $(k+1) \times k$ matrix $\mathbf{B}$ with elements that are 0s and 1s such that for every column there are two rows of $\mathbf{B}$ that differ in only one element. In particular, $\mathbf{B}$ may be chosen to be a strictly lower triangular matrix of 1s. Consider the transposed matrix $\mathbf{B}'$ given by

$$\mathbf{B}' = \mathbf{J}_{k+1,1}\,\mathbf{x}^* + \Delta\mathbf{B}, \tag{4.4}$$

where $\mathbf{J}_{k+1,k}$ is a $(k+1) \times k$ matrix of 1s, and $\mathbf{x}^*$ is a randomly chosen 'base value' of $\mathbf{x}$. This $\mathbf{B}'$ could be used as a design matrix, since it would provide k elementary effects, one for each input factor, with a computational cost of $k+1$ runs. However, the problem is that the k elementary effects $\mathbf{B}'$ produces would not be randomly selected. A randomized version of the design matrix is given by

$$\mathbf{B}^* = (\mathbf{J}_{k+1,1}\mathbf{x}^* + (\Delta/2)[(2\,\mathbf{B} - \mathbf{J}_{k+1,k})\mathbf{D}^* + \mathbf{J}_{k+1,k}])\mathbf{P}^*, \tag{4.5}$$

where $\mathbf{D}^*$ is a k-dimensional diagonal matrix with elements either $+1$ or -1 with equal probability, and $\mathbf{P}^*$ is a $k \times k$ random permutation matrix, in which each column contains one element equal to 1 and all the others equal to 0, and no two columns have 1s in the same position. $\mathbf{B}^*$ provides one elementary effect per factor that is randomly selected.

4.4.1 Example

We suppose that $p = 4$, $k = 2$, and $\Delta = \frac{2}{3}$; that is, we examine two factors that may have values in the set $\{0, \frac{1}{3}, \frac{2}{3}, 1\}$. Then $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix},$$

and the randomly generated $\mathbf{x}^*$ $\mathbf{D}^*$ and $\mathbf{P}^*$ happen to be

$$\mathbf{x}^* = (0, \tfrac{1}{3}), \qquad \mathbf{D}^* = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \mathbf{P}^* = \mathbf{I}.$$

This gives

$$(\Delta/2)[(2\mathbf{B} - \mathbf{J}_{k+1,k})\mathbf{D}^* + \mathbf{J}_{k+1,k}] = \begin{bmatrix} 0 & \Delta \\ \Delta & \Delta \\ \Delta & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & 0 \end{bmatrix},$$

and, from Equation (4.5),

$$\mathbf{B}^* = \begin{bmatrix} 0 & 1 \\ \frac{2}{3} & 1 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix},$$

or

$$\mathbf{x}^{(1)} = (0, 1), \qquad \mathbf{x}^{(2)} = (\tfrac{2}{3}, 1), \qquad \mathbf{x}^{(3)} = (\tfrac{2}{3}, \tfrac{1}{3}).$$

To estimate the mean and variance of the distribution $F_i (i = 1, \ldots, k)$, Morris takes a random sample of r elements; that is, he samples r mutually independent orientation

matrices (corresponding to r different trajectories, each with a different starting point). Since each orientation matrix provides one elementary effect per factor, the r matrices together provide *rk-dimensional* samples, one for each $F_i (i = 1, \ldots, k)$.

This design gives k correlated estimators per trajectory (orientation matrix), whereas the r independent trajectories give r independent estimators. Therefore, the mean μ and standard deviation σ for each of the k factors can be estimated through the classic estimators for an independent random sample.

The main advantage of Morris' design is its relatively low computational cost. The design requires about one model evaluation per computed elementary effect, and a number r of elementary effects is computed for each factor. Thus, the design requires a total number of runs that is a linear function of the number of examined factors, k. The economy of the design is $rk/r(k+1) = k/(k+1)$.

The main disadvantage of the method is that individual interactions among factors can not be estimated. The method can provide an 'overall' measure of the interactions of a factor with the rest of the model, but it does not give specific information about the identity of the interactions.

An analytical test case, proposed by Morris (1991, see Chapter 2), is presented below.

4.4.2 Analytical Test Case

The computational model constructed by Morris (1991, Chapter 2) contains 20 input factors and has the following form:

$$y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{i,j} w_i w_j + \sum_{i<j<l}^{20} \beta_{i,j,l} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{i,j,l,s} w_i w_j w_l w_s,$$

where $w_i = 2 \times (x_i - \frac{1}{2})$ except for $i = 3, 5$, and 7, where $w_i = 2 \times (1.1 x_i/(x_i + 0.1) - \frac{1}{2})$. Coefficients with relatively large values are assigned as

$$\beta_i = +20 \quad (i = 1, \ldots, 10), \qquad \beta_{i,j} = -15 \quad (i, j = 1, \ldots, 6),$$
$$\beta_{i,j,l} = -10 \quad (i, j, l = 1, \ldots, 5), \qquad \beta_{i,j,l,s} = +5 \quad (i, j, l, s = 1, \ldots, 4).$$

The remaining first- and second-order coefficients are independently generated from a normal distribution with zero mean and unit standard deviation; the remaining third- and fourth-order coefficients are set to zero.

Parameters of the experiment were set respectively to $l = 4$, $\Delta = \frac{2}{3}$, and $r = 4$. Using the same representation as in Morris (1991), the values obtained for the sensitivity measures μ and σ are displayed in Figure 4.2. The pattern described in Figure 4.2 almost reproduces the one shown in Figure 4.1 of Morris (1991). Input variables $1, \ldots, 10$, which are supposed to have a significant effect on the output, are well separated from the others. In particular, as shown in Morris (1991), variables 8, 9, and 10 are separated from the others because of their high value of the mean (abscissa). Hence, considering both means and standard deviations together, one can conclude that the first ten factors are important; of these, the first seven have significant effects that involve either interactions or curvatures; the other three are important mainly because of their first-order effect.

The present volume also contains two examples of application of the Morris method to environmental models: a chemical kinetics model of the tropospheric oxidation pathways of dimethylsulfide (DMS), a sulfur-bearing compound of interest in climatic studies (see Chapter 18), and a model of fish population dynamics (Chapter 19).

18
16
14
12
10
σ
8
6
4
2

Figure 4.2 Estimate
the analytical example

A software packa
described in the soft

4.5 COTTER'S

Cotter (1979) propo
about interactions, s
systematic fractional

- one initial run wit
- k runs with each
their low levels;
- k runs with each f
levels;
- one run with all f

Denote the resul
denote the sum of
expected values esti

Then the measures
for the factors.

A major problem
remain undetected.