SKB TECHNICAL REPORT 85-18

Procedures for uncertainty and sensitivity analysis in repository performance assessment

Kurt Pörn, Ove Åkerlund

PROCEDURES FOR UNCERTAINTY AND SENSITIVITY ANALYSIS
IN REPOSITORY PERFORMANCE ASSESSMENT

Kurt Pörn, Ove Åkerlund

Studsvik Energiteknik AB, Nyköping, Sweden
1985-10-10

Procedures for Uncertainty and Sensitivity
Analysis in Repository Performance Assessment

Kurt Pörn and Ove Åkerlund

PN5836A

ABSTRACT

This report was prepared as an account of work
sponsored by the Swedish Nuclear Fuel and Waste
Management Co. The objective of the project was
mainly a literature study of available methods
for the treatment of parameter uncertainty pro-
pagation and sensitivity aspects in complete
models such as those concerning geologic disposal
of radioactive waste. The study, which has run
parallel with the development of a code package
(PROPER) for computer assisted analysis of func-
tion, also aims at the choice of accurate, cost-
affective methods for uncertainty and sensitivity
analysis. Such a choice depends on several factors
like the number of input parameters, the capacity
of the model and the computer resources required
to use the model.

Two basic approaches are addressed in the report.
In one of these the model of interest is directly
simulated by an efficient sampling technique to
generate an output distribution. Applying the
other basic method the model is replaced by an
approximating analytical response surface, which
is then used in the sampling phase or in moment
matching to generate the output distribution. Both
approaches are illustrated by simple examples in
the report.

LIST OF CONTENTS

1.   INTRODUCTION

In the context of this paper, the goal of uncer-
tainty analysis is to measure the uncertainty in
the outcomes of the model of interest as a func-
tion of the the uncertainty in the input vari-
ables, while the goal of sensitivity analysis is
to identify the major contributors to this uncer-
tainty.

In this context the model of interest is con-
cerned with the geologic disposal of radioactive
waste. Thus the outcomes of the model might be
doses for various nuclides as functions of time,
maximum dose, time to maximum dose and/or inte-
grated doses. There are many input parameters,
the values of which are not exactly known. There-
fore, this uncertainty is expressed as proba-
bility distributions. The objective of this paper
is to give some guidelines for the selection of
good, cost-effective methods to be used in the
study of how the input uncertainty is propagated
through the model resulting in an uncertainty of
specific output variables. Several methods are
also closely related to the sensitivity aspects,
i e how much each input variable contributes to
the total variation of the model output.

There is a multitude of methods for the investi-
gation of uncertainty propagation. The selection
of efficient methods depends on several factors
such as number of input variables, the complexity
of the model (linear, non-linear, continuous,
discontinuous) and the computer resources re-
quired to use the model. Some possible approaches
are outlined in Figure 4.1.

Based on extensive studies of the literature in
this area, the authors of this paper believe
that some kind of Monte Carlo technique probably
must be used for the uncertainty analysis. To
enhance the efficiency of the simulation tech-
nique there are many variance reduction methods
available. One of these could be Latin Hypercube
Sampling (LHS). LHS is a type of stratified Monte
Carlo sampling, which in many cases yields a con-
siderable variance reduction compared with simple
random sampling. LHS is also relatively easy to
implement and, thanks to its stochastic nature,
it can be used for direct estimation of the cumu-
lative distribution function (cdf). If the LHS
technique is combined with a regression analysis,
the resulting regression function is useful from
the sensitivity analysis point of view.

2. TWO BASIC APPROACHES

A large system such as PROPER is too complex to permit a simple analysis of uncertainties and sensitivities associated with the model. Presumably one must start with a stepwise analysis, where the impact of uncertainty on individual submodels in the system is determined. There are many different methods for uncertainty and sensitivity analyses. After a comprehensive study of the appropriate literature we want in this report to focus the attention on two basicly different approaches, namely the use of response surface replacement (RS) and the non-use of response surface replacement (NRS).

For this purpose it is convenient to think of the model as a function $Y = g(X_1 \ldots, X_k)$, where the variables $X_1, \ldots, X_k$ are used to represent a variety of phenomena in the model. Uncertainty about these phenomena is assumed to be expressed as probability distributions. The objective of the uncertainty analysis is then to determine the variation or probability distribution of Y that results from the collective variation of the model variables $X_1, \ldots, X_k$. Thus the uncertainty in Y may be displayed as the cumulative distribution function (cdf) of Y as a measure of the variability range of Y. In addition to the cdf one may be interested in quantities like the mean or median of Y, the variance of Y, and the lower and upper 5 % quantiles of Y.

Another natural question is which variable or variables contribute most to the uncertainty in Y? Sensitivity analysis is performed in order to identify the important contributors to uncertainty in Y.

The PROPER system under development is expected to be characterized by

-        many input variables

-        correlations exist within couples or larger groups of input variables

-        the model is time consuming to run on a computer

-        model outputs are nonlinear and time-dependent functions of the input variables.

This paper shortly describes two basic approaches to uncertainty and sensitivity analysis, and discusses their advantages and limitations:

-        The model $Y = g(\underline{X})$ is replaced with an approximating response surface

-        Direct estimate of the cdf for Y without response surface replacement

2.1   Response surface replacement (RS)

The variation of the computed model output as a function of the input perturbations is called the response surface. The equation representing this surface is called the response surface equation. In practice, this true response surface $Y = g(\underline{X})$ very often must be replaced with a simpler and approximating surface. In the following we use the term "response surface" to denote the approximating function while the original function is simply called the model. A common approach is a second-order polynomial

$$\tilde{Y} = b_0 + \sum_{i=1}^{k} b_i Z_i + \sum_{i=1}^{k} \sum_{j=1}^{k} b_{ij} Z_i Z_j \qquad (2.1)$$

where $\tilde{Y}$ is the response surface estimate and Z_i denotes the i'th input variable X_i or its dimen-

sionless value $(X_i-\mu_i)/\sigma_i$ or some other transformation of X_i. The coefficients b_o, b_i and b_{ij} are estimated by least-squares fitting

$$\sum_{i=1}^{k} (Y_i - \tilde{Y}_i)^2 = \min \qquad (2.2)$$

or by generalized interpolation schemes. The normalization quantities μ_i and σ_i denote the nominal (best-estimate) value and standard deviation respectively of the i'th input variable.

The equation (2.1) is a multivariable Taylor series expansion of second order, which in principle allows a second-order error propagation analysis. For nonlinear models, this kind of analysis is required at the very least.

When the coefficients of Eq (2.1), which are proportional to the partial derivatives of a Taylor series expansion about the nominal values, have been evaluated, one has an approximating response surface equation, which can be used to produce a cdf for Y and various central characteristics of that distribution. The coefficients themselves can be used for ranking of the input variables. Thus if we succeed to find a satisfactory response surface, most of the information about the original model is concentrated into that surface. All subsequent uncertainty and sensitivity analysis is therefore based on the response surface. For instance, mean, variance and even higher moments for Y are easy to derive from the polynomial form (2.1), based on the corresponding moments of the input variables. A practical way to compute the entire distribution of Y is described below. The most difficult step in

the response surface generation is to select
values of the input variables in such a way that
the necessary information is obtained with as few
computer runs as is practical. Different statis-
tical designs of experiments are available: com-
posite design with sequential implementation,
Resolution IV fractional factorial designs supp-
lemented with star points (those obtained when a
single parameter is perturbed while the remainder
are fixed at their nominal values). References to
these designs are given by Cox (1977), who also de-
scribes an application of the Resolution IV design
to the determination of peak cladding temperature
with 7 input variables and 57 computed responses.
Resulting responses were fed into a linear step-
wise regression analysis subroutine resulting in
a satisfactory response surface equation with an
R-square statistic of 0.89. The subsequent deter-
mination of the cdf was performed either by a
crude Monte Carlo Sampling from the response sur-
face equation or by a moment matching technique.
In case of the latter, the moments of the distri-
bution of Y were computed by the use of the SOERP
code (Cox & Miller 1975) and matched to those of
a specific Pearson distribution family.

Another, basicly different experimental design
is the knot-point selection scheme by Vaurio &
Mueller (1977) and Vaurio (1981). Starting from
a second-degree response surface of the type
(2.1), the unknown coefficients are determined
by a multivariate generalization of the Lagrange
interpolation scheme. This means that a set of
$1+2k+k(k-1)/2$ knot-points are selected at which
the approximation $\tilde{Y}$ is made equal to the actual

values of Y. The set of knot-points consists of
a central reference point, two extreme points
for each input variable and $k(k-1)/2$ interjacent
points for the description of the interaction
effects. There are different variants of the
knot-point selection schemes. A single-quadrant
response surface (SQ) is a single polynomial
covering the entire parameter space. A multi-
quadrant surface (MQ) is obtained when region-
wise (one per quadrant) surfaces are combined.
This approach leads to a better approximation at
the price of a nearly fourfold number of knot-
points. Even random selection of the knot-points
with least-squares fitting is possible in the
PROSA-code developed by Vaurio (1981). This fea-
ture allows both second-degree response surfaces
(RF2) and third-degree response surfaces (RF3)
to be fitted to the data. Maximum positive and
negative errors as well as the mean-square error
of the fitted response surfaces can be calculated
in the knot-points.

In Vaurio (1981) it is stated that the simplest
response surfaces, SQ and RF2, are normally used
when k is large, while the more refined surfaces,
MQ and RF3 are used with small k, possibly after
the less important variables have been eliminat-
ed. Concerning the accuracy, one can expect (and
has been verified) RF2 to be more accurate in the
central part of the distribution while SQ may
be more accurate in the tail area, as a conse-
quence of the knot-point selection schemes. MQ
and RF3 should be superior to the others, since
they are more flexible in predicting the true
functionality of the model.

The Latin Hypercube Sampling (LHS) technique, which is described in the next section, can also be used as a method to select the values of input variables for the determination of the response surface. One advantageous feature of LHS with respect to the response surface approach is the forced spread of sampled values inherent in the method. However, as Iman & Helton (1985) assert, there is little incentive to fit a response surface based on the LHS-technique because this sampling method is efficient enough to directly obtain the desired estimates.

In general, the accuracy of the response surface can be improved by making functional transformations of the consequence and/or input variables. Thereby one is able to use a simple polynomial form for the response surface rather than a more complicated function in the original variables. However, the search for suitable and efficient transformation is by no means any simple task. It requires a relatively good knowledge of the dependency between the model output and at least the most important input variables. It is quite clear that such a knowledge must be achieved in an iterative way, a procedure that scarcely can be automated.

As a summary of the above, the following advantages and limitations with the response surface methodology may be mentioned:

Advantages:

The response surface itself is very useful in the sensitivity analysis involving determination of the response $\tilde{Y}$ to changes in the input vari-

ables X. There are several ways to measure this
influcence of individual variables, as e.g. stan-
dardized regression coefficients, partial corre-
lation coefficients, normalized partial deriva-
tives, and contribution to variance. A discussion
and comparison of these measures can be found in
Iman & Helton (1985). The purpose of these dif-
ferent measures is to facilitate a ranking by
importance of the input variables.

Limitations:

The necessary number of points in the input
parameter space is growing fast with increasing
dimensionality k.

It is difficult to choose, from a set of many
different experimental designs, one input value
selection scheme which best suits the features
of the appropriate model.

It is also a difficult step to look for effici-
ent functional transformations, which very often
will be necessary for the approximation of com-
plex models with simple polynomial forms.

The uncertainty distribution of $\tilde{Y}$ cannot be
estimated directly from the responses derived
from the input values in cases where these are
selected systematically.

2.2   No response surface replacement (NRS)

Trying to estimate the distribution of Y without
using a response surface the aim is to describe
the variability in Y with as few computations of
Y as possible. As in generating a response sur-
face the problem is to select values for the in-

put parameters. In 2.1 deterministic models to
select input values were described and some ran-
dom sampling models were mentioned. Here a method
called Latin Hypercube Sampling (LHS) will be ex-
plained. This method can be used both with and
without the generation of a response surface but
here we will only examine the method of point
selection.

Given a function $Y = g(\underline{X})$ and a known distribu-
tion for each one of the input variables
X_i (i = 1,...,k) we want to assign values to
each X_i, N number of times, to be able to make N
calculations of Y. In LHS the range of each X_i
is then divided into N equal probability inter-
vals, where the range is determined according to
the distribution of each X_i. The following
example will illustrate the LHS method.

Suppose k = 3 so $Y = g(X_1, X_2, X_3)$ and N = 4,
the number of intervals

X_1: |————+—+—+————|

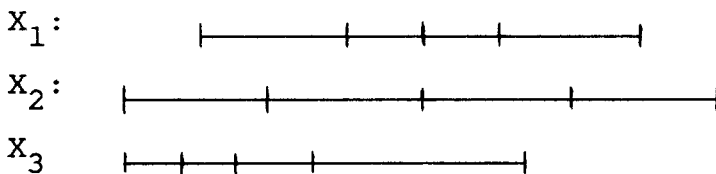X_2: |————+————+————+————|

X_3 |—+—+——+————|

Figure 2.2.1
The range for the three input variables. Each
range is divided into four intervals with equal
probability.

As can be seen from Figure 2.2.1 the intervals
can have different lengths depending on the
distribution of each X_i. The size of the inter-
vals are selected to give the same probability
to each one of them. For example a uniform dis-
tribution will have equally long intervals and

a normal distribution will have shorter intervals
in the middle of the range. Now the LHS method
selects one interval randomly for each variable
and then the actual value is randomly selected
within each interval. In the next step the inter-
vals already chosen can not be selected so this
is a random sampling without replacement. In
this way the method assures a selection also of
extreme input values even for small samples. The
method is illustrated in Figure 2.2.2 . The
method is described in detail by Mc Kay, Conover
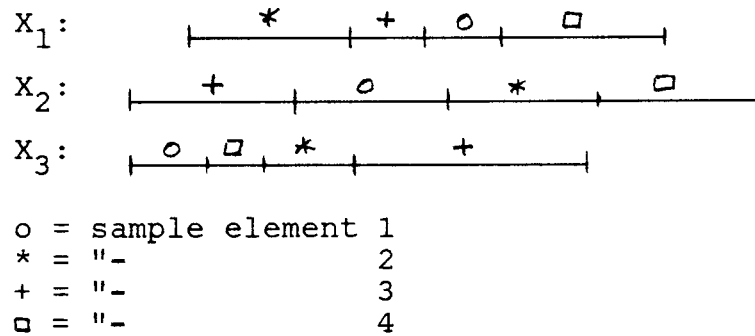and Beckman (1979).

X_1:

X_2:

X_3:

o = sample element 1
* = "-   2
+ = "-   3
□ = "-   4

Figure 2.2.2

One possible outcome of a LHS for three variables
and four intervals.

Since our aim is to estimate the distribution of
Y = g($\underline{X}$) we do need a sample size which gives a
desired confidence in the estimated distribution
parameters. In practice, the choice of the sample
size is dependent on both the number of variables
k and how complicated the function g($\underline{X}$) is. In
chapter 5, the precision of the estimates for
the 5 % upper and lower percentiles for a speci-
fied function g($\underline{X}$) is examined when using LHS
with 20 intervals. Also a comparison between the
precision in the LHS and a crude random sampling
is being done.

The LHS method has been used in several analysis and is reputed to be a very good and useful tool as a sampling technique. In McKay, Conover, Beckman (1979) there is a comparison of three sampling methods.

1   Crude random sampling
2   Stratified sampling
3   Latin hypercube sampling

Some conclusions in that report are

-        LHS gives an unbiased estimator of the mean and the cumulative distribution function (cdf) of Y

-        the estimates based on LHS appear to be more precise in general than the other two types of estimates.

Also Iman & Helton (1985) prefer the LHS because

-        it is more efficient than simple random sampling (the degree of superiority varies from case to case)

-        having a probabilistic basis, it can provide direct estimates for the mean, variance and cdf

-        it allows the input of correlated input variables to the model.

3.　　SOME GENERAL CONCEPTS RELATED TO MONTE CARLO TECHNIQUES

In principle a Monte Carlo procedure can be interpreted as a method for evaluating an integral. An intuitive justification for this interpretation can be given by the way in which the Monte Carlo method works. The simulation is exercised numerous times and conclusions about the process are drawn by averaging the individual outcomes. But averaging is a means for estimating particular types of integrals known as expectations, McGrath & Irving (1975).

Let us denote the model by

$$Y = g(\underline{X}), \tag{3.1}$$

where the vector $\underline{X}$ represents the k input parameters, the uncertainty of which affects the outcome Y. If $f(\underline{X})$ denotes the probability density function (pdf) of $\underline{X}$ then the objective of the Monte Carlo simulation may be to estimate the integral

$$I = E[g(\underline{X})] = \int g(\underline{X})f(\underline{X})d\underline{X} \tag{3.2}$$

A crude Monte Carlo method means that a random sample $\underline{X}_1, \ldots, \underline{X}_N$ is selected from $f(\underline{X})$ and I is estimated by the sample mean

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} g(\underline{X}_i) \tag{3.3}$$

According to the law of large numbers, the estimate $\hat{I}$ converges to I in most cases. It is also true that $\hat{I}$ is a random variable and that the expected value satisfies

$$E [\hat{I}] = I, \tag{3.4}$$

i.e. $\hat{I}$ is an <u>unbiased estimator</u> of I.

The <u>population variance</u>

$$\sigma^2 = E [g(\underline{X}) - I]^2 \tag{3.5}$$

is commonly estimated by the <u>sample variance</u>

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} [g(\underline{X}_i) - \hat{I}]^2 = \frac{N}{N-1} \{\frac{1}{N} \sum_{i=1}^{N} g^2(\underline{X}_i) - \hat{I}^2\} \tag{3.6}$$

A basic measure for the effectiveness of the estimator is

$$\sigma^2_{\hat{I}} = E [\hat{I} - I]^2 \tag{3.7}$$

For stochastically independent histories it is easy to see that

$$\sigma^2_{\hat{I}} = \frac{\sigma^2}{N}$$

and since σ^2 is estimated by S^2, an unbiased estimate of the efficiency measure is obtained by

$$s^2 = \frac{S^2}{N} = \frac{1}{N-1} \{\frac{1}{N} \sum_{i=1}^{N} g^2(\underline{X}_i) - \hat{I}^2\} \tag{3.9}$$

The question of efficiency can also be considered in terms of uncertainty intervals. From basic statistics it is known that the estimate $\hat{I}$ with high probability will fall between $I - z\sigma/\sqrt{N}$ and $I + z\sigma/\sqrt{N}$, where z is some constant. Thus for fixed z, the precision of the estimate is related to the number of histories N, and the variance of $g(\underline{X})$.

The disadvantage of increasing the accuracy of
the estimator by increasing the number of histor-
ies is quite obvious. The other possibility is
to reduce the variance (σ) associated with each
observation.

Let us assume there are two simulation methods
for estimating the same parameter I. Let the
variances per history be σ_1^2 and σ_2^2 respectively
and the computing time per history t_1 and t_2.
Then the relative efficiency of the two simu-
lation methods is given by the ratio of the com-
puting times required to achieve the same pre-
cision:

$$\varepsilon = \frac{t_1 \; \sigma_1^2}{t_2 \sigma_2^2} \qquad (3.10)$$

Because in most applications a variance reduction
method is being compared to crude sampling, t_1
and $\sigma_1^{\;2}$ would be the quantities obtained when
crude Monte Carlo is used. Further, because
σ_1^2 and σ_2^2 are rarely known, the relative ef-
ficiency must be estimated by

$$\hat{\varepsilon} = \frac{t_1 \; s_1^2}{t_2 \; s_2^2} \qquad (3.11)$$

In case of random sampling it is well known that
the estimators (3.6 and 3.9) are unbiased esti-
mators of the variance of $g(\underline{X})$ and $\hat{I}$ respect-
ively. However, the use of many variance re-
duction techniques, such as LHS and other strati-
fied sampling methods, Russian Roulette and
splitting, correlated sampling or antithetic
variates, will not produce histories that are
stochastically independent. Then in many cases

it is possible to group the histories into
batches such that the batches are independent
and equivalent, while the samples within a batch
may be correlated with each other. If the samples
$g(\underline{X}_1), \ldots g(\underline{X}_{N_i})$ are contained in batch i, then
the average in that batch is

$$\hat{I}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} g(\underline{X}_j) \qquad (3.12)$$

A final estimate of I is

$$\hat{I} = \frac{1}{N_B} \sum_{i=1}^{N_B} \hat{I}_i \qquad (3.13)$$

where N_B is the number of batches. An estimate
of the corresponding σ^2 can now be obtained from
the "batch sample" variance

$$S^2 = \frac{N_B}{N_B - 1} \left[\frac{1}{N_B} \sum_{i=1}^{N_B} \hat{I}_i^2 - \hat{I}^2 \right] \qquad (3.14)$$

4. A COMPARISON BETWEEN RS AND NRS APPLIED TO SIMPLE EXAMPLES

The problem to estimate the distribution of a complex function $Y = g(\underline{X})$, with known distributions of the independent parameters $\underline{X}$, is going to be further discussed in this chapter. The two basic approaches to the problem, Response Surface Replacement (RS) and No Response Surface Replacement (NRS) are described in chapter 2. To further highlight the differences between methods we will apply them to four simple models.

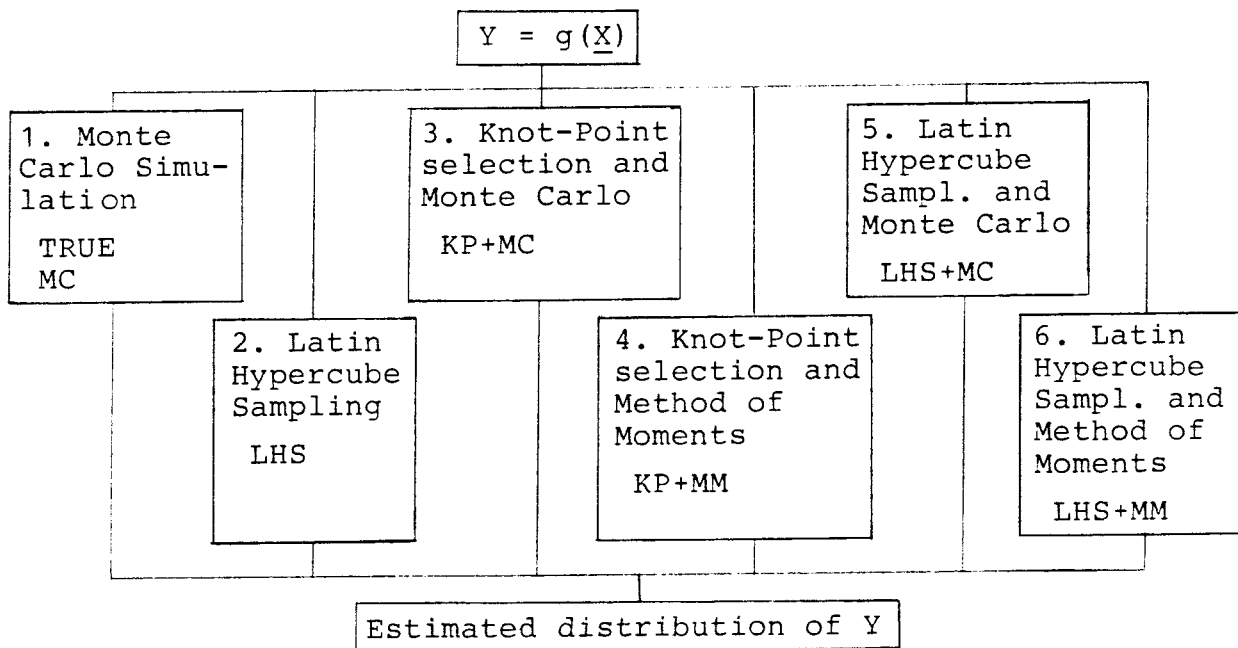Six different methods will be tried to assess the distribution of Y, see Figure 4.1.

Figure 4.1

Six different ways to assess the distribution of Y.

Below the different methods are described in detail.

1)   TRUE or MC, which is a Crude Monte Carlo simulation on the original expression $Y = g(\underline{X})$. This means that random sampling has been made for each parameter X_i according to its distribution and corresponding values of Y have been calculated. For TRUE this is done 20 000 times to get the "true" distribution of Y and for MC 2 000 samples are used.

2)   LHS, Latin Hypercube Sampling where the variation of each input-parameter is split into 50 intervals. This means that there are 50 calculations of Y to estimate its distribution characteristics.

3)   KP + MC, Knot-Point selection and Monte Carlo simulation on the response surface $\tilde{Y} = k(\underline{X})$. Here 28 points (knot-points) are used in the interpolation scheme to calculate the coefficients in the response surface. Then 20 000 Monte Carlo simulations are made on the response surface.

4)   KP + MM, Knot-Point selection and interpolation to get the response surface $\tilde{Y} = k(\underline{X})$, which is the same surface as for the KP + MC method. With the "Method of Moments" the four first moments of $\tilde{Y}$ are determined. In this case we have not calculated the distribution of $\tilde{Y}$, but using the calculated moments we could have adjusted a distribution of the Pearson family to $\tilde{Y}$.

5)   LHS + MC, Latin Hypercube Sampling and Monte Carlo simulation on the response surface $\tilde{Y} = h(\underline{X})$. In this case 28 "LHS-points" are used in a least square fitting to estimate the coefficients in the response surface $h(\underline{X})$. The Monte Carlo simulation has been made 20 000 times on $\tilde{Y}$.

6)   LHS + MM, the response surface $\tilde{Y} = h(\underline{X})$ is assessed with Latin Hypercube Sampling and least square fitting. By the "Method of Moments" the four first moments of $\tilde{Y}$ are calculated and in the same way as for method 4, one can also adjust a Pearson family distribution to these moments.

Methods 3 - 6 are included to illustrate that
one must be cautious when using a response sur-
face to estimate the distribution of Y. The use-
fulness of response surfaces in general is
treated in chapter 2.

In the Response Surface Replacement methods 3
and 5 a Monte Carlo (MC) simulation is made with
$\tilde{Y}$ to estimate the distribution of Y. An interest-
ing alternative to the MC-simulation is to use
the Method of Moments, as in methods 4 and 6. In
case of a second-order error propagation approach
this method requires that

1. $\tilde{Y}$ is a function of at most second order
        with known coefficents and

2.   The first 8 moments shall be given for
        the independent variables.

With these two requirements fulfilled the first
four moments of $\tilde{Y}$ can be calculated. There is a
code SOERP to make these calculations, see COX
(1977). With the four first moments given for $\tilde{Y}$
one can find a distribution within the Pearson-
family having the same moments. This adjustment
from moments to coefficients of a Pearson-type
distribution is described by Cox and Miller
(1976).

The four examples mentioned earlier will now be
penetrated and in chapter 6 some conclusions
will be given.

For each example the results are shown in a
table where the following sample characteristics
for Y will be examined:

22

Method	Six different methods
Sample size	Sample size where 1 000 = 1'
Mean	Estimated mean value
Std dev	Estimated standard deviation
5 % perc	Estimated 5 % percentile
95 % perc	Estimated 95 % percentile
Skewness	The ratio of the squares of the third and the second central moments, μ_3^2/μ_2^3. (For the normal distribution skewness = 0.)
Kurtosis	The ratio of the fourth and the squared second central moments, μ_4/μ_2^2. (For the normal distribution kurtosis = 3.0.)
CPU-time	Computer execution time. It is to be noted that the execution times are not directly comparable because different codes with varying preprocessing have been used.

In some of the examples not all of these characterictics are calculated.

4.1 Example 1

In this example Y has the form

$$Y = X_1^2 + \frac{(X_2 - X_3)^2}{2} + \frac{(X_4 + X_5 + X_6)^2}{3}$$

where all X_i are normally distributed with mean 0 and variance 1.

The functional form of Y has no physical meaning. Instead it is selected in order to have an exact distribution, namely chi-square with 3 degrees of freedom. Then we know that

$$E(Y) = 3 \qquad \text{mean}$$
$$D(Y) = 2.45 \qquad \text{standard deviation}$$
$$Y_{.05} = 0.35 \qquad 5\ \%\ \text{percentile}$$
$$Y_{.95} = 7.81 \qquad 95\ \%\ \text{percentile}$$
$$\text{Skewness} = 2.67$$
$$\text{Kurtosis} = 7.0$$

In Table 4.1 the corresponding values are calculated according to the six different methods described earlier.

In Figure 4.2 the cumulative distribution function is plotted for four of the methods.

Table 4.1

Distribution characteristics for the six methods in example 1.

Method	Sample size	Mean	Std dev	5% perc	95% perc	Skew-ness	Kurto-sis	CPU time (S)
TRUE	20'	3.0	2.5	.35	7.8	3.1	8.4	21.
MC	2'	3.0	2.4	.37	7.6	3.2	8.9	2.4
LHS	50	3.1	2.5	.22	7.2	-	-	7.1
KP+MC	28+20'	3.0	2.4	.55	7.9	2.7	6.8	24.
KP+MM	28	2.8	2.4	-	-	2.7	7.0	2.3
LHS+MC	28+20'	3.0	2.5	.22	7.9	2.9	8.1	30.
LHS+MM	28	3.0	2.5	-	-	2.5	6.9	8.3

A comparison of the mean and standard deviation shows there is no difference between the methods nor to the exact chi-square distribution. Coming further out in the tails of the distribution the agreement between the methods is not so good for the 5% percentile but nevertheless they seem to concentrate around the chi-square value. The 95 % percentile is better concerning both comparisons except for the LHS-method.

Skewness and kurtosis seem to indicate that for
the surface methods the distribution is less skew
and not so peaked as the distribution obtained by
direct Monte Carlo simulation. On the other hand,
in the former case those is a better agreement
with the exact values. In particular, the agree-
ment is complete for the knot-point method com-
bined with moment matching, which is quite natural
when the resulting response surface is considered.

At this stage it is interesting to investigate
the response surfaces estimated by the KP
(interpolation) and LHS (regression) methods
respectively. Notice that in both cases there
are 28 points behind the estimation of $\tilde{Y}$ and
that $\tilde{Y}$ is a function of at most second order.

KP: $\tilde{Y} = k(\underline{X}) = X_1^2 + \frac{1}{2} (X_2^2 + X_3^2) + \frac{1}{3} (X_4^2 + X_5^2 + X_6^2) -$

$- X_2 X_3 + \frac{2}{3} (X_4 X_5 + X_4 X_6 + X_5 X_6)$

LHS: $\tilde{Y} = h(\underline{X}) = -0.02 - 0.13 X_4 - 0.02 X_6 +$

$+ 0.98 X_1^2 + 0.58 X_2^2 + 0.53 X_3^2 + 0.39 X_4^2 +$

$+ 0.51 X_6^2 - 1.12 X_2 X_3 + 0.42 X_4 X_5 +$

$+ 0.50 X_4 X_6 + 0.80 X_5 X_6$

From the equations we see that the Knot-Point
method gives back the original expression for Y
which it should do. The LHS-method shows a dif-
ference to the original Y, specially the X_5
squared term is missing. This is due to the cut
off value, selected in the regression process,
to stop the inclusion of further variables. In
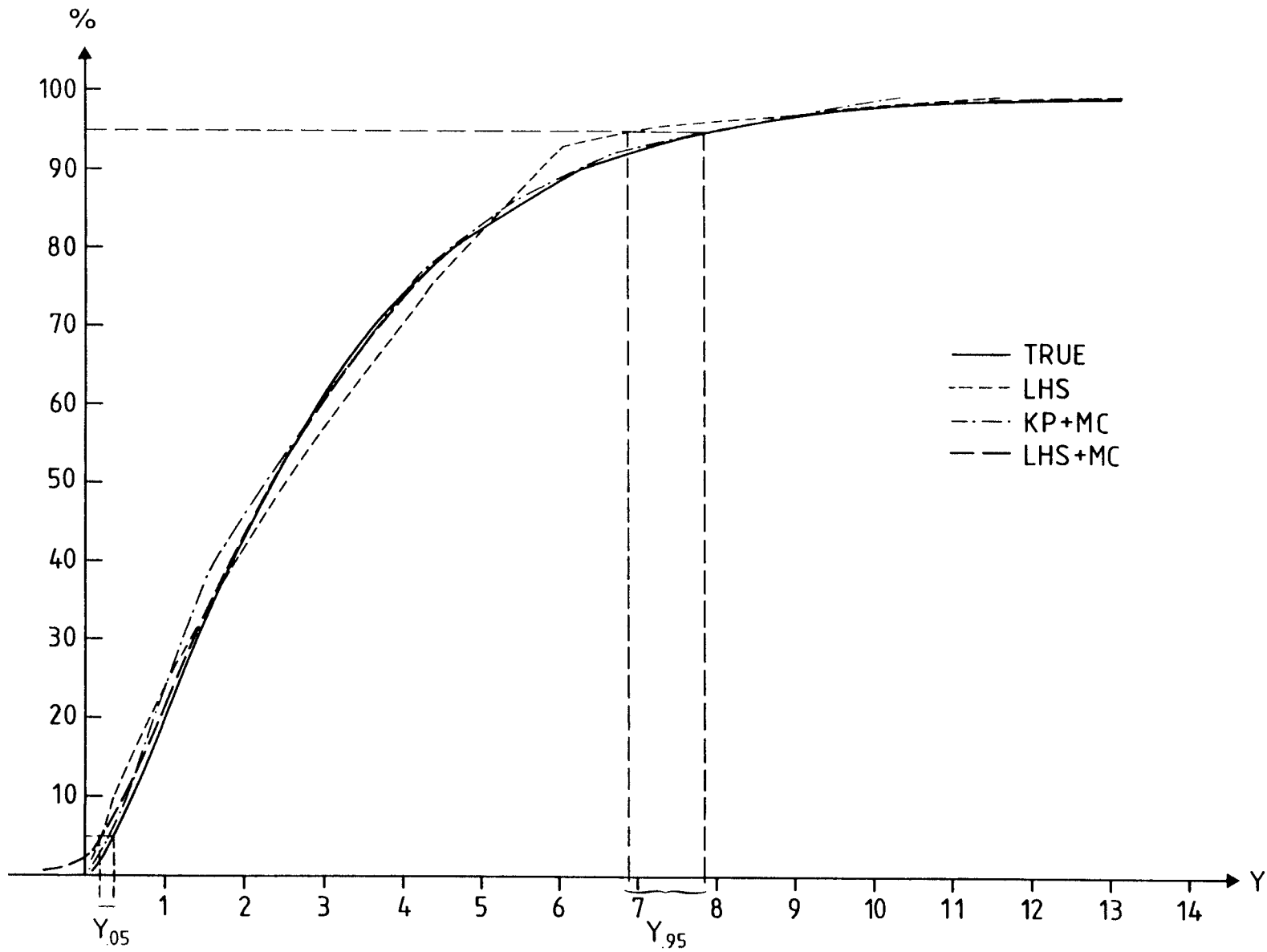this case the X_5 squared term did not contribute

Figure 4.2

Distribution plots for four of the methods in
example 1

enough to the explanation of the variation of Y.
Still the coefficients are of the right magni-
tude and of the right sign and the Monte Carlo
simulation on h($\underline{X}$) gives a very good correspon-
dence to the TRUE-distribution.

Of course, in this case with such a simple form
of Y, there is no need for a response surface
but it is illustrative to see the results of the
methods.

From Figure 4.2 we can make the same conclusions
as from table 4.1 that there are small differen-
ces between the methods. In the figure two "un-
certainty"-intervals are drawn, one for each of
the 5% and 95% percentile. The 5% value has a
narrow interval even though the relative devia-
tion is greater than for the 95% percentile. The
95% percentile has a wider interval due to the
low "LHS-value". One explanation to this, of
course, is that behind the LHS-curve are only
50 points which makes the curve more uncertain
than the others with 20 000 points. By mere
chance the curve of LHS + MM seems to agree
better although it is based on only 28 obser-
vations.

Remembering the separate expressions for g($\underline{X}$)
and its response h($\underline{X}$) they are still just about
the same in Figure 4.2

4.2   Example 2

The function Y has the form

$$Y = X_1 + 2X_2 + X_3^2 + 2X_4^2 + X_5^3 + X_6^4$$

where all X_i are uniformly distributed over the
interval (0,2).

Here the form of Y has been chosen of order 4 to see how well a response surface of second order without variable transformation can "replace" it.

The exact distribution of Y is not known so the Crude Monte Carlo simulation 20 000 times will be the best knowledge we have, the TRUE method in Table 4.2.

As for example 1 a Table 4.2 and a Figure 4.3 show the characteristics and the cumulative distribution respectively for the methods.

Table 4.2

Distribution characteristics for four of the methods in example 2.

Method	Sample size	Mean	Std dev	5% perc	95% perc	Skew- ness	Kurto- sis	CPU time (s)
TRUE	20'	12.2	5.6	4.5	22.9	.5	3.3	8.6
MC	2'	12.2	5.6	4.5	22.9	.5	3.1	1.1
LHS	50	12.2	5.9	4.3	23.4	-	-	5.9
KP+MC	28+20'	12.3	6.3	3.5	24.1	.3	2.9	8.7
LHS+MC	20+20'	11.4	4.4	5.2	19.6	.3	2.9	15.

In this example there is a good resemblance between the methods. The difference in $Y_{.95}$ between TRUE and LHS is 0.5 which is about 1/4 of the statistical uncertainty in the estimation of $Y_{.95}$ itself (see chapter 5). This means that the uncertainty due to method is smaller than the uncertainty due to sampling in this case. Of course the sample size in the LHS method, here 50, could be increased to achieve a better precision.

The KP + MC method is practically identical with TRUE which is explained below when investigating the response surfaces $\tilde{Y}$.

Here the least good agreement with TRUE is the LHS + MC method. As can be seen from the 5% and 95% percentile values the LHS+MC distribution is more narrow and the mean is lower than for TRUE, see further in figure 4.3

Some of the similarities and disagreements with TRUE just mentioned can be explained by looking at the response surfaces $\tilde{Y}$.

KP: $\tilde{Y} = k(\underline{X}) = 1.X_1 + 2.X_2 - 2.X_5 - 6.X_6 + 1.X_3^2 +$

$$+ 2.X_4^2 + 3.X_5^2 + 7.X_6^2$$

LHS: $\tilde{Y} = h(\underline{X}) = 4.0 + 3.4\ X_5X_6 + 1.2\ X_2\ X_4 +$

$$+ 1.9\ X_1X_6 + 1.8\ X_4^2 - 1.4\ X_1X_4$$

For the KP method the generalized interpolation scheme yields an exact resemblance between the original Y and $\tilde{Y}$ in the 28 knot-points. Notice that Y is of order four but in these 28 points it can for a certain set of coefficients, determined by the interpolation scheme, be exchanged by a function $\tilde{Y}$ of order two. Then the overall resemblance is dependent on the behaviour in the area between the knot-points. In this example with well restricted variable ranges the resemblance seems to be good enough, at least regarding the lower order moments.
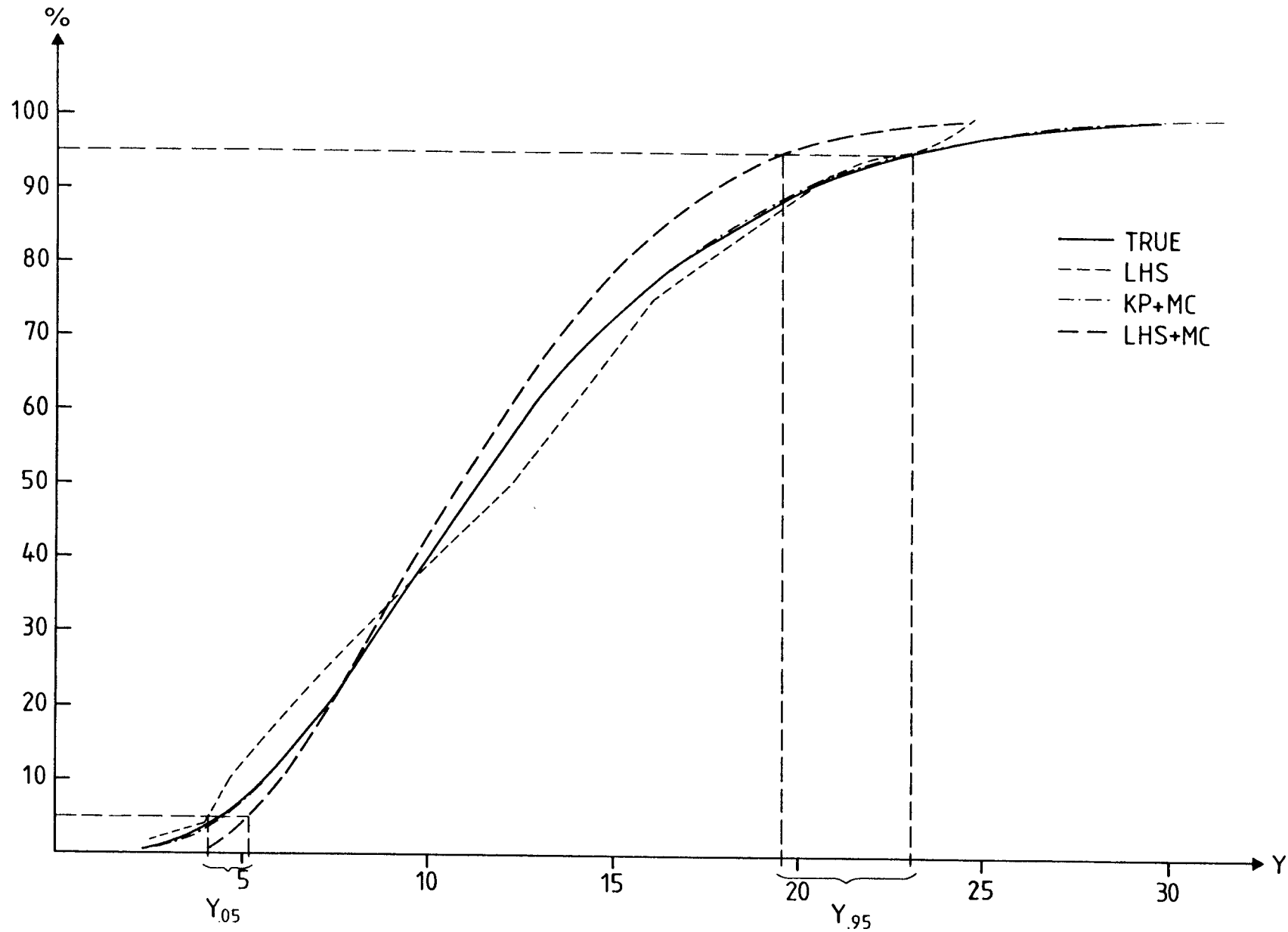
Figur 4.3

Distribution plots for four of the methods in example 2

The determination of a response surface with
regression in the LHS + MC method was not so
successful. Obviously, the sample size of 20
points used in the LHS was not enough to ad-
equately describe the variation in Y. The degree
of explanation in the regression was only 93 %
(R^2) compared to 99 % in example 1.

In Figure 4.3 there are three things to point
out.

1.   The distribution plot of LHS is not as
smooth as the others, which depends on
both the small number of calculations
of Y and that all the necessary data are
not available in the computer printout.

2.   The earlier mentioned systematic dis-
agreement for the LHS + MC method is
illustrated very well.

3.   Also in this example the greatest error
in $Y_{.\alpha}$, due to method, is about 20 %.

The uncertainty in estimating $Y_{.\alpha}$ is not only
due to the method but also to sampling technique
i.e how to select values for the input parameters
to get the best possible knowledge of the
variation of Y. The precision of $Y_{.\alpha}$, due to
sampling will be further discussed in the next
chapter.

4.3   Example 3

The function Y has the form

$$Y = \frac{1 + 3X_1 + 2X_2^2 + X_3^3}{80 + 9X_4 + 4X_5^2 + 2X_6^3}$$

where all X_i have a normal distribution with mean
0 and variance 1.

In this example the aim is to make Y more complex by introducing a ratio, but still keep the response surface $\tilde{Y}$ of at most second order. The results are shown in Table 4.3.

Table 4.3

Distribution characteristics for the six methods in Example 3.

Method	Sample size	Mean	Std dev	5% perc	95% perc	Skew- ness	Kurto- sis	CPU time (s)
TRUE	20'	.04	.07	-.06	.15	.02	49.	22.
MC	2'	.04	.07	-.06	.15	.64	12.	2.5
LHS	20	.04	.05	-.08	.12	-	-	7.
KP+MC	28+20'	.04	.07	-.07	.16	.10	3.8	24.
KP+MM	28	.04	.07	-	-	.14	3.9	2.3
LHS+MC	28+20'	.04	.05	-0.4	.12	.07	5.1	28.
LHS+MM	28	.04	.05	-	-	.12	5.3	7.3

In Table 4.3 it is interesting to notice how well the LHS method estimates the characteristics with only 20 samples, despite the relatively complex expression for Y. The kurtosis is much greater for TRUE than the other methods, which indicates that information is lost in the tails of the distribution of Y.

In this case the response surfaces are more interesting to investigate since Y is a ratio and $\tilde{Y}$ is only a polynominal of second order.

KP: $\tilde{Y} = k(\underline{X}) = 10^{-3}$ [13. + 38.X_1 + 48.X_3 - 1.X_4 -

$$- 1.X_6 + 25.X_2^2 - 1.X_5^2 - 3.X_1X_4 - 3.X_1X_5 -$$

$$- 3.X_1X_6 - 5.X_2X_4 - 4.X_2X_5 - 4.X_2X_6 -$$

$$- 4.X_3X_4 - 4.X_3X_5 - 4.X_3X_6]$$

LHS: $\tilde{Y} = h(\underline{X}) = 10^{-3}$ [16. + 27.X_1 + 22.X_3 +

$$+ \ 22.X_2^2 \ - \ 9.X_3^2 \ + \ 6.X_4^2 \ - \ 15.X_3X_6]$$

For KP only terms with coefficients greater than or equal to 0.001 are shown in $\tilde{Y} = k(\underline{X})$. In the LHS case only the terms shown in $h(\underline{X})$ were selected by stepwise regression rules for including variables in the response surface. The degree of explanation is 96 % so the resemblance should be quite good. Nevertheless there is a considerable disagreement with respect to higher moments, which is shown, e.g. by the decrease in kurtosis.

Doing a comparison of the response surfaces for KP and LHS it is noticeable that the constant term is about the same and that the three greatest coefficients belong to the same variables.

4.4 Example 4

This is practically the same model as in example 3 except that the distribution for the input variables is different. The Y model has the form

$$Y = \frac{5 + 3X_1 + 2X_2^2 + X_3^3}{1 + 9X_4 + 4X_5^2 + 2X_6^3}$$

where $X_1 \sim$ log N (-.69,1) log normal

$X_2 \sim$ U (2,10) uniform

$X_3 \sim$ N (0,1) normal

$X_4 \sim$ U (50,100)

$X_5 \sim$ N (5,3)

$X_6 \sim$ N (0,1)

One reason for applying these three distributions was to test this facility in the codes. The three distributions normal, lognormal and uniform are

the only distributions that are available in all
of the four codes used. The results for this Y
model are shown in Table 4.4.

Table 4.4
Distribution characteristics for the six methods
in Example 4.

Method	Sample size	Mean	Std dev	5% perc	95% perc	Skew- ness	Kurto- sis	CPU time (s)
TRUE	20'	.13	.10	.02	.32	1.4	4.4	20.
MC	2'	.12	.09	.02	.30	1.1	4.0	2.4
LHS	20	.12	.08	.02	.27	-	-	8.3
KP+MC	28+20'	.13	.09	.03	.33	1.4	4.2	20.
KP+MM	28	.12	.08	-	-	.6	2.9	2.3
LHS+MC	28+20'	.11	.11	-.02	.32	.8	3.2	28.
LHS+MM	28	.00	.08	-	-	.1	2.0	8.6

Once again it is shown how well the different
methods acts in the middle of the distribution
and in this case they are fairly well even in the
tails. The skewness and kurtosis indicate that
the distributions get less skewed and flatter
for the response surface methods.

In this case with different distributions for the
input parameters it is more difficult to interpret
the coefficients of each term. To make the inter-
pretation easier each variable X_i has been re-
placed by $Z_i = (X_i - \mu_i)$ where $\mu_i = E(X_i)$.

KP: $\tilde{Y} = k(\underline{Z}) = 10^{-3} [100. + 4.Z_1 + 31.Z_2 + 5.Z_3 -$

$$- 1.Z_4 - 4.Z_5 - 1.Z_6 + 3.Z_2^2 - 2.Z_2Z_5]$$

LHS: $\tilde{Y} = h(\underline{Z}) = 10^{-3} [101. + 31.Z_2 - 2.Z_4 -$

$$- 6.Z_5 + 2.Z_2^2 - 0.3Z_2Z_4 - 1.Z_2Z_5]$$

As in example 3 only terms with coefficients
greater than or equal to 0.001 are shown in the
KP case. Both methods seem to agree that X_2 has
a great influence in the variability of Y.

In the LHS case the stepwise regression stops
after including 4 of the 27 possible terms of up
to 2'nd order $(X_1, \ldots, X_6, X_1^2, \ldots, X_6^2, X_1X_2, \ldots, X_5X_6)$.
It is interesting to notice that neither X_1, X_3
nor X_6 has any significant contribution to the
variation in Y, and that the explanation value
for this expression is 98 %.

5   MORE ABOUT PRECISION

In the previous chapter methods were examined trying to find a method that estimates a distribution as close to the "true" distribution as possible at a low cost (small computer time). Figures 4.2 and 4.3 show the uncertainty in the cumulative distribution function, due to the different methods. Even given a method, there is another uncertainty in this estimation due to the randomness in the sampling technique. This problem will be discussed under the following two points.

1   Estimate the precision of the 5 % and 95 % percentile values of Y for a given method and

2   Compare the precision of Latin Hypercube Sampling (LHS) and simple random sampling as is used in Crude Monte Carlo methods (MC)

To investigate point 1 the same system as in example 2 will be used

$$Y = X_1 + 2X_2 + X_3^2 + 2X_4^2 + X_5^3 + X_6^4$$

where all X_i are uniformly distributed over the interval (0,2).

In the case of a crude MC method the precision of the percentile values can be estimated, see Leverenz (1981). For the LHS method this seems to be more difficult to do. There is one simple but time consuming way, which have been tried in the following.

The LHS method have been repeated 25 times with 20 intervals (calculations of Y) and different seed (random selection seed) each time. In this way 25 estimates of $Y_{.05}$ and $Y_{.95}$ have been made which enables the calculation of the mean and standard deviation for the percentiles.

In Appendix 2 sample characteristics for the 25 samples are shown.

$\bar{Y}_{.05} = 4.9$ $\bar{Y}_{.95} = 23.6$ mean

$S_{.05} = 0.8$ $S_{.95} = 1.8$ standard deviation

One way to illustrate $\hat{Y}_{.\alpha}$ and it's precision is by $\bar{Y}_{.\alpha} \pm S_{.\alpha}$. These intervals are illustrated in Figure 5.1.

Comparing Figure 4.3 and 5.1 we see that the uncertainty due to modelling is of the same size as the uncertainty due to sampling.

Under point 2 we want to compare the precision of the percentile estimators obtained above by repeated LHS to the corresponding precision measure for simple random sampling. In order to ease the comparison, let us consider

the 'true' percentile values (taken from the "TRUE" of example 2) $Y_{.05} = 4.51$ and $Y_{.95} = 22.91$.

and

the probability estimators $\hat{p}_\alpha = \upsilon_\alpha / N$, where υ_α observations in each sample of size N are less than or equal to Y_α.

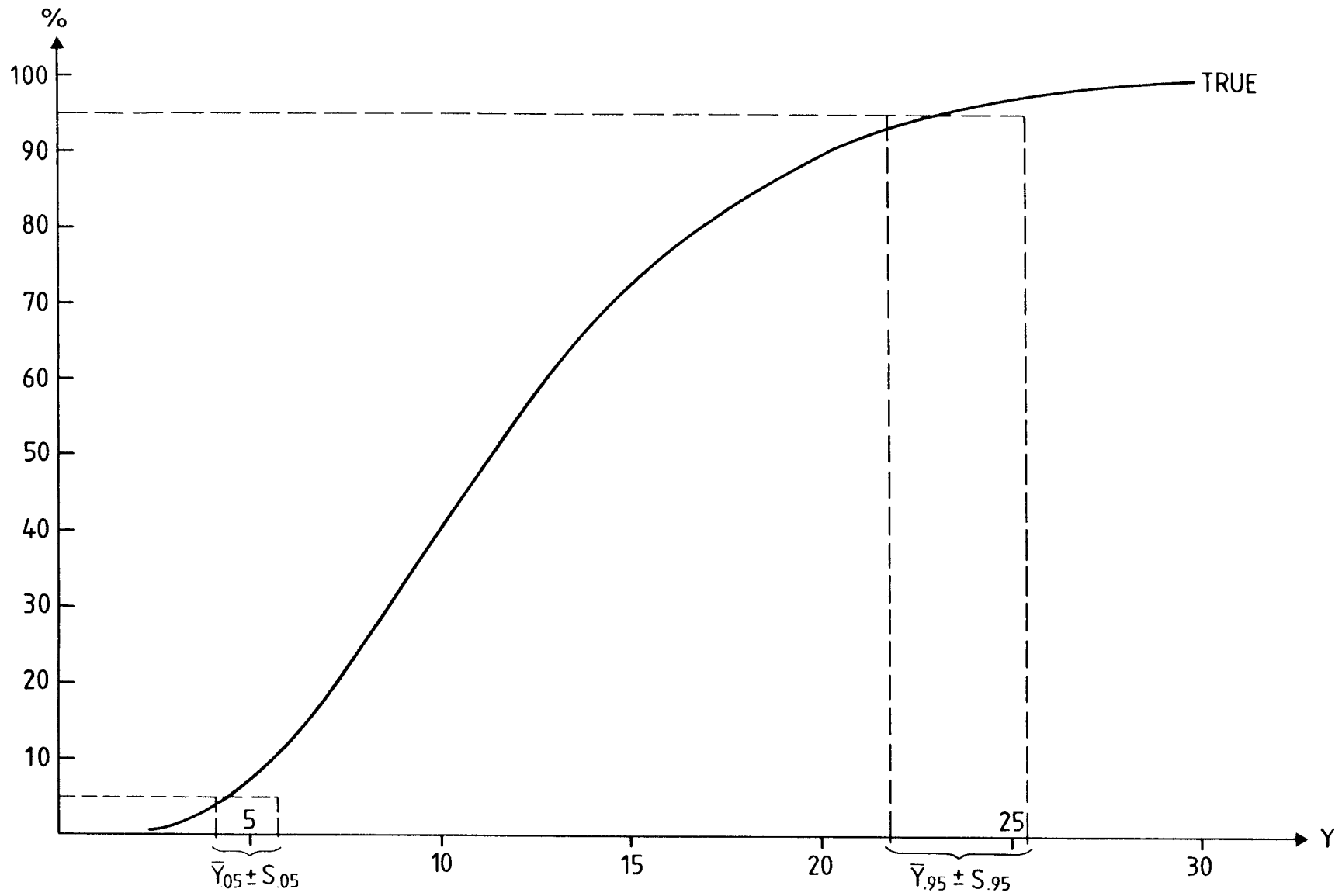From the 25 replicated LHS samples we obtain the following mean values and standard deviations

Figure 5.1

Precision in $Y_{.\alpha}$ for LHS method with 20 intervals

α	$E(\hat{p}_\alpha)$	$D(\hat{p}_\alpha)$	
5 %	0.038	0.039	
95 %	0.942	0.028	(5.1)

In the MC case the variable υ_α is binomially distributed $\upsilon_\alpha \sim B(N,\alpha)$. Thus the corresponding mean values and standard deviations can be written

α	$E(\hat{p}_\alpha)$	$D(\hat{p}_\alpha)$	
5 %	0.005	$\sqrt{\dfrac{0.05(1-0.05)}{N_{MC}}}$	
95 %	0.95	$\sqrt{\dfrac{0.95(1-0.95)}{N_{MC}}}$	(5.2)

By comparing the standard deviations in (5.1) and (5.2) as precision measures we can calculate the sample size N_{MC} required to yield the same precision as the LHS method with sample size $N = 20$. The result of such a calculation is $N_{MC} = 31$ for $\alpha = 5$ % and $N_{MC} = 62$ for $\alpha = 95$ %.

This result confirms what can be derived analytically, namely that LHS based estimators of a simple type, such as estimators for the mean and the cdf, have a smaller variance than corresponding simple random sampling estimators. The degree of variance reduction varies from case to case. In the example above we notice a difference of factor 3 in sample size for the estimation of the 95 % percentile.

6. SUMMARY AND CONCLUSIONS

The scope of this report is to investigate methods to be able to have control over the uncertainties in a complex function, $Y = g(\underline{X})$. The work has been done through literature studies, examples etc described in the previous chapters. Here a summary is made over the most important results and questions, which we want to pay attention to.

1) The objective of the uncertainty analysis for a model $Y = g(\underline{X})$ is to determine the cumulative distribution function (cdf) of Y on the basis of the distribution functions for the input variables $\underline{X}$. The purpose of the sensitivity analysis is to mutually rank the variables $\underline{X}$ according to their contributions to the spread in the distribution of Y.

2) For economical reasons and from the sensitivity analysis point of view it may be useful to determine an approximating response surface (RS) $\tilde{Y}$ for the model $g(\underline{X})$, $\tilde{Y} \cong g(\underline{X})$. The calculation of a response surface can be done either with randomly selected points (simple random sampling, Latin Hypercube Sampling LHS) and stepwise regression or systematically selected knot-points (KP) and interpolation/least squares fitting. The cdf of Y can be determined either directly through randomly selected values of the input variables or by utilizing the simpler model $\tilde{Y}$.

3) In this report, the different approaches mentioned above have been applied to four simple examples. Calculation of a RS seems to be more efficient with the KP-method than by the LHS-

method in the very simple examples 1 and 2. In
the ratio models, examples 3 and 4, it is more
difficult to rank the methods from an efficiency
point of view. Because of the simple nature of
the examples, the experiences gained are far
from being generally valid. However, the KP-
method could be applicable even for more complex
models since there are possibilities to trans-
form both the input variables $\underline{X}$ and the response
Y. The same transformation technique can be used
for the LHS method too. Then the strength of the
stepwise regression procedure can be enhanced
considerably. If the cdf is determined directly
by LHS, the regression curve is used for ranking
purposes only. In that case the requirements on
an accurate RS are less severe.

4) The estimation of the cdf of Y via a response
surface $\tilde{Y}$ must be done with great care. If $\tilde{Y}$
does not reflect the variability of Y well
enough, systematic disagreements in the distri-
bution functions will occur (see 4.2, Figure 4.3).
The determination of the response surface can
probably not be automated.

5) Estimators of the form $\Sigma f(Y_i)$ are unbiased when
Y_i are selected both by simple random sampling
and LHS. This type of estimator is used for the
mean response E(Y) and the cdf of Y. If
$Y = g(X_1, \ldots, X_k)$ is monotonic in each of its
arguments, and f(Y) is a monotonic function of Y,
the LHS-based estimator has a smaller variance
than corresponding estimator based on simple
random sampling.

Estimators for the variance V(Y) and the percen-
tiles $Y_{.\alpha}$ are not of this simple form. However,
simulation studies (McCay, Conover, Beckman,

1979) have shown that the bias of the variance
estimator (3.6) is small and that this estimator
for the LHS method has a better precision com-
pared to simple random sampling. Analogous re-
sults have been obtained in this report
(Chapter 5) with regard to LHS-based estimators
of the percentiles $Y_{.05}$ and $Y_{.95}$. Thus there are
certain evidences indicating that the LHS-method
is an efficient variance reduction method.

6) A comparison of methods should not only include
the measured precision in some sample character-
istics but also the computer time to make the
appropriate calculations. A measure of ef-
ficiency ε is suggested,

$$\varepsilon = \frac{t_1 \sigma_1^2}{t_2 \sigma_2^2}$$

where t_i = computer time by method i

σ_i = standard deviation by method i

(see Chapter 3, page 17).

Until a better procedure is established, batch-
wise repetition of the sampling can be used to
obtain estimates of the sampling precision (see
Chapter 3 (3.14) in cases where variance re-
duction methods are used.

7) The "contribution to variance" concept, being
used in the stepwise regression analysis, is use-
ful for ranking purposes. In fact, it is equiv-
alent to standardized regression coefficients.

8) In a pathway model analysis by Iman & Helton
(1985) (which is much more realistic than the
simple examples reported here) one draws similar
conclusions. They state for instance, that a re-

sponse surface based on a fractional factorial design (FFD) selection of points is good in some cases and poor in others.* The direct LHS method N = 50 and random sampling (N = 500) are in good agreement.

9)   A further conclusion in the analysis mentioned above is that direct variance estimates of type (3.6) tend to produce a more reliable measure of spread than do variance estimates based on response surfaces.

10)   Response surface replacement based on systematic selection of knot-points is difficult to use in case of dependent input variables. For multi-variate input structures LHS can be used to produce required rank correlations, according to Iman & Conover (1982).

---

*   The model contains 20 input variables, behaving independently and producing nonlinear output which is a monotonic function of each of the input variables. A FFD with N = 128 points was used. The generalized interpolation scheme for a second-degree response surface would have required N = 231 points.

REFERENCES

COX, N D
Comparison of Two Uncertainty Analysis Methods.
Nuclear Science and Engineering 64, pp 258-265,
1977.

COX, N D and MILLER, C F
User's Description of Second-Order Error
Propagation Computer Code for Statistically
Independent Variables (SOERP).
RE-S-76-138, Idaho National Engineering
Laboratory, 1976.

VAURIO, J K and MUELLER, C
Probabilistic Analysis of Liquid-Metal Fast
Breeder Reactor Accident Consequences with
Response Surface Techniques.
Nuclear Scienec and Engineering 65, pp 401-413,
1977.

VAURIO, J K
PROSA-2: A Probabilistic Response-Surface
Analysis and Simulation Code.
Argonne National Laboratory, ANL-81-33, 1981.

IMAN, R L and HELTON J C
A Comparison of Uncertainty and Sensitivity
Analysis Techniques for Computer Models.
NUREG/CR-3904, SAND84-1461, 1985.

McKAY, M D, CONOVER, W J and BECKMAN, R J
A Comparison of Three Methods for Selecting
Values of Input Variables in the Analysis of
Output from a Computer Code.
Technometrics 21 (2) pp 239-245, 1979.

IMAN, R L and CONOVER W J
A Distribution-free Approach to Inducing Rank
Correlation Among Input Variables.
Commun Statist-Simula Computa, 11(3) pp 311-334,
1982.

LEVERENZ, F L
SPASM. A Computer Code for Monte Carlo System
Evaluation.
NP-1685, EPRI, 1981.

McGRATH, E J and IRVING, D C
Techniques for Efficient Monte Carlo Simulation.
Vol III, Variance Reduction.
ORNL-RSIC-38, 1975.

COMPUTER CODES USED IN THE ANALYSIS

In this report most of the calculations are done with special computer codes for the different methods. A brief discussion is presented in Appendix 1 on each one of the following codes

Code	Method	Ref
PRISM	LHS	1
SPASM	Crude Monte Carlo	2
PROSA-2	Knot Point	3
SOERP	Method of Moments	4

Within this project these four codes have been implemented at the Studsvik computer, where the calculations for the examples have been made. Both the PROSA-2 and the SOERP codes are also available at the NEA data bank.

PRISM

This code is developed to be a systematic method for determining the effect of parameter uncertainties on model predictions. The code is a system of programs which is composed of three steps.

In PRISM 1, the means, variances and covariances, type of distribution, and upper and lower limits of each parameter are given as input to define their probability density functions. With a known distribution for each parameter the Latin Hypercube Sampling method is used to generate sets of random parameter values which are output to a file.

PRISM 2 reads each set of parameter values, evaluates the model $Y = g(\underline{X})$, and outputs the model predictions.

PRISM 3 statistically evaluates and summarizes the joint set of model parameters and predictions. Also a stepwise regression analysis is performed to calculate a best response surface $\tilde{Y} = h(\underline{X})$.

SPASM

Originally the code was prepared along with the development of the WAM-series of codes, which is a code package for dealing with fault tree analysis. In our application SPASM has been given the function $Y = g(\underline{X})$ and the distribution $f_i(X_i)$ for each parameter X_i as input. Then with a crude Monte Carlo technique a sample of size N has been performed. The crude Monte Carlo technique means

1   Generate a random value R_V from a uniform distribution over $(0, 1)$

2   Set $R_V = F_i (Z_i) = \int_{\infty}^{Z_i} f_i (X_i) \, dX_i$

    where $f_i(X_i)$ is the given probability density for the parameter X_i

3   Solve expression above for Z_i, which is one sample value for X_i.

From the N evaluations of Y the distribution is estimated. As output the code gives the cumulative distribution at specified levels.

PROSA-2

This code was originally developed because the use of so called response surfaces leads to considerable savings in computer time in comparison to direct simulation with long running accident analysis codes within the nuclear reactor field.

The code contains a vast number of possibilities for different types of uncertainty and sensitivity analysis. For the purposes of this report the following facilities of the code have been used. PROSA-2 uses a response surface technique for obtaining a probability distribution for a response $Y = g(\underline{X})$. In this technique probability distributions are assigned to the parameters X_i. A limited number of parameter values (called knot points) are selected and input to $Y = g(\underline{X})$. The results of the deterministic analysis are used to generate an analytical function $\tilde{Y} = h(\underline{X})$, called response surface, which approximates the original function $Y = g(\underline{X})$. The analytical function $\tilde{Y}$ is then used in a Monte Carlo Type simulation to estimate the distribution of Y.

The response surface methodology of the code includes both systematical and random knot point selection schemes, second- and third-degree response surfaces, functional transformations of both input parameters and consequence variables, smooth synthesis of regionwise response surfaces and the treatment of random conditions for conditional distributions.

SOERP

The SOERP code is developed for determining the error propagation of uncertainties in the input parameters. The input for running the code is

1   A function of at most second order, in our case the response surface $\tilde{Y}$ is this function.

2   The parameters of Y are independent and for each one of them the first eight moments shall be given.

The output is the first four moments of the resulting distribution.

References

1. GARDNER, R H, RÖJDER, B and BERGSTRÖM, U
    A systematic method for determining the
    effect of parameter uncertainties on
    model predictions.
    STUDSVIK/NW-83/555.

2. LEVERENZ, F L
    SPASM, a computer code for Monte Carlo
    system evalutation.
    Palo Alto, NP-1685, Jan 1981.

3. VAURIO, J K
    PROSA-2, a probabilistic response-sur-
    face analysis and simulation code.
    Argonne National Laboratory, ANL-81-33.

4   COX, N D, and MILLER, C F
    User's descripction of second-order
    error propagation (SOERP) computer code
    for statistically independent variables.
    Idaho National Laboratory, March 1978.

25 REPETITIONS OF LHS

To analyse the precision in the estimation of the $\alpha\%$ percentile value, $Y_{.\alpha}$, using the LHS method, 25 repetitions of the LHS method with 20 intervals were made. The system used is the same as in example 2 and the only difference between the repetitions is a new random seed value. In each repetition some of the distribution characteristics of Y were estimated and are listed in Table 1.

In Table 1 also the number of observations less than a given $Z_{.\alpha}$-value is noticed. This is done to estimate the precision of the probability of finding $Y_{.\alpha}$-values less than a given $Z_{.\alpha}$-value.

Table 1

25 repetitions of the LHS method with 20 inter-
vals. The same system Y = g($\underline{X}$) is used as in
example 2.

Repetition number	Mean	5% perc	95% perc	Nr of obs< *)$Z_{.05}$	Nr of obs < **)$Z_{.95}$
1	12.3	5.74	23.07	0	19
2	12.2	5.82	22.74	0	19
3	12.1	3.60	23.44	1	18
4	12.2	6.02	26.53	0	18
5	12.2	4.86	22.93	0	19
6	12.1	4.60	21.32	1	19
7	12.2	4.44	21.17	1	19
8	12.2	5.19	22.67	1	19
9	12.2	6.77	25.98	0	19
10	12.3	4.42	23.13	1	19
11	12.0	3.97	22.03	3	19
12	12.2	4.15	22.47	1	20
13	12.3	5.70	22.99	0	19
14	12.2	4.94	27.21	1	18
15	12.2	5.15	24.89	0	19
16	11.9	3.98	22.16	2	19
17	12.3	5.89	22.89	0	19
18	12.2	3.99	22.26	2	19
19	12.1	5.64	26.25	0	19
20	12.3	5.66	22.16	0	19
21	12.1	4.44	26.27	1	18
22	12.2	4.94	25.64	1	19
23	12.1	4.42	23.69	1	19
24	12.3	4.54	21.82	1	20
25	12.2	4.26	25.50	1	18

*) $Z_{.05}$ = 4.51 taken from TRUE, example 2.

**) $Z_{.95}$ = 22.91 " -

List of Technical Reports

TR 85-08
In situ one-year burial experiments with simulated nuclear waste glasses
Larry L Hench, Derek Spilman and T Buona-
quisti
College of Engineering, Univ. of Florida,
Gainesville, USA
Alexander Lodding
Chalmers Univ. of Technology, Gothenburg,
Sweden
Lars Werme
SKB, Stockholm, Sweden

TR 85-09
Concentration and distribution of natural radionuclides at Klipperåsen and Bjulebo, Sweden
Björn Sundblad, Ove Landström, Rune Axelsson
Studsvik Energiteknik AB, Nyköping, Sweden

TR 85-10
Chemical interactions between the bentonite and the natural solutions from the granite near a repository for spent nuclear fuel
Bertrand Fritz and Marie Kam
Université Louis Pasteur de Strasbourg, Insti-
tut de Géologie, France
July 1985

TR 85-11
Hydrochemical investigations in crystalline bedrock in relation to existing hydraulic conditions: Experience from the SKB test-sites in Sweden
John Smellie, Nils-Åke Larsson
Swedish Geological Company, Uppsala,
Sweden
Peter Wikberg
Royal Institute of Technology, Stockholm
Sweden
Leif Carlsson
Swedish Geological Company, Göteborg,
Sweden
November 1985

TR 85-12
Hydrogeological investigations and tracer tests in a well-defined rock mass in the Stripa mine
Peter Andersson
Carl-Erik Klockars
Swedish Geological Company
Division of Engineering Geology
Uppsala 1985-11-29

TR 85-13
Analysis of hydrodynamic dispersion in discrete fracture networks using the method of moments
Anders Rasmuson
Dep of Chemical Engineering, Royal Inst of
Technology Stockholm
June 20, 1985

TR 85-14
Radionuclide migration in strongly fissured zones — The sensitivity to some assumptions and parameters
Anders Rasmuson, Ivars Neretnieks
Dep of Chemical Engineering, Royal Inst of
Technology Stockholm
August 7, 1985

TR 85-15
Diffusion measurements of cesium and strontium in biotite gneiss
Kristina Skagius, Ivars Neretnieks
Dep of Chemical Engineering, Royal Inst of
Technology Stockholm
1985-12-30

TR 85-16
The corrosion of spent UO_2 fuel in synthetic groundwater
R S Forsyth
Studsvik Energiteknik AB, Nyköping, Sweden
L O Werme
The Swedish Nuclear Fuel and Waste Manage-
ment Co (SKB), Stockholm, Sweden
J Bruno
Royal Institute of Technology, Dept of Inor-
ganic Chemistry, Stockholm, Sweden
October 1985

TR 85-17
**Sealing of rock fractures
A survey of potentially useful methods and substances**
Roland Pusch*, Mikael Erlström, Lennart
Börgesson
Swedish Geological Co, Lund
* also Lund University of Technology and
Natural Sciences, Lund
December 1985