ELSEVIER

Journal of Hazardous Materials A91 (2002) 1–24

Journal of Hazardous Materials

www.elsevier.com/locate/jhazmat

Use of the most likely failure point method for risk estimation and risk uncertainty analysis

B. Mitchell*

WS Atkins Consultants Ltd, Woodcote Grove, Ashley Road, Epsom, Surrey KT185BW, UK

Accepted 31 October 2001

Abstract

The most likely failure point (MLFP) method, developed within the field of structural reliability analysis (where it is known as the FORM/SORM method) is a technique for estimating the risk (probability) that a calculated quantity Q exceeds a set limit Q_{lim} when some or all of the inputs to the calculation are uncertain.

It can be used as an efficient stand-alone method for this type of risk calculation. However, for application within the field of toxic hazards, it is proposed as a means for performing sensitivity analyses, possibly in parallel with a risk calculation carried out by conventional methods.

The basis of the method is outlined and its use is demonstrated by means of an example calculation of the risk arising from an installation containing chlorine. The calculation uses, as a consequence model, commercial software for the prediction of dense gas transport. The risk estimate is shown to be acceptably close to that obtained by the Monte Carlo method. The use of a proposed screening procedure utilising the sensitivity formulas that the method provides, in order to identify the most significant uncertainties, is demonstrated.

The identification of a single set of input values containing sufficient information to summarise (at least approximately) the entire risk analysis is considered to be an important feature of the method and is proposed as the basis of a means for assessing the validity of the consequence model.

Keywords: Risk; Toxic; Hazard; Uncertainty; Sensitivity

1. Introduction

This paper describes an efficient method for performing sensitivity analyses for certain types of risk calculation, i.e. those based on a (usually complex) deterministic model

* Tel.: +44-1372-387768; fax: +44-1372-740055.
E-mail address: brian.mitchell@wsatkins.com (B. Mitchell).

0304-3894/02/$ – see front matter
PII: S0304-3894(01)00378-8

Nomenclature

c	chlorine release rate (kg s^{-1})
C	chlorine concentration (ppm)
d	downwind distance (m)
D	toxic dose (ppm^2 min)
D_d	dangerous toxic dose
f	frequency distribution
f_D	fraction of deaths arising from toxic dose
f_T	frequency of a toxic release
F	subscript denoting failure region
i	index for uncertain variable; $i = 1 \ldots N$
I_i	importance of uncertain variable i
M_g	inventory of tank (kg)
N	number of uncertain input variables to consequence model
p	probability density
P	probability
P_{f1}	first-order estimate of failure probability
P_{ij}	joint probability assigned to cell i, j in quadrature method
P_X	cumulative probability distribution applying to variable X
P_1	first-order estimate of probability integral over region beyond failure surface, relative to best-estimate point
Q	quantity predicted by consequence model
Q_{lim}	limit for acceptable values of Q
R	risk
R_c	conditional risk
R_t	risk per unit time
S	spread of toxicity tolerance
T	mean toxicity tolerance
u	value of standard normal variable
u_{im}	value of u_i at the most likely failure point
v	wind speed (m s^{-1})
V	toxic dose variable, $V = \ln(D)$
x	value of input quantity to consequence model
z_0	ground roughness length (m)

Greek letters

β	distance (number of standard deviations) from best-estimate point to most likely failure point
δ	diameter of leak hole (mm)
λ_i	parameter in probability distribution assigned to input variable i
μ	mean of distribution
σ	standard deviation of distribution
τ_m	duration of release (min)
ϕ	standard normal density distribution
Φ	cumulative standard normal distribution

which calculates the consequence of a hazardous event. The concept of risk in such a case arises because of the requirement that the consequence (however quantified) should not exceed a certain acceptable limit, given that the values of the inputs to the model are uncertain.

As an example, consider the assessment of risk arising from the release of a toxic material from a storage site. Deterministic models exist which can predict downwind concentration arising from such a release. However, neither the release characteristics nor the atmospheric conditions at the time can be predicted, and the model will often require the input of other data whose values may be subject to some uncertainty. It is common practice to represent the former group of these uncertain quantities by probability distributions and to assign best-estimate values to the latter.

The problem then facing anyone performing this type of analysis is that of demonstrating that the result is not sensitive to uncertainties in the input. This can be a particular problem if the calculation contains a significant number of input parameters, all of which can be considered uncertain to an extent and whose exact role within the calculation is not always understood in detail. It is normal to perform a standard sensitivity study, whereby various input parameters are altered and the problem re-run. However, this is often unsatisfactory and can be a laborious process for a risk analysis, requiring that the full analysis be repeated a number of times.

This paper describes a method which produces not only an estimate of risk, but also, as a by product, the sensitivities of the risk to all the inputs, including the parameters of the probability distributions assigned to them. This method has been used routinely in the field of structural reliability for the past two decades or so, where it is used to predict the risk of structural collapse. It goes by the name of the FORM/SORM method, but is referred to here by the more descriptive name of the most likely failure point (MLFP) method, for reasons which will become apparent.

Despite its strong association with structural reliability, the MLFP method is quite general in nature and can be used for any application where it is necessary to estimate the probability P that some quantity Q (calculated using uncertain data) exceeds a set limit Q_{lim}. The method is described in detail in [1–3] and more recent lists of references can be found in [4,5].

The following therefore describes:

- an example of a simple risk calculation for chlorine release,
- the basis of the MLFP method,
- how the MLFP method can be used in a real case,
- the interpretation of results from the example case described above.

2. Example of a toxic risk calculation

Fig. 1 shows a fictitious but characteristic plot of predicted toxic dose D (at a selected downwind location) as a function of chlorine release rate c and wind speed v, with all other quantities input to the consequence model assumed to be fixed. The predictions are represented by contours of constant D, with one contour being of particular importance since it represents the value of the 'dangerous dose' D_{d}. Doses below D_{d} are considered to be acceptable ('pass'); those above D_{d} unacceptable ('fail').

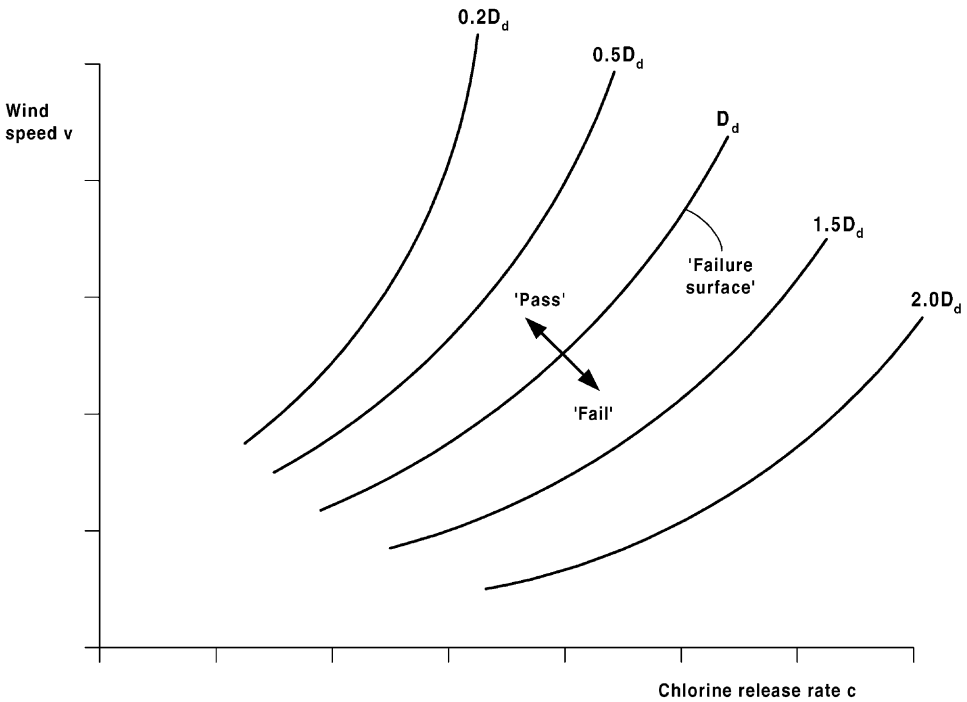

Fig. 1. Contour plot of notional toxic dose calculation, showing predicted dose D as a function of chlorine release rate c and wind speed v. D_d represents the 'dangerous dose'.

The values of the other contours are represented as multiples or fractions of D_d. The region below and to the right of the D_d contour represents all those combinations of v and c which cause the dangerous dose to be exceeded, and is termed the failure region. The region above and to the left therefore represents the pass region and the line $D = D_\mathrm{d}$, which is the boundary between the two, is called the failure surface. Though illustrated here with only two axes, the approach and terminology can be applied to any number of variables.

Assume now that a very simple quantitative risk assessment is being carried out with this consequence model. The wind speed and the chlorine release rate at the time of any postulated release are, of course, unknown and have to be represented by frequency or probability distributions. If the risk R is defined as the probability that someone located downwind receives a dose $D > D_\mathrm{d}$, then, with reference to Fig. 1, this can be visualised as the probability that the combination of c and v will lie somewhere in the fail region when the release occurs. In other words, if $f(c)\,\mathrm{d}c$ is the frequency (probability per unit time) of a release of rate between c and $c + \mathrm{d}c$, and $p(v)\,\mathrm{d}v$ is the probability that the wind speed lies between v and $v + \mathrm{d}v$, then the risk R_t (per unit time) is given by

$$R_t = \int_F f(c)\,p(v)\,\mathrm{d}v\,\mathrm{d}c \tag{1}$$

where the F indicates that the integration is carried out over the failure region—below and to the right of the $D = D_d$ contour. If $f(c)$ is defined as:

$$f(c) = f_T p(c) \tag{2}$$

where f_T is the frequency (per unit time) of a release of any size; $p(c)$ probability; given that a release has occurred, that the rate is between c and $c + dc$ then:

$$R_t = f_T \int_F p(c)\, p(v)\, dv\, dc = f_T R_c \tag{3}$$

where

$$R_c = \int_F p(c)\, p(v)\, dv\, dc \tag{4}$$

is the conditional risk (i.e. given that a release actually occurs) predicted by the calculation. In what follows, the term 'risk' always refers to this conditional risk R_c, unless stated otherwise.

Calculation of R_c therefore requires evaluation of integrals of the type given in Eq. (4). Two conventional methods for performing this integration: the quadrature method and the Monte Carlo method, are contrasted with the MLFP method in the next section.

3. Comparison of risk assessment methods

The MLFP method, in common with other methods for quantitative risk assessment, starts with the assignment of probability distributions to any quantities which are uncertain, either because they vary with time (atmospheric variables or toxic inventory) or are otherwise inherently unpredictable (e.g. release rate). In addition, it is desirable, though not always feasible (see below) to include probability distributions for other model input parameters (e.g. ground roughness length), reflecting the user's uncertainty as to the most appropriate value to use.

In terms of the simple example introduced above, both the release rate and the wind speed require representation by probability distributions, whereas the other input quantities are assumed (for now) to be fixed. The following contrasts the MLFP method with two conventional risk calculation methods: quadrature and Monte Carlo.

3.1. The quadrature method

In the quadrature method, the (c, v) plane is subdivided into regions or cells defined by suitable intervals over the normal ranges of the variables. Fig. 2a gives a simple example Let cell (i, j) be the cell formed by the intersection of the ith interval on the c-axis and the jth interval on the v axis, where i runs from 1 to i and j from 1 to J so that there are I, J cells altogether. Each cell is assigned a probability P_{ij} that the actual values of c and v will simultaneously fall within it, where the probability P_{ij} are seen to sum to 1.0.

The method proceeds as follows:

For every cell (i, j), a representative pair of values $(\bar{c}_i, \bar{v}_j)$ is selected (usually at the centre of the cell) and a dose calculation $D(\bar{c}_i, \bar{v}_j)$ is performed. If $D > D_d$ (a 'fail'), the value of

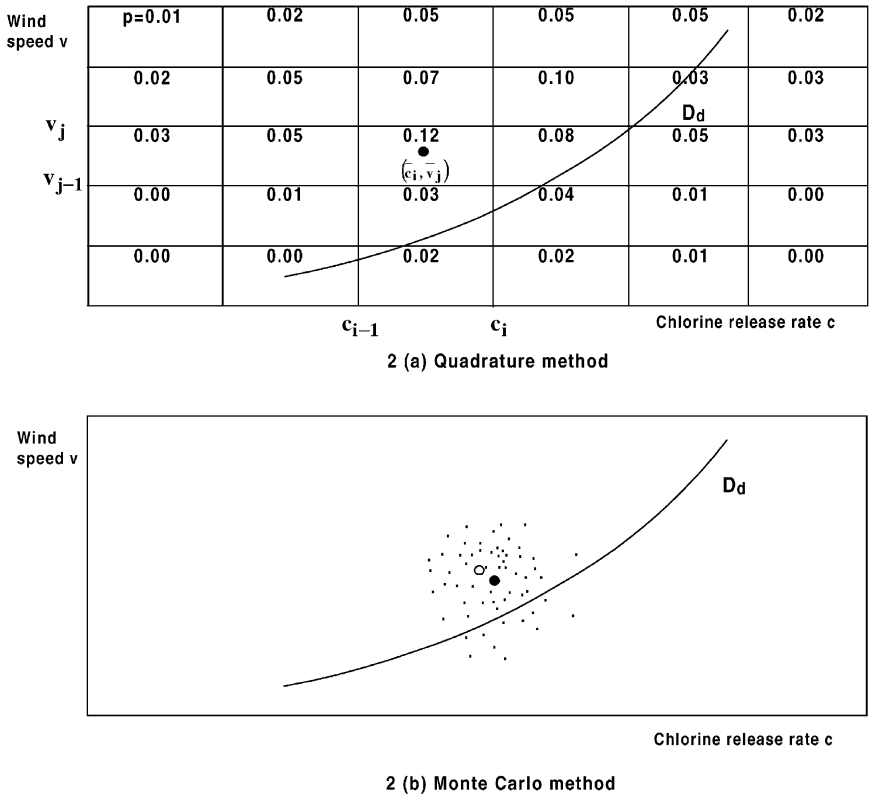

Fig. 2. Illustration of (a) quadrature and (b) Monte Carlo methods, for calculation of risk.

P_{ij} is added to an accumulator. When all the cells have been covered, the sum of P_{ij} over all the fail cells is the estimate of the risk R_c.

The quadrature method is straightforward and robust and can cope with any kind of probability distributions. Its disadvantages are that the subdivision of the probability distributions usually has to be done 'by hand', that the method's accuracy depends upon the size of the cells, and that only a limited number of uncertain variables can be treated in this way. An extra variable implies an extra dimension: the product of *I*, *J* cells which have to be visited in the above example becomes *IJK* cells if another variable is added, so that there is a severe penalty for adding other variables.

3.2. The Monte Carlo method

The Monte Carlo method is illustrated in Fig. 2b. Values of c and v are generated randomly in pairs from their probability distributions and, for each pair (c_k, v_k), the dose $D(c_k, v_k)$ is calculated. The number of times that D exceeds D_d is counted and, when a sufficient number of fails has occurred, this value is divided by the total number of evaluations to give R_c.

The Monte Carlo method is reliable and a useful technique for bench-marking the other methods. Its well-known disadvantage is that, if the risk is fairly small, a very large number of evaluations may be required to give a converged estimate of R_c. It may be seen that the addition of further uncertain variables is quite straightforward in the Monte Carlo method, requiring only the generation of an extra random value per consequence model calculation.

In neither of the above methods it is easy to perform an estimate of the sensitivity of the predicted risk to the parameters assumed for each of the probability distributions. In the MLFP method however, this is straightforward, as will now be described.

3.3. The MLFP method

3.3.1. Estimating R_c

The MLFP method works on a different principle from those techniques described above. Referring to Fig. 2b, assume that the medians of the distributions are located at point O, which in this example lies in the pass region. O is termed here the best-estimate point (BEP). The reason for the choice of median, rather than mean, will become evident in what follows. It is reasonably clear that the further the point O lies from the contour $D = D_d$, the smaller the probability that D will exceed D_d. The principle of the MLFP method is first, to define a standardised coordinate systems in which this distance can be expressed, then to calculate the distance (by finding the point of closest approach of the contour to O) and then finally to estimate R_c by performing an analytic integration over a region which approximates the failure region. This is in contrast to the conventional methods, which produce an approximate numerical integral over the true failure region.

The transformation to the standard coordinate system is achieved as follows. Assume that an input quantity x to the consequence model is uncertain. This uncertainty is represented by a continuous probability density distribution $p_x(x)$ or, alternatively, by the cumulative probability distribution (the integral of $p_x(x)$) $P_x(x)$. The value of x is converted to the value of an alternative variable u_x by means of solving the equation:

$$\Phi(u_x) = P_x(x) \tag{5}$$

where

$$\Phi(u) = 0.5\left(1 + \mathrm{erf}\left(\frac{u}{\sqrt{2}}\right)\right) \tag{6}$$

is the cumulative standard normal distribution; i.e. the cumulative distribution corresponding to the normal (i.e. Gaussian) density distribution ϕ for a variable of mean 0.0 and standard deviation 1.0, i.e.

$$\phi(u) = \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \tag{7}$$

Each variable u_x thereby has a standard normal probability distribution and is referred to as the standard normal variable corresponding to x.

The best-estimate values for the uncertain variables are taken to be those defined by the medians of the probability distributions. For the example given above, this is the point where:

$$P_c(c) = 0.5 \quad P_v(v) = 0.5 \tag{8}$$

Assume that this point (the best-estimate point, or BEP) lies somewhere in the pass region. From (6) and (8), the point corresponds to:

$$u_c = 0 \quad u_v = 0 \tag{9}$$

i.e. to the origin of the standard normal co-ordinate system.

The relationship between the best-estimate point and the failure surface in the new co-ordinate system is illustrated in Fig. 3. Since each of the transformed variables now has the same (standard normal) probability distribution, the combined distribution is a function

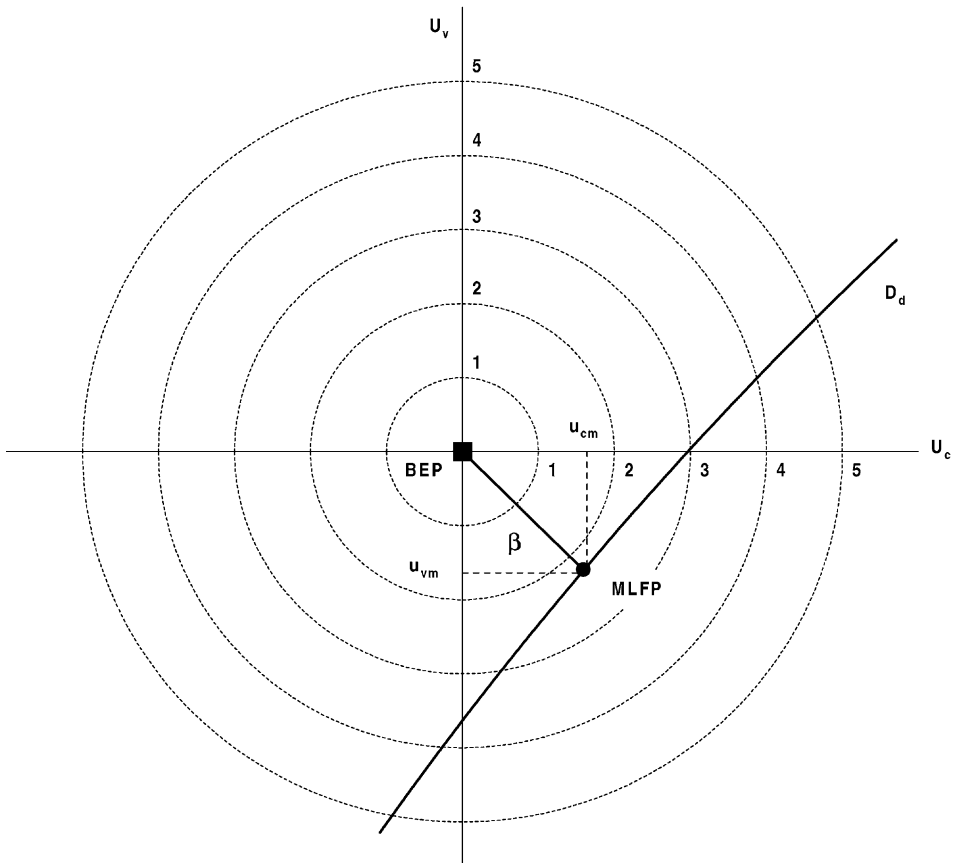

Fig. 3. Dose contour plot in standard normal co-ordinates (schematic). Location of best-estimate (median) point (BEP) and most likely failure point (MLFP).

only of distance from the BEP and the circles shown in the figure also represent contours of constant probability density. The point of closest approach therefore also possesses the maximum probability density (in the transformed system), and is therefore known as the most likely failure point, or MLFP.

If the distance (essentially the combined number of standard deviations) from the BEP to the MLFP is denoted by β, then the integral of the probability density over the region beyond the failure surface (relative to the position of the BEP) can be estimated to first-order by the quantity P_1, where

$$P_1 = \Phi(-\beta) \tag{10}$$

which applies to any number of uncertain variables. Therefore, the first-order failure probability P_{f1} is given by

$$P_{f1} = P_1 \quad \text{if the BEP lies in the pass region,} \tag{11a}$$

and

$$P_{f1} = 1 - P_1 \quad \text{if the BEP lies in the fail region} \tag{11b}$$

P_{f1} is the first-order reliability method, or FORM, approximation to R_c, and assumes that the failure surface can be approximated by a straight line (or flat plane, or hyperplane) tangent to the actual failure surface at the MLFP. A second-order reliability method (SORM) approximation, which uses the curvature(s) of the failure surface to obtain an improved approximation to R_c, can be obtained (see [1–5]). However, the method is proposed here not specifically to calculate an accurate estimate of R_c; it may even be that the full risk analysis is actually performed using one of the conventional methods described above, but to identify the quantities (if any) to which R_c is particularly sensitive. For this purpose, the first-order estimate is sufficient, as described below.

The process of finding the MLFP is an example of a *constrained minimisation* technique [6]. This procedure is not trivial and requires the use of specialised software. Some practical advice to this end is given in Appendix A.

3.3.2. Sensitivity analysis

Assume the general case in which there are N uncertain input variables x_i; $i = 1 \ldots N$ to the consequence model and that the quantity calculated by the consequence model is represented in general by the symbol Q (the 'quantity of interest', replacing dose D in the example). Assume also that the location of the MLFP has been found and that its co-ordinates in the standard normal system are u_{im}; $i = 1 \ldots N$, the subscript m denoting the MLFP values. Suppose that we want to estimate how sensitive R_c is to one of the parameters of the probability distribution assigned to variable i. Using P_{f1} as a reasonable estimate of R_c, it is straightforward to derive the rate of change of P_{f1} with respect to this parameter (call it λ_i) as

$$\frac{\mathrm{d}P_{f1}}{\mathrm{d}\lambda_i} = \mp\left(\frac{\phi(\beta)u_{im}}{\phi(u_{im})\beta}\right)\left.\frac{\mathrm{d}P_{xi}}{\mathrm{d}\lambda_i}\right|_{\mathrm{MLFP}} \tag{12}$$

where the sign depends upon whether (11a) or (11b) applies. Since P_{xi} is usually one of the standard forms of distribution, the value of $dP_{xi}/d\lambda_i$ can be readily derived for the type of distribution used for variable i.

However, it is usually the case (see Section 5) that the sensitivities vary by orders of magnitude from variable to variable, so now assume that we wish to take a more pragmatic approach and require instead only a rough estimate of the sensitivity, so that unimportant variables can be identified with the minimum of effort. The following describes the basis of a screening procedure which identifies the variables to which the risk estimate is clearly insensitive, allowing the user to apply (if required) more accurate methods at subsequent steps in the process.

For a normal distribution with parameters σ_i (standard deviation) and μ_i (mean), the form of (12) is particularly simple.

$$\frac{dP_{f1}}{d\sigma_i} = \pm\frac{\phi(\beta)u_{im}^2}{\beta\sigma_i} \tag{13a}$$

$$\frac{dP_{f1}}{d\mu_i} = \pm\frac{\phi(\beta)u_{im}}{\beta\sigma_i} \tag{13b}$$

For each uncertain variable, let

$$I_i = \left(\frac{u_{im}}{\beta}\right)^2 \tag{14}$$

I_i is termed here the 'importance' of the uncertainty in variable i to the risk calculation, for reasons which will become apparent below. From Pythagoras's theorem it is obvious (see Fig. 3), that

$$\sum I_i = 1 \tag{15}$$

Now, assume that we wish to estimate the sensitivity of R_c to the uncertainty in each variable, as expressed by its standard deviation. For a normal distribution, from (13a) and (14)

$$\frac{dP_{f1}}{d\sigma_i} = \pm\frac{\beta\phi(\beta)I_i}{\sigma_i} \tag{16}$$

In addition, the change in calculated risk arising from the replacement of a normally distributed variable by its median (and hence, mean) value can be estimated as [7]

$$\Delta P_{f1} \approx \mp\frac{\beta\phi(\beta)I_i}{2} \tag{17}$$

Although (16) and (17) apply strictly only to a normal distribution, they are accurate enough (say, within a factor of 2 or so) for any symmetric distribution to help determine whether the sensitivity is likely to be of significance. The following therefore describes the proposed screening procedure. This procedure is seen in operation in Section 5.

Step 1. Inspect the values of I_i for each uncertain variable:

- If $I_i \leq 5 \times 10^{-3}$, the uncertainty in the input quantity is of no significance. The uncertain variable may be replaced by a best-estimate value.

- If $5 \times 10^{-3} \leq I_i \leq 5 \times 10^{-2}$, the uncertainty may be of significance. The sensitivity of the risk estimate may be checked by an appropriate sensitivity formula (Step 2).
- If $I_i \geq 5 \times 10^{-2}$, the uncertainty is probably having a significant influence on the calculated risk. The sensitivity of the risk estimate should be checked by an appropriate sensitivity formula (Step 2).

Step 2. Use the appropriate formula ((17), or (12) if required) to determine the sensitivity of the calculated risk to the uncertainty in the distribution parameter. If ΔP_{f1} is significant compared with P_{f1}, go to Step 3.

Step 3. Alter the parameter and re-run the MLFP calculation (using the original MLFP as a starting guess for the iteration). If ΔP_{f1} is significant compared to P_{f1}, go to Step 4.

Step 4. Review the derivation of the probability distribution for the variables which reach Step 4. Re-run the risk calculation as necessary.

It may be noted that the risk estimates (10) and (11), the importances (14) and the sensitivity formula (17) require, for their evaluation, only the values u_{im}; $i = 1 \ldots N$. In other words, the co-ordinates of one single point, the most likely failure point, contain sufficient information to characterise (at least approximately) the full risk assessment and therefore, provide a compact and standardised format for expressing the result.

4. Example case

The following example illustrates the use of the MLFP method for a risk calculation for an installation containing chlorine. Although something of an idealisation, the calculation contains many of the elements of a conventional risk analysis.

The case assumes the presence of a single 20 t storage tank and the MLFP method is used to calculate the conditional risk R_{c} at a sequence of downwind distances d ranging from 200 m to 20 km, based on an assumed distribution of leak hole size. In order to broaden the scope of the calculation, the quantity of interest is selected now to be not the toxic dose D but the fraction of deaths f_{D} resulting from D. The risk is calculated as the (conditional) probability that $f_{\mathrm{D}} > 0.005$.

The physical model employed is the commercially available dense gas transport software GASTAR, Version 3 [8]. GASTAR uses an integral, or 'box' model to predict the transport of a dense gas plume, averaged over sections of the plume.

The following describes the probability (i.e. frequency or uncertainty) distributions assigned to a selection of the input quantities of the model. These are summarised in Table 1. For the purposes of this example case, all the distributions are assumed to be independent of one another (see Section 4.2) and the remaining input data are assumed to have fixed values.

It should be emphasised that the purpose of the calculation is to produce the sensitivity information which will allow the user to make an informed decision as to which distributions need to be defined more precisely and which can either be left as rough estimates or replaced

Table 1
Summary of probability distributions assigned to consequence model input quantities

Quantity	Distribution type (see foot of table)	Parameters (see foot of table)	Median value
Tank inventory (kg)	Uniform	$x_1 = 0, x_2 = 20{,}000$	10,000
Leak hole diameter (mm)	Truncated exponential	$\lambda = 0.095, x_1 = 2, x_2 = 66$	9.27
Droplet fraction	Uniform	$x_1 = 0.0, x_2 = 1.0$	0.5
Multiple by mass of air entrained at source	Log-normal	$m = 2.26, s = 0.294$, (Gives $\mu = 10.0, \sigma = 3.0$)	9.58
Aspect ratio of the source	Normal	$\mu = 1.0, \sigma = 0.2$	1.0
Wind speed ($m\,s^{-1}$)	Weibull	$k = 1.5, w = 3.0$, (Gives $\mu = 2.71, \sigma = 1.84$)	2.35
Air temperature (K) (the ground temperature is fixed at a value of 290 K)	Uniform	$x_1 = 295.0, x_2 = 300.0$	297.5
Humidity (%)	Uniform	$x_1 = 10, x_2 = 90$	50
Roughness length (m)	Log-normal	$m = -2.65$, s = 0.833, (Gives $\mu = 0.1, \sigma = 0.1$)	0.0707
Toxicity tolerance mean	Normal	$\mu = 14.45, \sigma = 1.5$	14.45
Toxicity tolerance spread	Normal	$\mu = 1.087, \sigma = 0.15$	1.087
Type	Probability density		
Normal (Gaussian)	$(1/\sigma\sqrt{(2\pi)})\exp(-0.5((x-\mu)/\sigma)^2)$		
Log-normal	$(1/s\sqrt{(2\pi)})\exp(-0.5((y-m)/s)^2);\ y = \ln(x)$		
Uniform	$1/(x_2 - x_1) : x_1 \le x \le x_2;\ 0,$ elsewhere		
Weibull (2-parameter)	$(kx^{k-1}/w^k)\exp(-(x/w)^k)$		
Truncated exponential	$= A\exp(-\lambda x) \quad$ for $x_1 \le x \le x_2 = 0,$ elsewhere where $A = \lambda/(\exp(-\lambda x_1) - \exp(-\lambda x_2))$		

by fixed values. It is therefore not required, at this stage in the process, that the user be particularly accurate in selecting the probability distributions, other than that these should be a reasonable representation of either the expected frequency distributions or of the spread inherent in the uncertain quantities.

It should also be noted (Section 3.3.1) that the MLFP method requires the use of continuous distributions. However, a continuous distribution is usually a more realistic representation of the true state of affairs than a discrete distribution.

4.1. Source data

4.1.1. Chlorine inventory and release rate

The case assumes that the installation contains a single tank, whose inventory M_g at the time of the release may be anywhere between 0 and 20 t of chlorine; i.e. M_g is represented by a uniform probability between 0 and 2×10^4 kg. The release is assumed to occur via a hole in the tank or its associated pipework. The probability distribution of hole diameters δ (mm) is assumed to be:

$$p(\delta) = \begin{cases} 0.115 \exp(0.095\delta), & 2\,\text{mm} \leq \delta \leq 66\,\text{mm} \\ 0, & \text{elsewhere} \end{cases} \tag{18}$$

The discharge coefficient is set such that the release rate c (kg s^{-1}) is given by $c = 0.014\,\delta^2$.

Although the case to be modelled is that of a continuous release of chlorine, a toxic dose can be calculated by assuming that this release lasts for a period of length τ_m (minutes), at the end of which the release ceases because either

- the tank has emptied, or
- the release has been stopped by some sort of intervention.

If it is assumed that the release will be stopped by intervention after 20 min, then the release duration τ_m is given by:

$$\tau_m = \min\left(20, \frac{M_g}{60c}\right) \tag{19}$$

The use of the period τ_m in the calculation of the toxic dose D is described below.

4.1.2. Droplet fraction in released vapour/droplet mixture

This has been assumed to be represented by a uniform distribution between 0 and 1; in other words, as wide a spread as possible.

4.1.3. Air entrainment rate

The amount of air entrained into the released material at source is difficult to predict. A multiple by mass of 10 was considered a reasonable mean value, and the standard deviation (representing the uncertainty in this quantity) was set to 3.0. A log-normal distribution with these characteristics was used.

4.1.4. Plume shape at source

GASTAR requires plume width to be specified, and then calculates plume height in order to give the correct mass flow. Plume aspect ratio was selected as the uncertain input parameter, which was represented as a normal distribution with $\mu = 1.0$ and $\sigma = 0.2$. A subroutine was included within a pre-processing routine (Appendix A) in order to calculate the appropriate plume width.

4.2. Atmospheric variables

The frequency distributions of the atmospheric variables which are relevant to a dispersion calculation, these being

- wind speed,
- wind direction,
- atmospheric stability,
- air/ground temperature difference,
- humidity,

cannot in general be represented as independent of one another, nor can they be represented by simple analytic functions. Nevertheless, although techniques do exist within the scope of the MLFP method for treating interdependent variables (see, e.g. [2]), they are represented in a very simple way in this example case. The concentration is calculated directly downwind of the release, so the wind direction is not considered, and the atmospheric stability is set to Pasquill category 'E' (i.e. fairly stable).

The wind speed is represented by a Weibull distribution appropriate to a site where the winds are relatively light, having a mean speed of $2.7\,\mathrm{m\,s^{-1}}$. The ground temperature was fixed at 290 K (17 °C) with air temperature represented by a uniform distribution between 295 and 300 K. Humidity was represented by a uniform distribution between 10 and 90%.

4.3. Ground characteristics

The ground roughness length z_0 is represented by a log-normal distribution with a mean of 0.1 m and a standard deviation also of 0.1 m. The mean value is characteristic of a standard industrial site.

4.4. Toxicity

For a constant pollutant concentration C (ppm) persisting for a time τ_m (minutes) the toxic dose D ($\mathrm{ppm^2\,min}$) experienced at a location is calculated from the following expression:

$$D = C^2 \tau_\mathrm{m} \tag{20}$$

The toxicity can be expressed by means of a population tolerance distribution, which expresses the fraction of deaths f_D in a population as a function of the toxic dose received by an individual. It is found [9] that the distribution is approximately log-normal, i.e. if V is defined as

$$V = \ln(D) \tag{21}$$

then

$$f_{\mathrm{D}} = \Phi(\mathrm{v}) \tag{22}$$

where Φ is as in Eq. (5) and

$$\mathrm{v} = \frac{(V - T)}{S} \tag{23}$$

where T represents the mean tolerance and S the spread of tolerances in the population.

The quantities T and S, despite being parameters in a distribution (of toxic tolerance), can be considered as uncertain inputs to the consequence model and therefore, can themselves be represented by probability distributions in the MLFP calculation. Both T and S are therefore represented by normal distributions, reflecting (in a very approximate way) the variations in toxicity data found in the literature. From published chlorine toxicity (probit) data [9], T is given a mean of 14.45 and a standard deviation of 1.5 and S is given a mean of 1.087 and a standard deviation of 0.15.

5. Results

5.1. Risk estimate

The first property of the method to be demonstrated is that P_{f1}, the first-order estimate, is here a reasonable estimate of the conditional risk R_{c}. Fig. 4 shows the variation of both these quantities as a function of the distance d to the dose point, with R_{c} being calculated by the Monte Carlo method. Although the first-order estimate begins to drift away from the true value at large distances from the source, it remains sufficiently close such that an estimate of its sensitivity would be expected to be a reasonable measure of the sensitivity of the actual risk.

The number of consequence model calls for the cases shown in Fig. 4 varied from case to case, with a minimum of 61 and a maximum of 191. The required number will vary according to the convergence tolerance, the amount of ‘noise’ in the consequence model prediction and the proximity of the start point to the MLFP. The results of this and previous unpublished work indicate that, with the software used (Appendix A), the MLFP is usually located within 10–20 N calls, using the BEP as the start point.

5.2. Values of variables at the MLFP

The MLFP search algorithm adjusts the values of the model input quantities until it finds the nearest point on the failure surface to the BEP, with the distance (β) expressed as the combined number of standard deviations. This implies that, other things being equal, the inputs which have the most effect on the result will tend to be adjusted the most. Figs. 5 and 6 contrast the results of these adjustments for two variables: leak hole diameter and relative humidity.

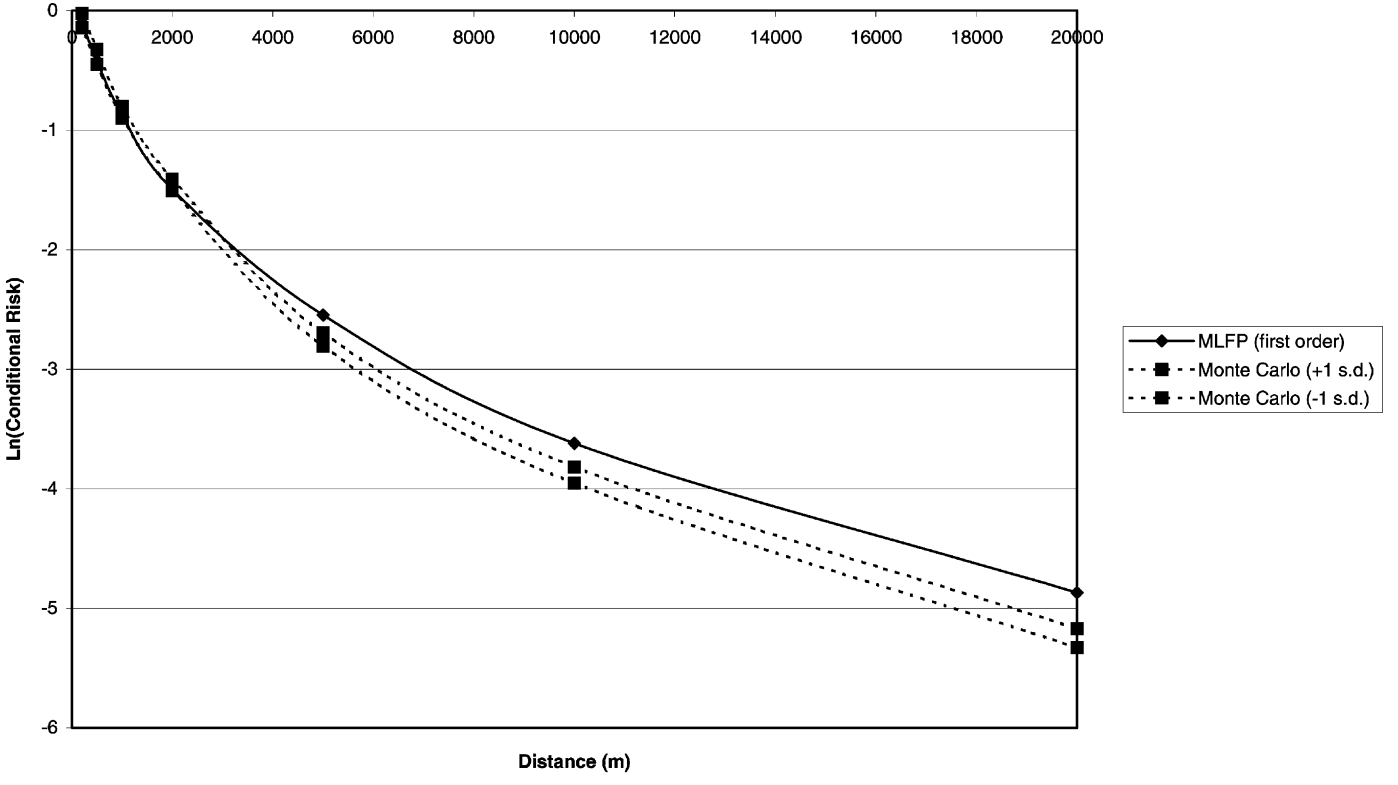

Fig. 4. Comparison of conditional risk R_c as calculated using Monte Carlo method with first-order (FORM) estimate using MLFP method.

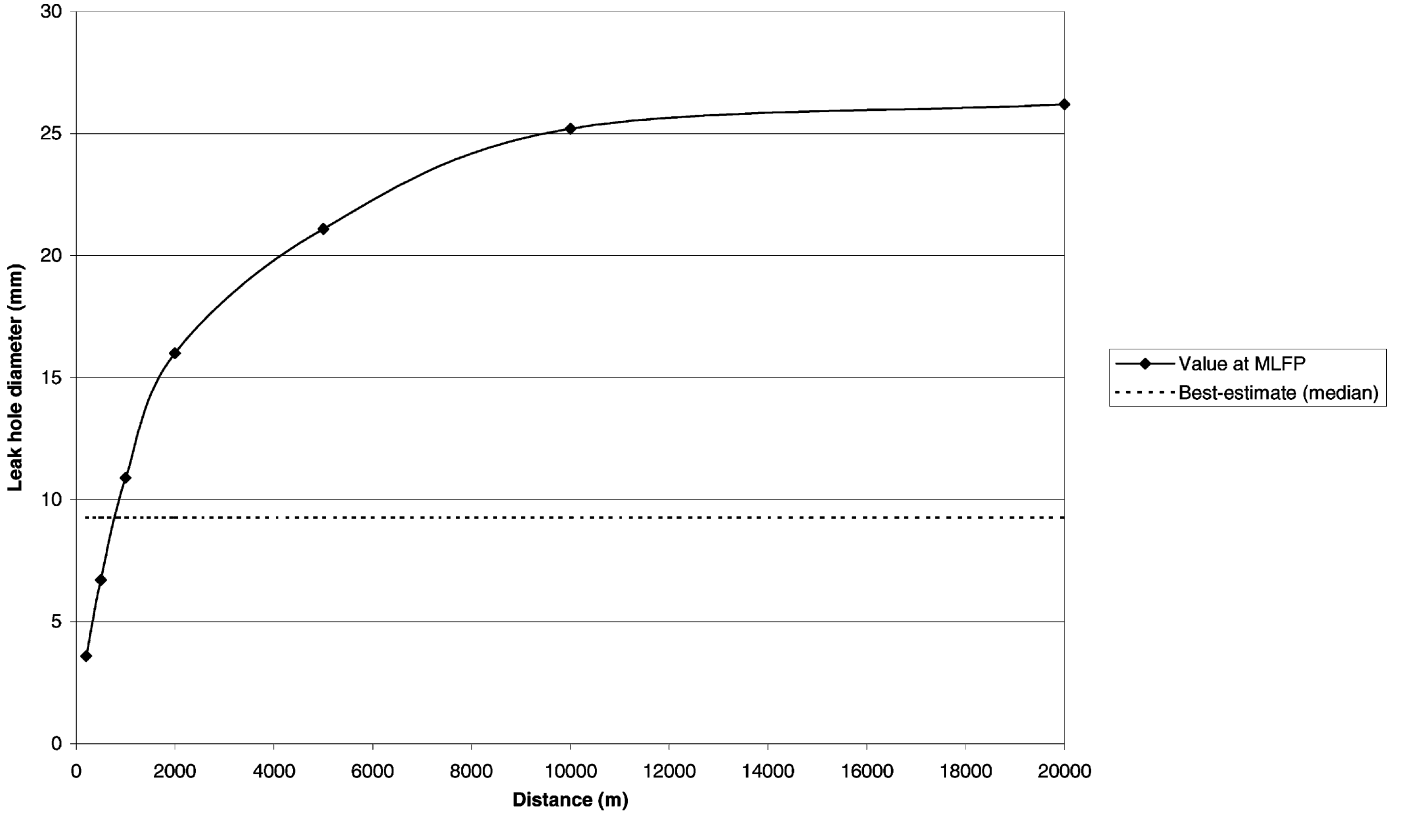

Fig. 5. Variation with distance of most likely failure point value of leak hole diameter.

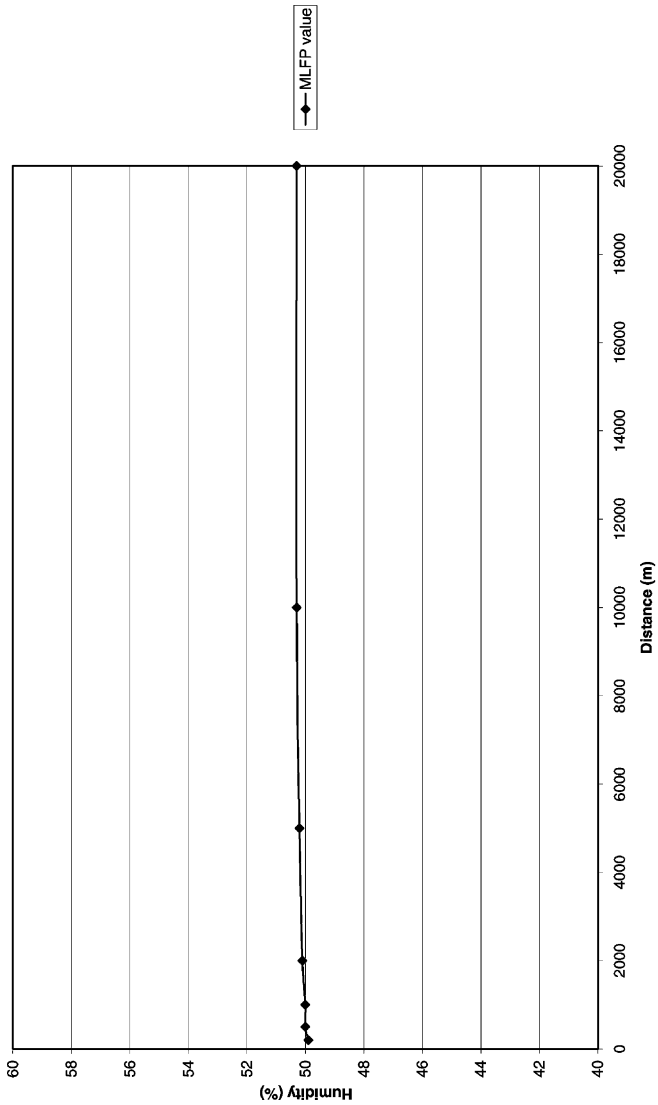

Fig. 6. Variation with distance of most likely failure point value of humidity.

As would be expected, the leak hole diameter has a strong influence on the toxic dose and this is reflected in the behaviour shown in Fig. 5. At distances of 200 and 500 m the BEP is in the failure region (i.e. the median values produce a value of $f_D > 0.005$) so the MLFP method must take steps to reduce f_D in order to find the failure surface. This is evident in Fig. 5, where the MLFP value of δ has been reduced below the median value of 9.27 mm. As the distance increases, the value of the leak hole diameter is progressively increased. The tendency to level out above 10,000 m is discussed in Section 5.3.

The large variation of leak hole diameter shown in Fig. 5 may be contrasted with the minimal variation of humidity shown in Fig. 6. This quantity has almost no effect on the toxic dose: the algorithm recognises this and produces no adjustment. One would expect, therefore, that the humidity could be represented by its mean value (i.e. 50%) without affecting the calculation. That is this is indeed the case demonstrated quantitatively below.

5.3. Importances

The net adjustments produced for all the variables are displayed in standardised form in Fig. 7, which shows the variation of the importances I_i (Eq. (14)) as a function of distance from the source. It is evident that these importances differ from variable to variable by orders of magnitude, so that an approximate method for estimating sensitivities is quite adequate for the screening procedure used for determining which variables are significant and which are not (Section 3.3.2).

At short distances, the most important variable is clearly the leak hole diameter δ, with that of the mean toxic tolerance T being about one half that of δ. The latter indicates that the calculated risk is being influenced by the significant possibility (to be quantified below) that the exposed population are actually more vulnerable than the best-estimate value of T would suggest.

The other variables are much less important, though the procedure suggested in Section 3.3.2 implies that the user might wish to assess the sensitivity to roughness length.

As d increases, the features of note in Fig. 7 are the rise in the importances of the wind speed and the inventory. A feature of the GASTAR plume transport model under the assumed conditions is a rapid rise in predicted airborne concentration with reductions in wind speed, giving rise to the increasing importance of this quantity as distance increases. The behaviour suggests that the low wind speed tail of the adopted distribution is extremely significant and that the user should confirm that the wind profile at the site in question is correctly represented and that the consequence model is valid at the low wind speeds predicted (Section 5.5).

It is noticeable that the importance of the inventory is zero for d up to 2000 m and non-zero for 5000 m and above. At 2000 m, the MLFP value of hole size produces a leak rate of about 3.6 kg s^{-1}, which is insufficient to empty the tank of its median value contents (10 t) within 20 min (Section 4.1.1). Above 5000 m however, the MLFP leak rate will empty the tank, so the inventory begins to influence the result (though not dramatically). Further increases in leak hole diameter empty the tank earlier and hence only weakly affect the toxic dose. This is the reason why, in Fig. 5, the MLFP value of leak hole diameter does not increase much for d above about 10,000 m.

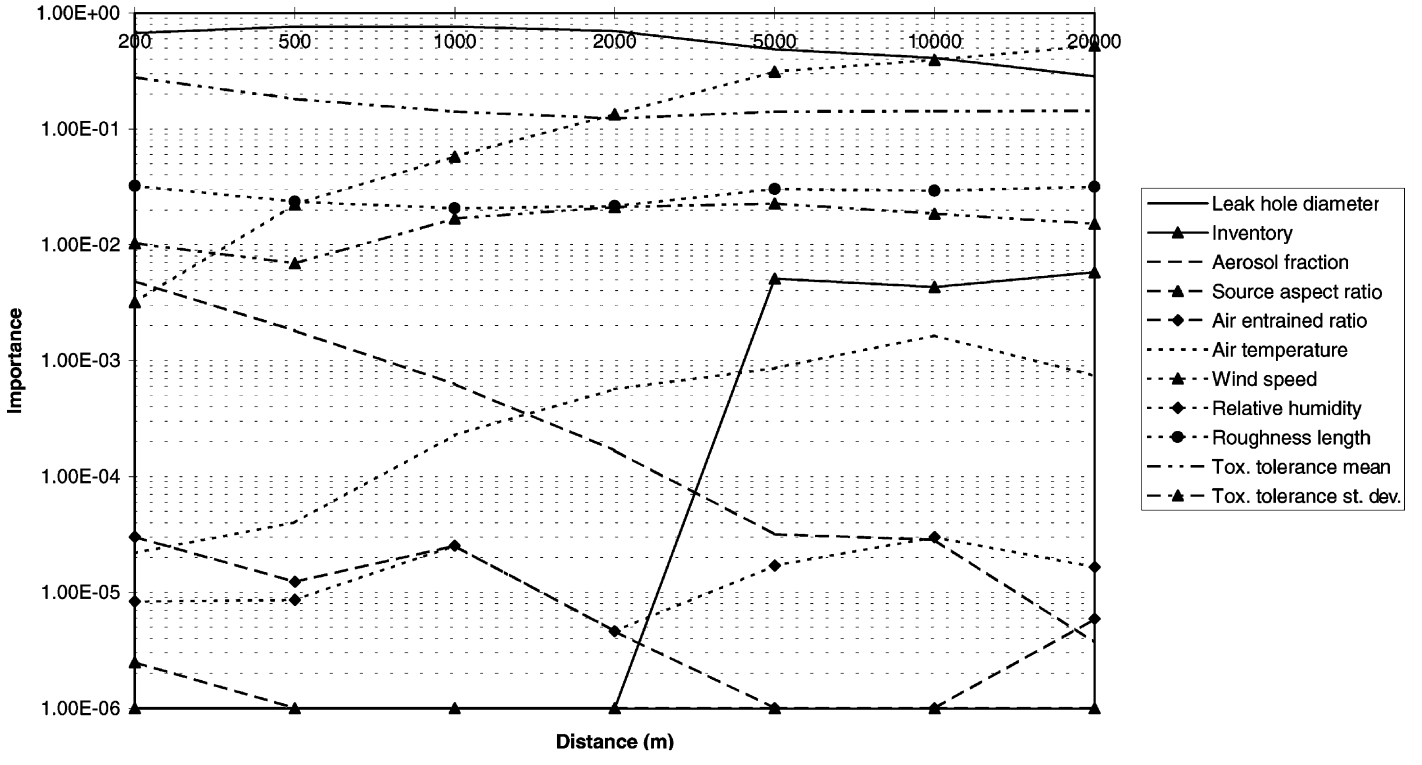

Fig. 7. Variation with distance of Importance I_i for uncertain variables.

Table 2
Sensitivity of the calculated risk: comparison of the predictions of the sensitivity formulas with the results of re-runs of the MLFP calculation

Quantity replaced by median value	Distance (m) = 200; ($P_{f1} = 9.06149 \times 10^{-1}$)		Distance (m) = 20,000; ($P_{f1} = 7.6836 \times 10^{-3}$)	
	Change in P_{f1} from base case	ΔP_{f1} as predicted by Eq. (17)	Change in P_{f1} from base case	ΔP_{f1} as predicted by Eq. (17)
Tank inventory	0.0	0.0	$<\lvert 2 \times 10^{-4}\rvert$	-1.66×10^{-4}
Droplet fraction	5.22×10^{-4}	5.34×10^{-4}	$<\lvert 2 \times 10^{-4}\rvert$	-1.41×10^{-8}
Air entrainment multiple	$<\lvert 2 \times 10^{-4}\rvert$	3.31×10^{-6}	$<\lvert 2 \times 10^{-4}\rvert$	-2.92×10^{-7}
Aspect ratio of source	$<\lvert 2 \times 10^{-4}\rvert$	2.74×10^{-7}	$<\lvert 2 \times 10^{-4}\rvert$	0.0
Air temperature	$<\lvert 2 \times 10^{-4}\rvert$	2.40×10^{-6}	$<\lvert 2 \times 10^{-4}\rvert$	-2.04×10^{-5}
Humidity	$<\lvert 2 \times 10^{-4}\rvert$	9.15×10^{-7}	$<\lvert 2 \times 10^{-4}\rvert$	-1.26×10^{-6}
Roughness length	2.65×10^{-3}	3.54×10^{-3}	-6.4×10^{-4}	-1.09×10^{-3}
Toxicity tolerance mean	5.32×10^{-2}	3.06×10^{-2}	-2.2×10^{-3}	-3.43×10^{-3}
Toxicity tolerance standard deviation	1.63×10^{-3}	1.13×10^{-3}	$<\lvert 2 \times 10^{-4}\rvert$	-5.54×10^{-4}

5.4. Sensitivity formulas

The use of the sensitivity formulas (Section 3.3.2) is illustrated in the results of Table 2. The two extreme distances; 200 and 20,000 m, have been selected to assess the predictions of the sensitivity formulas for those variables whose values are considered 'uncertain' and which are therefore represented by symmetric distributions (in the case of roughness length z_0, it is the distribution of $\ln(z_0)$ which is symmetric). The predictions of the formula were tested by re-running the MLFP calculation with the standard deviation set to a very low value. The MLFP search convergence tolerance (which is constrained by the noise in the value returned by the consequence model) is here equivalent to an uncertainty in P_{f1} of about 2×10^{-4}, so changes in P_{f1} below this value will not be accurate and have not been included explicitly.

It is evident from the table that in all cases except one, Eq. (17) provides a good prediction of the sensitivity of R_c to the process of replacing the distribution by its median value. In the aberrant case (toxicity tolerance standard deviation at 20,000 m), it is likely that the convergence tolerance is producing two similar values of P_{f1}, giving rise to a much smaller value of ΔP_{f1} than the formula predicts. Nevertheless, in all cases, Eq. (17) provides a reliable guide as to whether the uncertainty in the input quantity is significant or not.

The conclusion to be drawn from the values of ΔP_{f1} is clearly as follows. At 200 m, none of the uncertainties has any significant effect on the calculated risk: a risk calculation which assumes median values for these quantities will be satisfactory. At 20,000 m however, a risk calculation which assumes a best-estimate value of T may significantly underestimate the risk by neglecting the possibility that the population is actually more vulnerable than assumed. In the light of this information, the user would perhaps seek to obtain more specific information about the population downwind of the site under consideration. Furthermore, the user might wish to reassess the values used for roughness length, since this is seen to have some effect on the result.

5.5. Consequence model validity

Many consequence models are valid only over certain ranges of their inputs. When these models are used within conventional risk analysis methods (Section 3) the input values may stray outside these ranges. What criterion should be used to decide if the model is valid for the analysis as a whole? The MLFP method provides a clear answer: the model is valid for the analysis if it is valid at the most likely failure point. Nevertheless, if the MLFP turns out to be outside the valid range, this does not necessarily imply that the results cannot be used, so long as it can be demonstrated that the error is small or that the results are pessimistic. However, this will require the use of ad hoc arguments which are outside the scope of this paper.

6. Conclusions

The most likely failure point (MLFP) method, developed within the field of structural reliability analysis (where it is known as the FORM/SORM method) is a technique for estimating the risk (probability) that a calculated quantity Q exceeds a set limit Q_{lim} when some or all of the inputs to the calculation are uncertain.

The method works by first locating the most probable set of input values which produce the value Q_{lim} (i.e. the most likely failure point). Once the MLFP is found, its co-ordinates provide a simple means for estimating the value of the risk. In addition, the use of the very simple sensitivity formulas that the method provides, together with a screening procedure of the type proposed, readily identify the uncertainties to which the risk estimate is most sensitive.

Although the MLFP method can be used as an efficient stand-alone technique for the estimation of risk, for application within the field of toxic hazards it is proposed as a means for performing sensitivity analyses, possibly in parallel with a risk calculation carried out by conventional methods.

The use of the method and of the screening procedure has been demonstrated by means of example calculations of the risk arising from an installation containing chlorine. The calculation employs a commercially available stand-alone dense gas dispersion model. The results demonstrate that:

- the MLFP can be located in typically 10–20N executions of the consequence model, where N is the number of uncertain input variables;
- once the MLFP is found, an estimate of risk can be obtained by a simple formula and is sufficiently close to an ‘exact’ Monte Carlo estimate to be used as the basis of the sensitivity analysis;
- the proposed screening procedure and the sensitivity formula are easy, even trivial, to apply and provide a reliable guide to the sensitivity of the risk to the uncertainties in the input quantities;

The identification of a single set of input values containing sufficient information to summarise (at least approximately) the entire risk analysis is considered to be an important feature of the method. On this basis, it is proposed that if the MLFP is within the range of validity of the consequence model, the model can be considered to be valid for the whole analysis.

The most likely failure point method is therefore a generalised, efficient mathematical technique for certain types of risk calculation. Not only does it provide an estimate of risk, but it comes with a built-in sensitivity study. It deserves to be more widely known.

Acknowledgements

The above work was supported by a series of Research Contracts from the UK Health and Safety Executive.

Appendix A. Practicalities of performing an MLFP calculation

The execution of an MLFP calculation requires the use of specialised software. This software needs to include:

- a menu of available probability distributions,
- means for performing the transformation from standard normal to physical variables,
- a procedure for passing the physical variables to the consequence model and for obtaining in return the calculated value of the quantity of interest Q,
- a method for finding the most likely failure point using the above information,
- methods for calculating risk and sensitivity estimates.

This type of software is available commercially (e.g. [10]). However, the calculations described in the main part of the paper were performed using purpose-written experimental MLFP software written in Fortran and known as FARSIDE. The following discusses the specific implementation of the consequence function calls in FARSIDE, but the principles can be applied in general.

FARSIDE contains a statement $Q = \mathrm{CON}(N, X)$ where $\mathrm{CON}(N, X)$ is a Fortran function which defines the call to the consequence model. In the argument list, N is the number of uncertain physical variables and X is the array of values of these quantities which are being passed to the model.

The consequence model is assumed to be made up of the following components:

- a pre-processor,
- a stand-alone physical model,
- a post-processor.

The pre- and post-processors are assumed to be ‘shell’ routines which are user-written. The physical model is assumed to represent the main science/engineering content of the consequence model. It is, of course, preferable that the physical model should be available as source code so that it can be compiled and linked as part of the consequence model. However, it is possible that such a model will be available only in executable form. In this general case, the purpose of the shell routines is to:

- convert the values of the physical variables into suitable input quantities for the physical model,

- write a model input data file to disk,
- execute the physical model,
- read the relevant results from the model output disk file(s),
- interpret the results and calculate the value of the quantity Q.

It is likely to be the case that the majority of the time spent in performing the MLFP calculation is spent in running the physical model during the MLFP search. The efficiency with which the software performs the constrained minimisation in order to locate the MLFP will therefore determine the overall execution time. With FARSIDE, which uses an adapted form of the projected gradient method [11], the MLFP is usually located in 10–20N executions of the model, though this can be improved if a good initial guess of the location of the MLFP is available.

One feature which consequence models must possess is that the result (i.e. Q) must be a continuous function of the inputs x_i; $i = 1 \ldots N$. Discontinuities can arise from a number of causes:

- errors in the model,
- the use of different models for different data ranges without 'blending',
- poorly-converged solutions,
- low accuracy of output values.

These features can adversely affect the convergence of the MLFP search. Conversely, poor convergence behaviour can sometimes indicate the presence of a previously unsuspected error in the model.

References

[1] P. Thoft-Christensen, M.J. Baker, Structural Reliability Theory and its Applications, Springer, 1982.
[2] H.O. Madsen, S. Krenk, N.C. Lind, Methods of Structural Safety, Prentice Hall, 1986.
[3] P. Bjerager, On computation methods for structural reliability analysis, Struct. Safety 9 (1990) 79–96.
[4] Y.-G. Zhao, T. Ono, A general procedure for first- and second-order reliability method (FORM/SORM), Struct. Safety 21 (1999) 95–112.
[5] Y.-G. Zhao, and T. Ono, New approximations for SORM, Parts 1 and 2, J. Eng. Mech., Vol. 125(1), 1999, pp. 79–85 (Part 1) and pp. 86–93 (Part 2).
[6] P.E. Gill, W. Murray, M.H. Wright, Practical Optimisation, Academic Press, 1981.
[7] H.O. Madsen, Omission sensitivity factors, Struct. Safety 5 (1988) 35–45.
[8] GASTAR Version 3, User manual, Cambridge Environmental Research Consultants Ltd., Cambridge, UK, 1998.
[9] F.P. Lees, Loss Prevention in the Process Industries, Butterworth Heinemann, 1996.
[10] STRUREL, A Structural Reliability Analysis Program System, RCP GmbH, Munich, See www.strurel.de
[11] Dixon, L.C.W., Non-Linear Optimisation, English Universities Press, 1972.