Calculation of input data importances for toxic release risk assessments: Gastar and Crunch

Prepared by
WS Atkins Consultants Limited
for the Health and Safety Executive

CONTRACT RESEARCH REPORT
338/2001

Calculation of input data importances for toxic release risk assessments: Gastar and Crunch

WS Atkins Consultants Limited
Woodcote Grove
Ashley Road
Epsom
Surrey
KT18 5BW
United Kingdom

The Most Likely Failure Point (MLFP) method, often referred to in the literature as the FORM/SORM method, has been used with two stand-alone consequence models: GASTAR and CRUNCH, to calculate the toxic risk arising from a dense gas release.

A standard procedure associated with complex calculations is the performance of a sensitivity study, in which the values of selected input quantities are altered and the problem re-run. However, this is often unsatisfactory and can be a laborious procedure for a risk analysis, requiring that the full analysis is repeated a number of times.

A feature of the MLFP method is that it provides a built-in sensitivity analysis. This feature has been used to quantify the influence on the calculated risk of uncertainties in the input quantities, for each of the consequence models. A procedure has also been proposed, in which the unimportant quantities can be screened out by inspection, allowing the importance of the remainder to be quantified by means of progressively more accurate methods.

In a related piece of work, investigations were carried out into the best means of representing scatter plots of wind speed and atmospheric stability as smooth distributions. A technique known as the adaptive kernel method seemed to offer the best approach, through further work is required to optimise the representation. This work is reported in an Appendix to the main report.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the author alone and do not necessarily reflect HSE policy.

HSE BOOKS

First published 2001

ISBN 0 7176 2013 1

CONTENTS

Page

TABLES

FIGURES

APPENDICES

1. INTRODUCTION

1.1 BACKGROUND

This Report has been prepared for the Health and Safety Executive (HSE) by WS Atkins Science & Technology under the requirements of Research Contract No. 4040/R72.054. It describes the use of the Most-Likely Failure Point method (often referred to in the literature as the FORM/SORM method) for the calculation of input data importances associated with the calculation of risk arising from dense gas dispersion. The current work extends previous research reported in Refs. 1, 2 and 3.

One of the problems facing any analyst, whether performing a risk analysis or a single consequence calculation, is that of demonstrating that the result of the calculation is not unduly sensitive to uncertainties in the input parameters of the model. This can be a particular problem if the consequence model contains a significant number of input parameters, all of which can be considered uncertain to an extent and whose exact role within the model is not always understood in detail.

It is reasonably common practice 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 is repeated a number of times.

In Ref. 3, it was shown how the Most-Likely Failure Point (MLFP) method could be applied to the Quantitative Risk Assessment (QRA) of an industrial installation containing chlorine. The MLFP method can, in common with other techniques, be used for this type of calculation (such as Monte-Carlo, or the 'summation' (RISKAT-type) method), provide an estimate of overall risk, where risk is defined here as the probability that, at a given location near to an installation containing the toxic material, a hazard index (e.g. concentration, toxic dose or percentage of deaths) exceeds a set limit.

However, in obtaining its result (which it does very efficiently), the MLFP method can also provide useful additional information which tells the analyst by how much the uncertainties in the input data to the risk calculation affect the result (the data 'importances'). This was referred to in Ref. 3 as the 'built-in sensitivity study'.

In this document, the built-in sensitivity study facility of the MLFP method is used in order to calculate importances of the input variables for two consequence models used by HSE for dense gas dispersion calculations. These models are:

– CRUNCH [4]

– GASTAR [5]

The performance of these calculations for a range of conditions allows the identification of variables which are:

(a) always unimportant, so that best-estimate values can be used in future calculations in the knowledge that this will not significantly affect the calculated risk

(b) sometimes important, so that the conditions under which they become important can be identified

(c) always important, so that effort can be directed towards obtaining more precise values for subsequent use.

1.2 OBJECTIVES AND SCOPE OF THE WORK

The main objectives of the work were specified on the Research Contract as follows:

1. to investigate the effect of the consequence models (CRUNCH and GASTAR) on the performance of the MLFP calculation;

2. to calculate data importances for a range of chlorine releases;

The scope of the work is as follows:

1. Amend in-house MLFP software (FARSIDE), including improvements to structure and checks on global convergence (as in Para. 4 of Section 7.5 of Ref. 3)

2. Set up FARSIDE to 'call' CRUNCH by means of writing and reading interface files

3. Select suitable input parameters for CRUNCH, with probability distributions describing uncertainties

4. Repeat (2) and (3) for GASTAR

5. Investigate methods for smoothing measured wind speed / direction / stability distributions and select distributions for use

6. Run initial cases:

   - Compare first-order (FORM) and second-order (SORM) predictions

   - Evaluate importances and sensitivities

   - Compare CRUNCH and GASTAR

7. Investigate the effect of simpler representations of weather distributions on importances

8. Evaluate results, compare models, distinguish between results which apply generally and those which might be expected to be valid for only a range of circumstances. Discuss relative importances of atmospheric and other variables.

1.3 CONTENTS OF THE REPORT

The main part of this report therefore:

- describes the concept of the most likely failure point and how it can be used to summarise (approximately) the results of a risk analysis;

- describes the idea of input data importances and how they arise in a natural and straightforward way from the MLFP method;

- describes how the uncertainty in the risk estimate is related to the uncertainties in the input data of the consequence model, via the importances;

– provides practical guidance on how the information provided by the importances and related sensitivity can be interpreted and used;

– interprets the calculations described above and attempts to draw general conclusions concerning how the input quantities may be categorised as in (a) - (c) Section 1.1.

The additional work in the scope as listed above is presented in the form of appendices. Appendix A presents preliminary documentation of the FARSIDE software used to perform the MLFP calculations described within. The appendix describes the basis of the method used to search for the most likely failure point (Section 2) and summarises the routines which make up the code and the types of probability distributions available to the user.

Appendix B describes the results of investigations into the production of a smooth, interdependent frequency distribution of wind speed and atmospheric stability from a collection of simultaneous measurements of these quantities, obtained over a period of time.

The remaining appendices support the work described in the main part of the report. Appendix C summarises the CRUNCH input quantities (drawing on information in Ref. 4). Appendix D provides a description of the nomenclature used and Appendix E summarises the results of the FARSIDE MLFP calculations reported in Sections 5 and 6.

2. THE MOST LIKELY FAILURE POINT METHOD

The Most Likely Failure Point (MLFP) method was developed within the field of structural reliability, where it is used to predict the probability of structural collapse. It is known in structural reliability literature as the FORM/SORM method (First Order Reliability Method/Second Order Reliability Method).

The method is, however, not tied to its original field and may also be used to estimate the risk arising from a toxic release, where risk is defined here as the probability that, at a given location (the 'dose point') near to an installation containing the toxic material, a hazard index (e.g. concentration, toxic dose or probability of deaths) exceeds a set limit.

This section presents a brief description of the MLFP method. The principles and theory outlined below are treated in detail in previous reports [1-3]. The section on importances and sensitivities (Section 2.3) expands on material in these previous reports.

2.1 RISK ASSESSMENT OF A CHLORINE RELEASE - A SIMPLE EXAMPLE

Figure 1 shows a fictitious but characteristic plot of a predicted toxic dose D (at a selected downwind location) as a function of chlorine release rate c and wind speed v. 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.

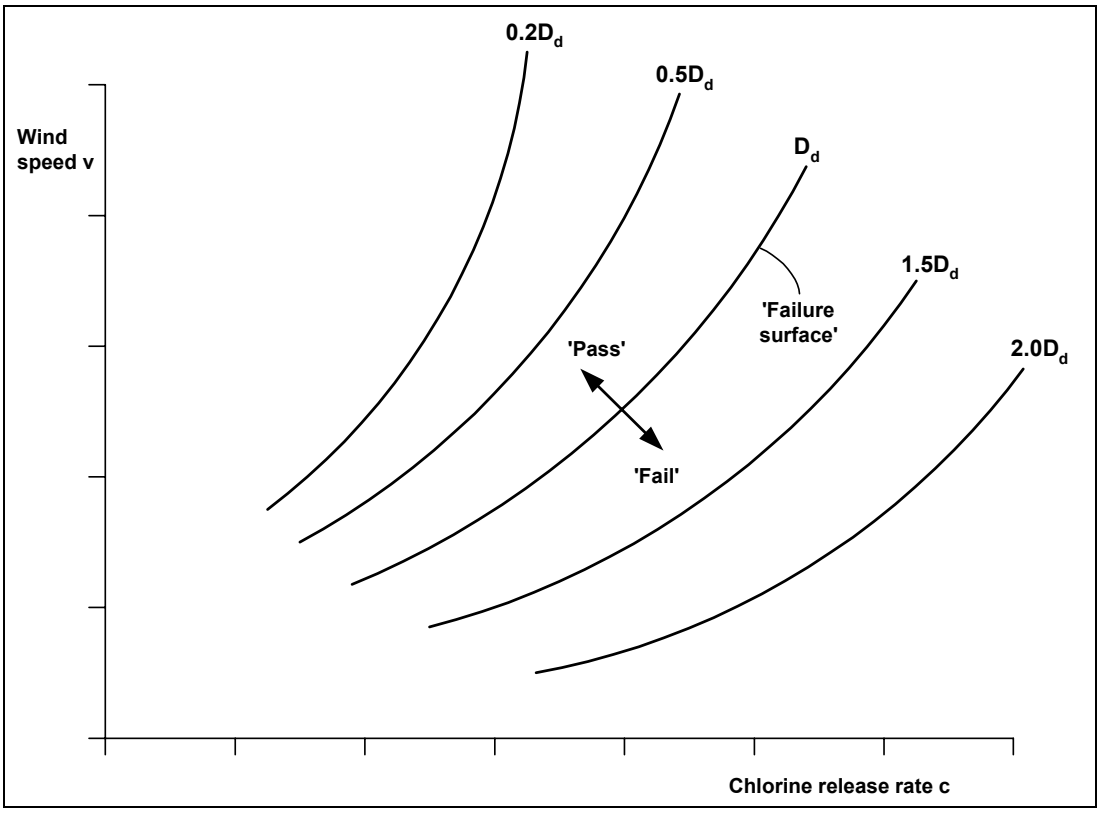

Figure 1
Contour plot of notional toxic dose calculation, showing predicted dose D as a function of chlorine release rate c and wind speed v

The values of the other contours are represented as multiples or fractions of D_d, with those below and to the right of the dangerous dose contour having $D>D_d$ and those above and to the

left having $D<D_d$. The former region is known as the fail or failure region, the latter as the pass region and the line $D=D_d$ as the failure surface. (In the general case (Section 2.3) the quantity being calculated is denoted as Q and the failure surface value is denoted by Q_{lim})

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 at the dose point receives a dose $D>D_d$, then, with reference to Figure 1, this can be visualised as the probability that the combination of c and v (i.e. the actual release rate and the wind speed when the release occurs) will be somewhere in the fail region. In other words, if f(c)dc is the frequency (probability per unit time) of a release of rate between c and c+dc, and p(v)dv is the probability that the wind speed lies between v and v+dv, then the risk R_t (per unit time) is given by

$$R_t = \int_F f(c)p(v)dvdc \tag{2.1}$$

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

$$f(c) = f_T p(c) \tag{2.2}$$

where

f_T is the frequency (per unit time) of a release of any size

p(c) is the probability, given that a release has occurred, that the rate is between c and c+dc

then

$$R_t = f_T \int_F p(c)p(v)dvdc \tag{2.3}$$

$$= f_T R_c \tag{2.4}$$

where

$$R_c = \int_F p_c(c)p_v(v)dvdc \tag{2.5}$$

is the conditional risk predicted by the calculation. In what follows, the term 'risk' always refers to conditional risk R_c, unless stated otherwise.

2.2 COMPARISON OF RISK ASSESSMENT METHODS

The MLFP method, in common with other methods for QRA, (specifically 'summation' (RISKAT type) and Monte Carlo methods) 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, reflecting the user's uncertainty as to the most appropriate value to use (e.g. roughness length).

In terms of the simple example introduced in Section 2.1, 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.

2.2.1 Summation (RISKAT) and Monte Carlo methods

The RISKAT method (broadly speaking) divides the c and v axes into a number of intervals, thereby subdividing the (c, v) plane into rectangles ('cells'). Let cell (i, j) be the cell formed by the intersection of the i^{th} interval on the c axis and the j^{th} interval on the v axis, where i runs from 1 to I and j from 1 to J so that there are IJ cells altogether. Let p_{ij} be the probability that the values of c and v fall within that cell. Then, the principle of the RISKAT method is to visit each cell in turn and perform a consequence model calculation for a representative pair of values (c_i, v_j) (say, at the centre of the cell). If the calculated value $D(c_i,v_j)>D_d$, then p_{ij} is added to an accumulator. The final summed value of all the probabilities for cells where $D>D_d$ is an approximate estimate of R_c.

The Monte Carlo method resembles the RISKAT method in that it accumulates an estimate of R_c by performing calculations over the (c,v) plane, but it differs in the means which it uses to do so. In this case, values of c and v are chosen randomly from their probability distributions and a consequence model calculation is performed at each point (c, v) so chosen. The method then simply counts the number of times that D exceeds D_d and divides by the number of calculations performed, to obtain its approximate value 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. For the summation method however, an extra variable implies an extra dimension: the total of IJ 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.

2.2.2 The most likely failure point method

The MLFP method works on an entirely different principle from those techniques described in Section 2.2.1. The method works by finding the distance to the failure surface from the best-estimate values of the variables, with this distance being expressed as the combined number of standard deviations.

This is achieved as follows. Assume that an input quantity x to the consequence model is uncertain. This uncertainty is represented by a 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:

$$\Phi(u_x) = P_x(x) \tag{2.6}$$

where

$$\Phi(u) = 0.5\left(1 + \mathrm{erf}\left(u/\sqrt{2}\right)\right) \tag{2.7}$$

is the cumulative standard normal distribution; that is: 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.

$$\phi(u) = \exp\left(-u^2/2\right)/\sqrt{2\pi} \tag{2.8}$$

The 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 of Section 2.1, this is the point where

$$\begin{aligned} P_c(c) &= 0.5 \\ P_v(v) &= 0.5 \end{aligned} \tag{2.9}$$

Assume that this point (the best-estimate point, or BEP) lies somewhere in the pass region. From (2.6) and (2.7), the point corresponds to

$$\begin{aligned} u_c &= 0 \\ u_v &= 0 \end{aligned} \tag{2.10}$$

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

The principle of the MLFP method is now to find, by a process of iteration (Appendix A), the point of closest approach of the failure surface to the BEP, as illustrated in Figure 2. Since each of the transformed variables now has the same (standard normal) probability distribution, the combined distribution is a function 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, and is therefore known as the most likely failure point, or MLFP. It is straightforward to show that if the distance (essentially the combined number of standard deviations) from the BEP to the MLFP is β, then an approximation to the conditional risk R_c is given by the first order fail probability P_{f1}, where

$$P_{f1} = \Phi(-\beta) \tag{2.11}$$

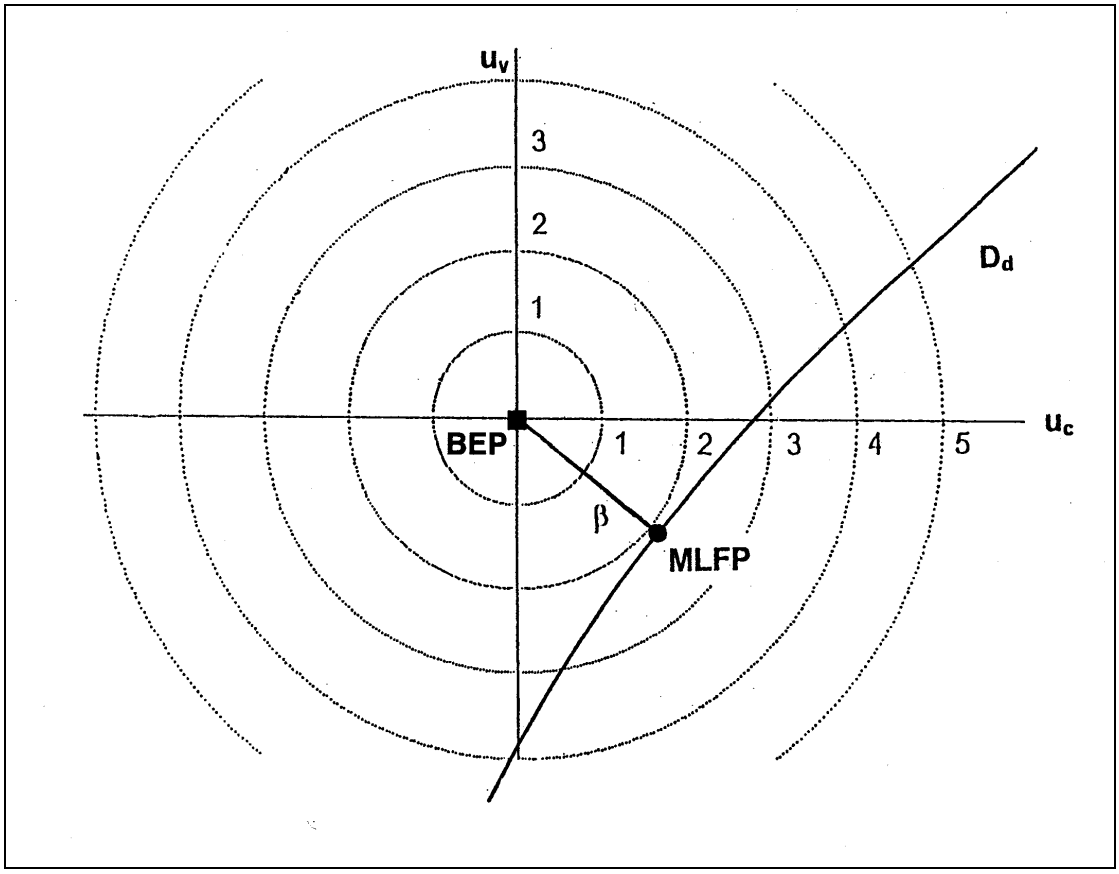

Figure 2
Location of most likely failure point: dose contour plot in standard normal coordinates (schematic)

The fundamental difference between the MLFP method and the RISKAT and Monte Carlo methods is therefore that, whereas the latter calculate R_c by numerical integration, the former gives an *analytic* approximation for R_c in the very simple form of Equation (2.11). Therefore, a problem (the risk analysis) which at first sight appears to require information about the behaviour of D over the whole (c,v) plane in order to be solved, can actually be expressed (as least approximately) by just the values of u_c and u_v (and hence c and v) which together define the MLFP.

It may be observed that, in the above discussion, it has been assumed implicitly that D is a smooth function of c and v. This is a requirement of the MLFP method, but one which is not always realised in consequence models, as will be seen later.

The expression (2.11) is the first order, 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 (SORM) approximation, which uses the curvature(s) of the failure surface to obtain an improved approximation to R_c, can be obtained, and has been discussed extensively in previous work. However, the purpose of the present work is 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 other methods described above. The purpose of this work is to estimate the quantities (if any) to which R_c is particularly sensitive. For this purpose, the first order estimate is sufficient, as described below.

2.3 IMPORTANCES AND SENSITIVITIES

Because R_c is estimated in the first-order MLFP approximation by an algebraic expression, (Equation 2.11) it is possible to use the analytic properties of this expression in order to investigate the sensitivity of R_c to a range of parameters. Once the location of the MLFP has been found, these subsidiary calculations are straightforward, as described below.

Assume the general case in which there are N uncertain input variables x_i; i=1…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 co-ordinate system are u_{im}; i=1…N.

First, for each uncertain variable, let

$$I_i = \left(\frac{u_{im}}{\beta} \right)^2 \tag{2.12}$$

I_i is termed the 'importance' of variable i (where, when referring to variable i, we also include the probability distribution assigned to variable i) to the risk calculation, for reasons which will become apparent below. From Pythagoras's theorem, it is obvious that

$$\sum I_i = 1 \tag{2.13}$$

Assume now that model variable i has a probability distribution characterised by mean value μ_i and standard deviation σ_i. It is shown in [2] that, under certain assumptions (see below), the sensitivities of β to variations in μ_i and in σ_i are as follows:

$$\left(\frac{\partial \beta}{\partial \mu_i} \right) = \pm \frac{I_i^{1/2}}{\sigma_i} \tag{2.14}$$

$$\left(\frac{\partial \beta}{\partial \sigma_i} \right) = -\frac{\beta I_i}{\sigma_i} \tag{2.15}$$

The sign in Equation (2.14) depends upon whether a positive change in the mean value moves the failure surface further away from (+) and or closer to (-) the BEP. The negative sign in Equation (2.15) arises because an increase in the standard deviation always moves the failure surface closer.

Furthermore as shown in [2], if variable i is replaced by its median value, this causes an increase in β given by:

$$\Delta\beta = \frac{\beta I_i}{2} \tag{2.16}$$

The above expressions were calculated on the assumptions that:

– the probability distribution of variable i is Gaussian, and

– the failure surface is represented by a tangent plane at the MLFP (i.e. the FORM approximation).

Despite the fact that these assumptions appear fairly restrictive, numerical experiments carried out in [2] and [3] indicate that, even for other types of symmetric distribution, these expressions are perfectly adequate as a guide to the sensitivity of the calculated risk (see below). That is to say: they are usually accurate to within a factor of about 2, whereas the sensitivities of the variables may differ by several orders of magnitude. As will be seen below, it is possible to determine easily, often simply by inspection, which uncertain variables are contributing strongly to risk and which can be replaced by constants without introducing significant error.

2.3.1 Useful versions of the sensitivity formulas

A particularly useful measure of sensitivity is that expressing the fractional change in risk arising from a given perturbation to the probability distribution of a given variable. Since (Equation (2.6))

$$P_{f1} = \Phi(-\beta) \tag{2.17}$$

then

$$\frac{dP_{f1}}{d\beta} = -\phi(\beta) \tag{2.18}$$

Substituting (2.17) and (2.18) into (2.14) to (2.16) and replacing the infinitesimals by finite changes, yields respectively:

$$\frac{\Delta P_{f1}}{P_{f1}} = \pm\Psi(\beta)\left(\frac{\Delta\mu_i}{\sigma_i}\right)I_i^{½} \tag{2.19}$$

$$\frac{\Delta P_{f1}}{P_{f1}} = \Psi(\beta)\left(\frac{\Delta\sigma_i}{\sigma_i}\right)\beta I_i \tag{2.20}$$

for small changes to μ_i and σ_i respectively and, for variable i replaced by its median value:

$$\frac{\Delta P_{f1}}{P_{f1}} = -\Psi(\beta)\frac{\beta I_i}{2} \tag{2.21}$$

where $\Delta\mu_i$ and $\Delta\sigma_i$ are the assumed changes to the mean and standard deviation of variable i and

$$\Psi(\beta) = \frac{\phi(\beta)}{\Phi(-\beta)} \tag{2.22}$$

Now, the function Ψ has a particularly simple form, having a value of 0.8 at β=0 and tending to $\Psi(\beta) = \beta$ as $\beta \to \infty$. For $\beta <\approx 2$, it is therefore possible to assume

$$\Psi(\beta) \approx 1 + \frac{\beta}{2} \tag{2.23}$$

and for $\beta \geq 2$

$$\Psi(\beta) \approx \beta \tag{2.24}$$

with errors of no greater than 20% (at $\beta=0$) which are not significant when used in (2.19) - (2.21).

Therefore, the fractional change in P_{fl} due to the replacement of the distribution by its median value (Eq. (2.21) is a reasonable measure of the contribution of a variable's probability distribution to the calculated risk. It may be seen that this measure is proportional to the quantity I_i, defined in Equation (2.12). Therefore, as stated above, I_i is a measure of the importance of the variable (together with its probability distribution) in terms of its contribution to the calculated risk.

2.3.2 Origin of the importances

It is worthwhile indicating how the importance of a variable arises as a result of its probability distribution and the properties of the consequence model. Consider again Figure 2, but now consider the general case where there are N uncertain variables, so that the failure surface is (N-1) dimensional. Since the failure surface is a surface of constant Q then, at the MLFP, the gradient $\nabla Q = \left[\partial Q/\partial u_1, \partial Q/\partial u_2, \ldots, \partial Q/\partial u_N\right]$ at the MLFP is aligned with the vector $\underline{u}_m = \left[u_{1m}, u_{2m}, \ldots, u_{Nm}\right]$. In other words:

$$\left(\frac{\partial Q}{\partial u_i}\right)_{MLFP} = \lambda u_{im} \qquad i = 1 \ldots N \tag{2.25}$$

where

λ is a constant

Now, assume that the uncertain variable x_i is represented by a normal distribution with mean μ_i and standard deviation σ_i. In this case

$$u_i = \frac{x_i - \mu_i}{\sigma_i} \tag{2.26}$$

so

$$\frac{\partial u_i}{\partial x_i} = \frac{1}{\sigma_i} \tag{2.27}$$

and since

$$\left(\frac{\partial Q}{\partial u_i}\right) = \left(\frac{\partial Q_i}{\partial x_i}\right)\left(\frac{\partial x_i}{\partial u_i}\right) \tag{2.28}$$

then (2.25) yields

$$\lambda u_{im} = \sigma_i \left(\frac{\partial Q}{\partial x_i} \right)_{MLFP} \tag{2.29}$$

so

$$I_i \propto \left[\sigma_i \left(\frac{\partial Q}{\partial x_i} \right) \right]^2_{MLFP} \tag{2.30}$$

giving the intuitively appealing result that the importance of a variable x_i to the risk analysis is proportional to (the square of) the product of its uncertainty (i.e. its standard deviation) and the sensitivity of the prediction of the consequence model to changes in x_i, evaluated at the MLFP.

2.4 USE OF THE IMPORTANCE AND SENSITIVITY FORMULAS

Section 2.3 above presented the MLFP importance and sensitivity formulas in various alternative forms. This section presents a number of recommendations as to how they may be used in practice.

It should be stressed that both the first-order estimate of risk (P_{fl}) and the sensitivity formulas are approximations. The sensitivity formulas, for example, tend to predict sensitivity to within about a factor of 2. They should therefore primarily be used as a guide and can only be used to support definite conclusions as to the significance of an uncertain variable if their predictions fall well away from borderline values. Fortunately, the very large range of importances means that this is often the case.

Based on the formulas derived above and the numerical experiments carried out in previous phases of the work [2,3], the following procedure is suggested as appropriate and prudent for the interpretation and use of the importance and sensitivity information provided by an MLFP calculation:

1. Inspect the values of I_i for each variable.

2. If $I_i \leq 5 \times 10^{-3}$, the uncertainty in the input quantity i is of no significance. It can be replaced by a suitable fixed value (i.e. the mean or median) without significantly affecting the risk.

3. If $5 \times 10^{-3} \leq I_i \leq 5 \times 10^{-2}$, the uncertainty in the variable may be of significance. It would be advisable to use one of the sensitivity formulas (2.19 - 2.21) to determine whether a change in the value of the quantity would have a significant effect on the calculated risk.

4. If $I_i \geq 5 \times 10^{-2}$, the uncertainty in the input quantity is having a significant effect on the risk. If this is a genuinely variable quantity (release rate or atmospheric value) this is what would be expected. Nevertheless, it would be reasonable to reconsider the assumptions used in the derivation of the frequency distributions for these quantities and to use one of the formulas to check the degree of sensitivity of the risk prediction.

If the input quantity is merely uncertain rather than genuinely variable, then the value of risk is being strongly influenced by this uncertainty. This is not a desirable state of affairs, since it means that (for example) any calculations of risk based on a best-estimate value of this quantity may be a significant underestimate.

The use of this suggested approach may be seen in the remainder of the report.

3. USE OF STAND ALONE MODELS CRUNCH AND GASTAR

3.1 OVERVIEW OF THE MODELS

CRUNCH [4] and GASTAR [5] are computer codes which predict the transport of dense gas. CRUNCH is relatively old, having been developed in the early 1980s. It uses a 'box' model representation and treats continuous releases only. GASTAR is also a box model but is fairly recent and can treat a range of source types. It also has facilities for the representation of gas flow in sloping terrain and around obstacles.

The facilities of each code which are represented in the MLFP calculations are discussed in Sections 4, 5 and 6. The remainder of this section describes the way in which each code was used as a consequence model in the manner required by FARSIDE.

3.2 USE OF STAND ALONE PROGRAMS AS CONSEQUENCE MODEL FUNCTIONS

The FARSIDE program (Appendix A) contains a function CON(N,X) which represents the value of the quantity Q (i.e. the prediction of the consequence model) as a function of the input parameters. In the above:

N is the number of uncertain variables

X is an array, containing the current values for each of the N uncertain variables. Each of these variables will previously have been assigned a probability distribution (see Section 2).

The general procedure carried out by FARSIDE in order to calculate a value of Q from any consequence model is as follows:

1. Produce new values u_i, i=1….N, where these are decided by the search algorithms (Appendix A).

2. Convert to model input quantities x_i, i=1…N (using Equation (2.6))

3. Run consequence model Q = CON(N,X)

4. Return the value of Q to FARSIDE.

If the consequence model is available as FORTRAN source code, then it is straightforward to represent the model as a subroutine. Furthermore, it is usually possible to identify calculation and output routines and tailor the result (i.e. the quantity calculated) to be suitable for FARSIDE. When the consequence model is (as here) available only as a stand-alone executable file, the function CON must itself produce the following sequence of actions:

1. (If necessary) pre-process the values of x_i to produce suitable input values for the stand-alone model.

2. Set default values for any fixed input parameters.

3. Write a disk file in the form of a standard input file for the stand-alone model.

4. Execute the stand-alone model.

5. Read the 'result' from the standard output file of the model (and include facilities for coping with failure of the model if necessary).

6. (If necessary) postprocess the result to produce the value of the required quantity Q.

There are two requirements for the above procedure to be possible:

(a) The FORTRAN compiler used for the calling program (i.e. FARSIDE) must allow a return to the calling program once the consequence model has finished executing.

(b) It must be possible for the model to be run in batch mode.

The compiler used in the previous phases of this work (Salford Fortran) did not allow (a) above to take place. However, Digital Fortran [6] does allow this process via the SYSTEMQQ command and therefore was used for the current work.

The following sections give details respectively of how CRUNCH and GASTAR were accessed in this way.

3.3 CRUNCH

3.3.1 Input file creation and CRUNCH execution

CRUNCH reads input data from a short file 'Runname.CRI' and produces two output files, 'Runname.CRO' and 'Runname.FCO', where the former is a condensed version of the latter. The format of the .CRI file was taken mainly from Ref. 4, supplemented by inspection of the correspondence between values in the .CRI and .FCO files. A summary of the input specification as inferred for the version used in this work is given in Appendix C.

3.3.2 Extraction of the value of Q from the CRUNCH output

The full output (.FCO) file written by CRUNCH is a formatted file in the old Fortran line printer style. It contains three sections.

1. a summary of the input data

2. a table of concentration vs. distance, applying to the range where the concentration of released material is such that the mixture of the material and the entrained air can be considered as a dense vapour.

3. a table of concentration vs distance applying to the range beyond that of 2) above where the concentration has reduced to the level where the mixture is represented as a neutrally buoyant (passive) gas.

There are two particular problems presented by the data in the .FCO file:

– the number of significant figures in the table entries is small (usually 3, but as low as 2).

– there is a significant discontinuity between the concentration profiles in the dense and passive tables.

The first of these gives rise to a small amount of noise in the value of Q, but the influence of this can be reduced by a suitable choice of the iteration parameters in FARSIDE (see Appendix A). The second is of greater consequence since the changeover from dense to neutrally buoyant (i.e. passive) tends to occur at concentrations comparable with those giving rise to around 50% fatality in around 20 minutes exposure (i.e. 300-1000 ppm). This is the sort of range from which one might typically select Q_{lim}. This means that FARSIDE will be attempting to find the most likely failure point in a region where Q_{lim} is subject to large (~30%) discontinuities.

The nature of the discontinuity can be seen in Figure 3, which shows a typical variation in predicted concentration at the dose point (1000 m from the source) with hole diameter. Evident in the curve is a 'staircase' appearance, plus two definite discontinuities: one at 21.5 mm and one at about 26 mm.

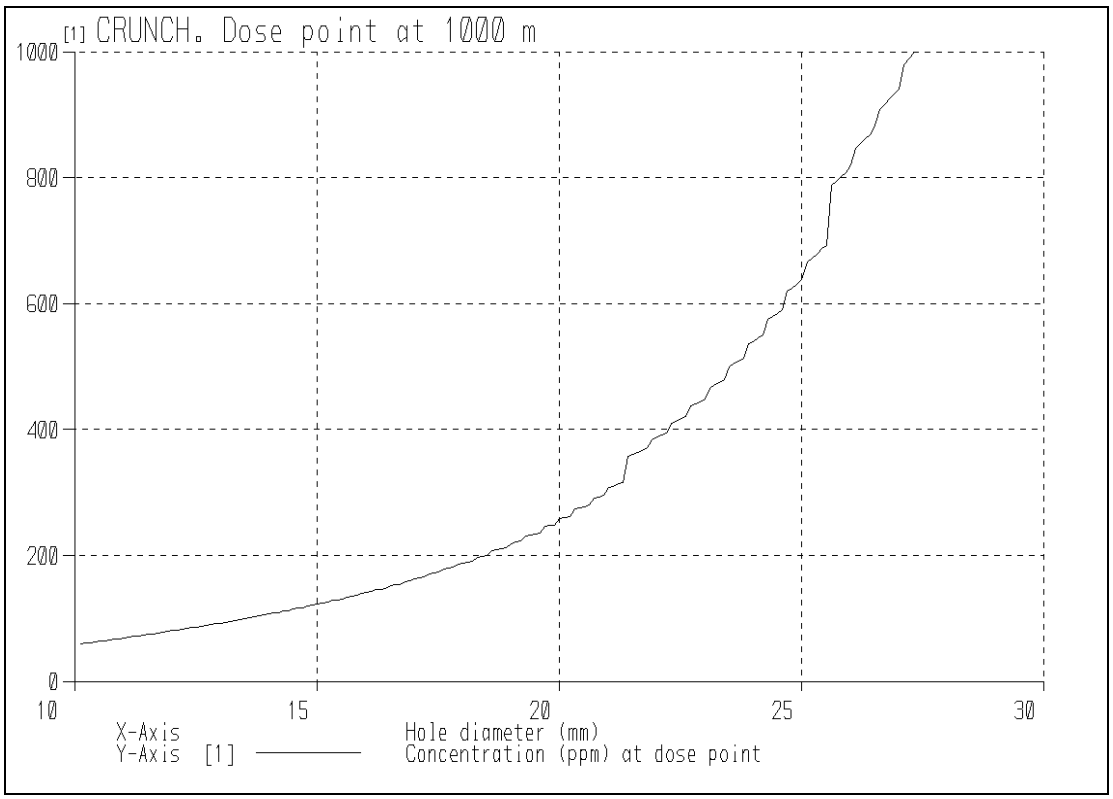

Figure 3
Concentration as a function of leak hole diameter at dose point 1000m from source: calculated by interpolation of CRUNCH results

The reason for this behaviour may be seen in Figure 4, which shows the variation of concentration with distance for two hole diameters on either side of the discontinuity at 21.5 mm. The release rates in the two cases differ by about ½% (6.34 kgs^{-1} and 6.37 kgs^{-1} respectively), so the two curves are indistinguishable. The cause of the discontinuous behaviour therefore lies not in the prediction of the model, but in the locations of the output points selected by CRUNCH.

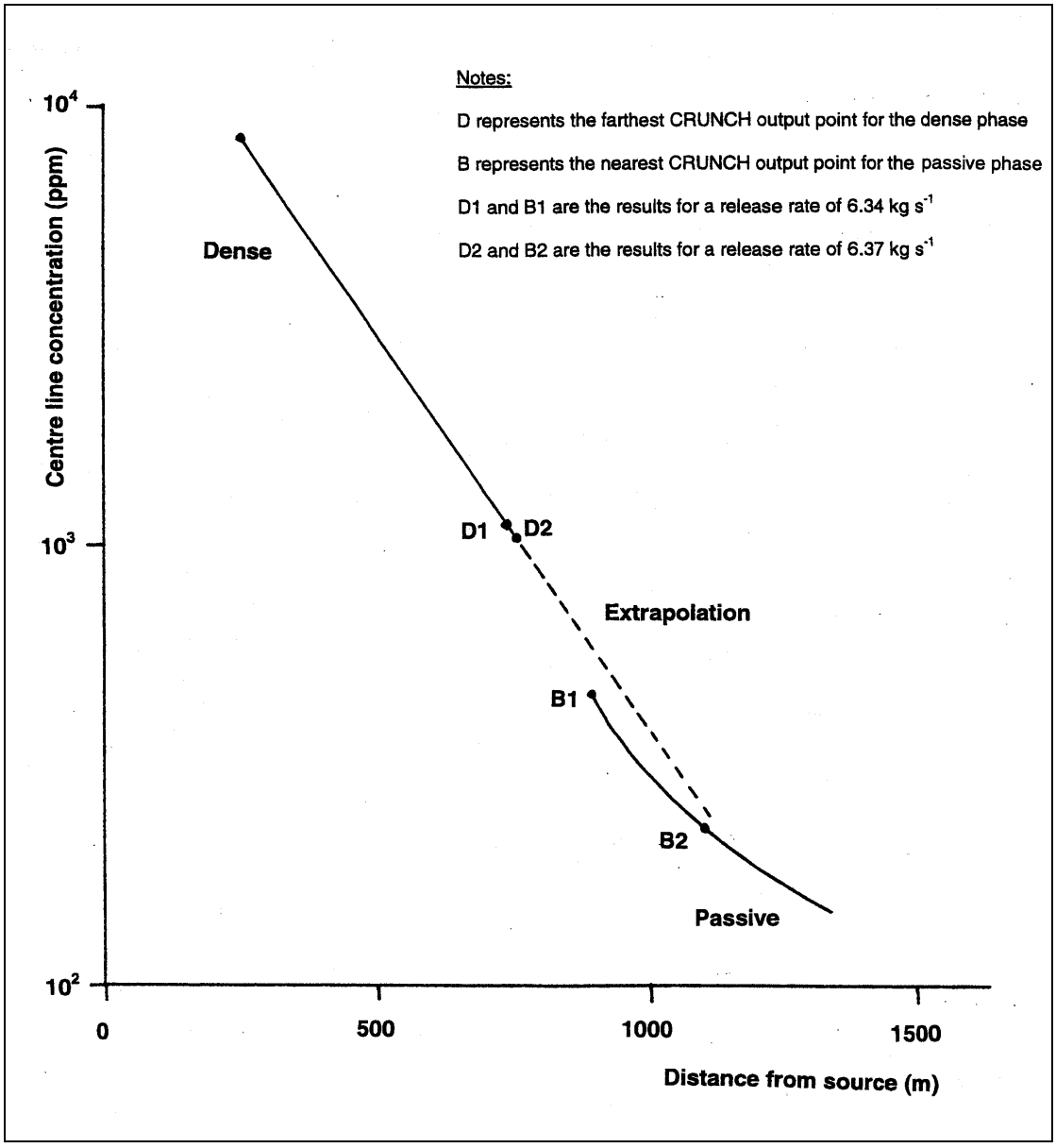

Figure 4
CRUNCH: variation in concentration with distance from source for release rate in the range 6.34 - 6.37 kgs^{-1}

The two branches of the illustrated curve: one for the dense phase and one for the passive phase, are different in character and quite widely separated. The points denoted D are the location of the farthest output point of the dense part of the plume, and those denoted B are the nearest points of the subsequent buoyant part. D1 and B1 are produced by the 6.34 kgs^{-1}case, whereas D2 and B2 are produced by the 6.37 kgs^{-1} run. FARSIDE (which has to interpolate between the D and B points) therefore perceives a discontinuity between the result for 6.34 kgs^{-1} and that for 6.37 kgs^{-1}. Analogous behaviour is also responsible for the staircase effect and the other discontinuity at the higher release rate.

Because of this discontinuous behaviour, an approximation was made to the CRUNCH predictions. It may be seen from Figure 4 (which is a log-linear plot) that the dense plume behaviour in this case is exactly exponential and that the extrapolation of this behaviour is a reasonable approximation to the buoyant plume predictions. The concentration at distances beyond the point where the plume ceases to behave as a dense gas is therefore calculated by

continuing the variation predicted in the last phase of the dense behaviour, assuming that this variation is exponential. It should be noted that this is not the only cause of discontinuities in the CRUNCH output (see Section 5).

3.4 GASTAR

GASTAR is normally provided with an interactive pre-processor, running under Windows, but can be run in batch mode in a similar way to CRUNCH. The input file is called 'Runname.GPL' and the corresponding formatted output file is 'Runname.GOF'.

The data requirements for the .GPL file were inferred from observations of the correspondence between values entered in the pre-processor and those observed in the file and in the output. These were supplemented by information in the user manual [5] and by the keyword lines in the input file itself, which are reasonably descriptive of the data which follow.

The problems described above in relation to the CRUNCH files are not present with GASTAR. The input file (being normally written by the pre-processor) represents the input data to high accuracy. The tables in the output are also written with a reasonable number of significant figures. However, interpolation of the tables is not necessary since concentrations at selected dose points are presented at the end of the file. The postprocessing routine forming part of CON in FARSIDE therefore need only step to the last line and read the result.

One complication which occurred with GASTAR was that an early run (using Version 3.05) for a sudden release highlighted a bug in the time integration of concentration at a fixed point. The calculations for the continuous release test case continued to be performed with Version 3.05.

When Version 3.08 was released with (among other things) the above bug corrected, the format of the .GPL (input) file had changed significantly and a useful input convergence parameter (present in the file, though not accessible from the pre-processor) had been omitted, presumably now being coded into GASTAR itself. The resulting default mesh of time points is quite coarse, resulting in a relatively inaccurate integration of the concentration and, consequently, a degree of scatter from case to case. This scatter is not noticeable normally, but does affect the convergence of FARSIDE on occasions, as will be discussed in Section 6.

4. SELECTION OF TEST CASES

4.1 INTRODUCTION

One of the goals of this work is to use the MLFP method to attempt to draw general conclusions concerning the importances of the main input quantities for the consequence models GASTAR and CRUNCH.

Section 2 described the basis of the MLFP method and, in particular, the expressions which define the sensitivities and importances of the individual model input variables.

The importance of an input variable is seen to be given (to a reasonable approximation) by the square of the product of its own uncertainty and the rate of change with respect to the variable of the consequence function Q, with this gradient being evaluated at the most-likely failure point (Section 2.3.2).

In order to achieve the above goal, a selection of relatively simple cases was first chosen in order to investigate the behaviour of the importances over a wide range of conditions. The types of release were selected as:

- continuous release with time limit (CRUNCH and GASTAR)
- sudden release (GASTAR only)

CRUNCH and GASTAR share a number of input variables which are either identical or can easily be translated from one to the other. The first set of test runs therefore represents a series of common CRUNCH/GASTAR cases (Section 5.1). Since this case already uses most of the CRUNCH options, further cases use GASTAR only.

The probability distributions assigned to each of the uncertain variables were selected in order to represent plausible ranges for each. The importances resulting from these distributions are expected therefore to be representative, though not necessarily definitive. Nevertheless, with the use of the expressions given in Section 2, it is possible to estimate (to within a reasonable degree of accuracy) how the importances would be expected to vary with changes in the assigned uncertainty distributions. In this way, general observations can be made from the results of a relatively small number of cases.

The remainder of this section outlines the test cases and then gives details of the probability distributions which have been selected for each of the common input variables. The results of these cases and of a number of variants are given in Section 5.

It should be noted that the algorithm for finding the MLFP and the expressions for the data importances both use numerical estimates of each component dQ/dx_i of the gradient. It is therefore necessary that Q varies reasonably smoothly with variations in its input quantities. As described in Section 3, and as will become more evident in what follows, this is not always the case.

4.2 CHOICE OF THE FAILURE SURFACE

As described in Section 2, one of the main parameters in the type of risk calculation carried out in this work is the value of Q which specifies the location of the failure surface (denoted Q_{lim}). In previous work (as for the example in Section 2) Q has been taken as the toxic dose D, and the value of Q_{lim} has been set to D_d, with a numerical value (for chlorine) of

1.08×10^5 ppm^2 min. In this work, in order to take account of the uncertainty in the toxicity, Q has been set equal to P_d, the fraction of the population killed by the exposure to the release. The relationship between this quantity and the airborne concentration χ (ppm) is discussed in Section 4.6. For the best-estimate values of toxicity, Table 1 summarises the relationship between P_d, D and χ.

Table 1
Correspondence between fraction of deaths, dose and concentration for chlorine

Fraction of deaths P_d	Dose D (ppm^2 min) (Note 1)	Concentration χ (ppm) (Note 2)
0.005	1.15×10^5	76
0.1	4.67×10^5	153
0.5	1.88×10^6	306
0.9	7.55×10^6	614

Notes: 1. Assuming best-estimate toxicity
2. Assuming 20 minutes exposure

In order to ensure that the results obtained by the test cases are as general as possible, the calculations have been carried out for Q_{lim} set to the range of P_d shown in Table 1. The lowest value (0.005) corresponds approximately to the value of D_d given above. Despite the large range of P_d and D, it may be seen that the concentration χ varies by only about a factor of 8.

4.3 Release rate and tank inventory

The release rate distribution was selected as a continuous distribution, based on the leak hole size frequency data used in Ref. 3. These data are reproduced in Table 2:

Table 2
Leak hole size vs expected frequency [3]

Hole dia δ (mm)	Frequency (10^{-6}y^{-1} per vessel)
6	40
13	10
25	5
50	5
(Total)	60

The discrete data were represented as a truncated exponential density distribution (see Appendix A) of the following form

$$\begin{aligned} p(\delta) &= a \exp(-\lambda\delta) \quad \delta_1 \le \delta \le \delta_2 \\ &= 0 \qquad\qquad\quad \text{elsewhere} \end{aligned} \tag{4.1}$$

where

δ_1, δ_2 are the lower and upper hole diameter limits

a is a constant which normalises the distribution so that $P(\delta_2)=1$

The limits δ_1 and δ_2 were chosen to be

$$\begin{aligned} \delta_1 &= 2\,\text{mm} \\ \delta_2 &= 66\,\text{mm} \end{aligned} \tag{4.2}$$

and the value of λ was calculated to give the same mean hole diameter as that implied by Table 2, yielding

$$\lambda = 0.0946\ \text{mm}^{-1}$$

and therefore

$$f(\delta) = 0.1146 \exp(-0.0946\delta) \qquad (2 < \delta < 66) \tag{4.3}$$

This distribution represents the tabular data fairly roughly, but is adequate for the test cases.

Both GASTAR and CRUNCH require the release rate c (kg s^{-1}) to be input, necessitating that the hole diameter is translated by the pre-processing routine (Section 3). This is carried out using an assumed discharge coefficient h of 0.0177 kg s^{-1} mm^{-2} [3], yielding

$$c(\text{kg s}^{-1}) = 0.0139(\delta(\text{mm}))^2 \tag{4.4}$$

This therefore gives a somewhat simplified representation of the actual process, since it assumes a single phase release. However, it is adequate for the purpose of providing, with (4.3), a reasonable spread of release rates.

Although the case to be modelled is that of a continuous release of chlorine, a toxic dose D can be calculated in the post-processing routine (Section 3) by assuming that this release lasts for a period of length τ_s (s), 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 minutes, then the release duration τ_s is given by:

$$\tau_s = \text{Min}(1200, c / M_g) \tag{4.5}$$

where

M_g is the tank inventory (kg)

Under this model, therefore, the release duration varies from case to case, but is not an independent uncertain parameter since it can be calculated from other uncertain quantities.

In the test cases, it is assumed that the tank inventory is represented by a uniform probability between 0 and $5x10^4$ kg. The use of the period τ_s in the calculation of the toxic dose D is described in Section 4.6.

In a real release, a fraction of the leaked material would form a liquid pool, which would evaporate relatively slowly. This is not represented in the common CRUNCH/GASTAR case: it is assumed that all the material leaking from the tank becomes airborne, either as vapour or droplets. Such an assumption will, of course, give rise to a pessimistic estimate of risk, but should not affect the calculated importances to a significant extent.

4.4 Other source data

As with the specification of the size of the source, both CRUNCH and GASTAR also require the following quantities to be specified:

- fraction of the release in the form of an aerosol
- rate of air entrainment at the source
- cross section shape of the plume at the source

4.4.1 Aerosol fraction

The release rate discussed in Section 4.3 may be considered as a mixture of vapour and droplets. Both CRUNCH and GASTAR require the droplet fraction to be specified. This has been assumed to be represented by a uniform distribution between 0 and 1; in other words: as wide a spread as possible.

4.4.2 Air entrainment rate

CRUNCH requires the air entrainment rate at source to be specified as a multiple of the release rate, whereas GASTAR specifies a rate in kg s^{-1}. The common input parameter (together with its probability distribution) was expressed in terms of the former for both codes, and converted into the latter in the GASTAR pre-processing routine in FARSIDE.

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. In order to avoid the possibility of negative values being produced during the FARSIDE iteration, a log normal distribution with these characteristics was used.

4.4.3 Plume shape at source

The plume shape at source is specified in CRUNCH as an aspect ratio r_a, where:

$$r_a = 2h/w$$

where

h is the plume height (m)

w is the plume width (m)

where the source area is calculated internally by CRUNCH in order to produce the correct downwind mass flow. GASTAR, on the other hand, requires plume width to be specified, and then calculates plume height in order to give the correct mass flow.

For both codes, 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 the GASTAR pre-processing routine in order to calculate the appropriate plume width.

4.5 Atmospheric variables

The atmospheric variables in question are:

- wind speed
- atmospheric stability
- wind direction

The work described in Ref. 3 developed a number of methods for the representation of the measured joint probability distribution of these quantities, and for the use of this information in a FARSIDE calculation. These techniques have been developed further during the course of this work (see Appendix B), but the results have not been used directly for this part of the analysis, because:

- the importances and sensitivities can be investigated using simpler methods
- neither CRUNCH nor GASTAR can fully represent continuous variations in atmospheric stability.

The second of these statements appears inconsistent with GASTAR's option of using Monin-Obukhov length L to quantify stability, rather than that of specifying Pasquill stability category. However, although L is used directly in the dense gas phase of the GASTAR calculation, the code appears to revert to Pasquill categories when deriving dispersion coefficients for the passive phase (Section 8.2.5 of [5]). Therefore the GASTAR predictions represent concentration at a dose point as a discontinuous function of L, despite the code's apparent capability of representing the behaviour in a continuous manner.

Because of this restriction, atmospheric stability is represented in the test cases by a selection of fixed Pasquill categories. These were selected in the neutral to stable range, since these tend to dominate the overall risk in a full QRA. It would be anticipated that the correct representation of the frequency distribution of stability would be fairly important for the correct calculation of risk, since higher doses are predicted for more stable conditions.

Therefore, if at some stage the consequence model was adapted to represent stability in a continuous manner, it would be desirable to revisit these calculations. Nevertheless, it would be expected that the treatment of stability in a discrete representation would not significantly affect the importances of the other input quantities. The conclusions drawn from the results of Sections 5 and 6 will therefore remain valid.

With atmospheric stability omitted as an uncertain variable, it is reasonable to approximate the wind speed frequency by an analytic distribution. As discussed in Ref. 3, a Weibull distribution is an appropriate representation. The distribution for this case represents a site where the winds are relatively light, with a mean wind speed of 2.7 m s^{-1} (compare with Section 6.1).

The use of a wind direction as one of the uncertain inputs was explored in Ref. 2, where it was observed that the inclusion of this variable sometimes gave rise to problems for the MLFP method. These problems arose from two sources, both related to the use of a 'box'-type consequence model:

(a) the type of model gave a sharp edge to the plume

(b) the plume shape for low wind speeds tended to be unrealistic.

The ways in which these features of the consequence model affected both the convergence and the accuracy of the MLFP model are discussed in detail in Ref. 2, but the main points are briefly outlined below:

Regarding (a), the basic heavy gas dispersion box model assumes a uniform concentration across the plume, the dimensions of which are calculated as a function of distance from the source. When searching for the failure surface (Section 2), the MLFP method must find the $D = D_d$ contour as a function of (amongst other things) the angle θ from the plume centreline. If the concentration changes abruptly with θ because the plume has a sharp edge, it may take the search algorithm some time to find this edge, since the variation of concentration with angle outside and inside the plume (both of which are essentially flat) give no clue as to where the plume edge lies.

Whereas (a) above tends to give rise to the observed increased number of model evaluations, (b) tends to cause a deterioration in accuracy. The plume model assumes that the vapour is carried along at the wind speed, with gravity-driven (slumping) flow only occurring perpendicular to the plume axis. For wind speeds which are large compared with the slumping velocity, this is a reasonable approximation, but for low wind speeds the vapour will tend to spread out as a disk, rather than in a plume.

The assumption that the vapour spreads out as a wide plume gives rise to a range of angles from zero to a certain limit where the variation of dose with windspeed (at a given distance) is independent on angle. This leads to a section of the failure surface which runs parallel to the axis representing angle. Conversely, because, at low wind speeds, the plume covers most of the half-plane downwind of the source, there exists a range of wind speeds from zero up to a certain limit where the variation of dose with angle is almost independent of windspeed. This leads to a section of the failure surface which runs parallel to the axis representing windspeed. The result of the combined behaviour is a failure surface which is not straight, as assumed in the FORM approximation (Section 2) but has a rounded 90° corner. This causes both the failure probability and the importance estimates to be less accurate than for a more realistic representation.

In light of these potential difficulties, it was considered that the following considerations gave sufficient justification for the omission of wind direction from these calculations:

1. the importance of wind direction itself can be inferred from the results of Ref. 3.

2. the omission of wind direction is unlikely to affect the importances of the other input quantities.

Regarding 1: it was shown in Ref. 3 that the importance of the wind direction distribution tends to be high for the near field and relatively low for the far field. This at first sight seems paradoxical, but the explanation is as follows:

The dose at any point tends to fall with increasing wind speed. For dose points in the near field, the MLFP tends to be at a relatively high wind speed, implying a fairly narrow plume. The calculated dose is therefore sensitive to whether the wind is blowing directly towards the dose point or not. At longer distances from the source, the MLFP occurs at lower wind speeds, implying a wide plume and a resulting reduction in the sensitivity to wind direction.

4.6 Toxicity

4.6.1 Basic equations

The toxic dose D experienced at any given location is calculated from the following expression:

$$D = \int_0^\infty \chi^n (t_m) dt_m \tag{4.6}$$

where

χ is the concentration (ppm)

n is an exponent which is usually (but not always) an integer (see below)

t_m is time in minutes

so the units of D are ppm^n min.

For a constant value of χ, (4.6) becomes

$$D = \chi^n \tau_m \tag{4.7}$$

where τ_m is the exposure duration in minutes. Clearly, (Equation (4.5))

$$\tau_m = \tau_s / 60 \tag{4.8}$$

It may be observed that, if n were set to 1.0 in Equations (4.6) and (4.7), then D would simply represent the total amount of chlorine which could be breathed by an individual located at the dose point, and would therefore include no information relating to the toxic effects. However, for a number of substances, the appropriate value of n exceeds 1.0 (see below), implying that some limited toxicity information is being included in the definition of D (i.e. that short exposures to high concentrations are worse than longer exposures to lower concentrations). Nevertheless, the dose D may be thought of as the 'physical' variable calculated by the consequence model, with the toxic impact being represented separately by the tolerance data.

The tolerance of individuals to toxic doses can be represented by a cumulative probability distribution $P_d(D)$, defined such that a dose D will cause a fraction $P_d(D)$ of the exposed population to die. Associated with $P_d(D)$ is the probability density $p_d(D)$, defined such that $p_d(D)dD$ represents the fraction with tolerances between D and D+dD.

Whereas a Gaussian distribution of tolerance amongst the population might be anticipated, it is perhaps unexpected that the distribution of this form tends to apply not to the dose, but to its logarithm. Furthermore, this type of rule tends to hold for a wide variety of causes of injury or death [7].

Therefore, let u be the variable defined as

$$u = \alpha + \beta Ln(D) \tag{4.9}$$

where α and β are constants. It is found empirically that, for suitable fitted values of α and β, the distribution of tolerance within the population can be represented by the standard normal density distribution

$$p_d(u) = \phi(u) = \left(1/\sqrt{(2\pi)}\right)\exp\left(-u^2/2\right) \tag{4.10}$$

which has a mean of 0.0 and a standard deviation of 1.0. Letting

$$x = Ln(D) \tag{4.11}$$

and substituting into (4.10) gives, after suitable rearrangement

$$p_d(x) = \left(1/(1/\beta)\sqrt{(2\pi)}\right)\exp\left(-1/2[(x-(-\alpha/\beta))/(1/\beta)]^2\right) \tag{4.12}$$

which is a normal distribution with mean μ_x and standard deviation σ_x, where

$$\mu_x = -\alpha/\beta \tag{4.13a}$$

$$\sigma_x = 1/\beta \tag{4.13b}$$

The quantities μ_x and σ_x, the mean and standard deviations of the population tolerance distribution, are considered here to represent two independent quantities characterising the chlorine toxicity (with possibly the exponent n being a third). Each of these quantities (despite being themselves parameters in a probability distribution) can be considered as uncertain inputs to an expanded consequence model which now predicts not dose, but number (i.e. fraction) of deaths. These quantities can be considered as candidates for representation by probability distributions in the MLFP calculation, which now calculates the risk that the number of deaths would exceed a selected critical value.

4.6.2 Use of published toxicity data

Best-estimate values of μ_x and σ_x can be derived from published toxicity data as follows. In risk analysis, injury (including toxicity) tolerance data is frequently published with reference to a derived 'harm' variable y, known as a probit, where

$$y = a + bLn(V) \tag{4.14}$$

where

V is the variable quantifying the physical cause of the injury (i.e. V is dose D, $=C^n t$, in this example)

The published values of a and b have been derived so that the population tolerance to y has a normal distribution with (for historical reasons) a mean value of 5.0 and a standard deviation of 1.0, which is just a standard normal distribution shifted by 5 standard deviations. It is therefore evident that:

$$\alpha = a - 5 \tag{4.15a}$$

$$\beta = b \tag{4.15b}$$

Typical values for chlorine as published in the American Center for Chemical Process Safety (CPSS) QRA guidelines ([8], see also [7]) are:

$$a = -8.29 \quad (4.16a)$$

$$b = 0.92 \quad (4.16b)$$

$$n = 2 \quad (4.16c)$$

which yield

$$\mu_x = 14.45 \quad (4.17a)$$

$$\sigma_x = 1.087 \quad (4.17b)$$

As discussed above, both μ_x and σ_x may be considered to be uncertain to some degree. For these example calculations, each of these quantities is represented by a normal distribution with means of -14.45 and 1.087 respectively.

4.6.3 Uncertainty in toxicity data

An accurate assessment of the uncertainties inherent in the values of μ_x and σ_x (Equation (4.17)) would require a review of basic toxicity data. This is beyond the scope of the current work, and it is considered that such effort would not be productive at this stage, since rough estimates of the uncertainties should be sufficient. To obtain these estimates, the uncertainties in the probit coefficients summarised in [7] have been reviewed. There are considered to be a number of possible sources of uncertainty:

– basic experimental data

– interpretation and processing of data into probit coefficients

– composition of the population at risk

The magnitude of the uncertainty in μ_x has been estimated from the variation in the 'a' coefficients given in Section 18.18.18 of [7] and summarised in Table 3.

Table 3
Variability of probit coefficients for chlorine due to differences in population

Population	Level of activity	a (Note 1)	μ_x
Regular	Base	-9.57	15.8
Regular	Standard	-8.29	14.4
Vulnerable	Base	-7.88	13.6
Vulnerable	Standard	-6.61	12.6

Notes
1. The value of b is constant at 0.92

The values of μ indicate a range of 3.2 between the smallest and largest values. Since these values are derived mainly from experiments on animals, it is likely that an additional

uncertainty will arise from the extrapolation of the results to humans. Furthermore, the behaviour of people (and therefore their exposure and respiration) under such extreme circumstances must also be uncertain. Based upon the above table and the rather qualitative arguments which follow it, it is considered that it would be reasonable to represent the uncertainty distribution of μ_x by a Gaussian with a mean (μ_μ) of 14.45 (based on a = -8.29, b = 0.92) and a standard deviation (σ_μ) of 1.5.

Since the value of σ_x (representing the spread of vulnerabilities within a population: see above) is assumed to be independent of μ_x, it can also be represented by a separate Gaussian distribution.

The mean of this distribution (μ_σ) can be set equal to 1/b (i.e. 1.087). There is little information upon which to base a value of the standard deviation (σ_σ), so a value of 0.15 has been selected as a trial value. As will be seen, this parameter is relatively unimportant relative to μ_x.

4.7 Plume transport data

The treatment of plume transport is one area where the specifications for the CRUNCH and GASTAR cases differed in detail. The following quantities characterise the plume transport:

- ground roughness length
- ground friction factor
- ground temperature
- air temperature
- humidity

The specification of roughness length is common to both codes and was 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.

A quantity referred to as the ground friction factor was included in the CRUNCH input, but no obvious equivalent is present in GASTAR. The use made of this quantity in the CRUNCH model was not evident in Ref. 4. Nevertheless, in order to investigate the importance of this parameter, the CRUNCH default value of 0.01 was selected for the mean, with a standard deviation of 0.002.

Atmospheric humidity was present in GASTAR but not in CRUNCH. A uniform distribution was assumed between limiting values of 10% and 90%.

The exchange of heat between ground and plume was represented differently between the cases using CRUNCH and those using GASTAR, due to some uncertainty as to how the specifications could be made to be equivalent between the two codes. In the CRUNCH input, the ground heat transfer coefficient was set to zero: for GASTAR, the ground temperature was set to 290 K under the assumption that the code calculates the heat transfer automatically (since it does not appear to be an input option). For both codes, a uniform distribution between 270 K and 300 K was assumed for the air temperature.

It may be noted that, in reality, the ground-air temperature difference ΔT is closely related to atmospheric stability, and hence it would not be accurate in a full QRA to treat these two quantities as independent. However, in this restricted analysis, the approach adopted is suitable to determine the importance of ΔT in terms of the extent to which uncertainties in its value are likely to affect the overall risk estimate.

4.8 Summary of the common CRUNCH/GASTAR case

Table 4 gives a summary of the assumptions made for the probability distributions in the common CRUNCH/GASTAR (continuous release) case. The distributions are described in Appendix A.

Table 4
Summary of input assumptions for common CRUNCH / GASTAR case

Quantity	Distribution	Parameters*
Leak hole diameter (mm)	Truncated exponential	$\lambda = 0.0946$ mm^{-1} $x_1 = 2$ mm $x_2 = 66$ mm
Mass fraction of release as an aerosol (-)	Uniform	$x_1 = 0.0$ $x_2 = 1.0$
Air temperature (K) (The GASTAR ground temperature was fixed at a value of 290 K)	Uniform	$x_1 = 270.0$ K $x_2 = 300.0$ K
Multiple by mass of air entrained at source (-)	Lognormal	$m = 2.26$ $s = 0.294$ (Gives $\mu = 10.0$ $\sigma = 3.0$)
Aspect ratio of the source (-)	Normal	$\mu = 1.0$ $\sigma = 0.2$
Wind speed (m s^{-1})	Weibull	$k = 1.5$ $w = 3.0$ (Gives $\mu = 2.71$ m s^{-1} $\sigma = 1.84$ m s^{-1})
Roughness length (m)	Lognormal	$m = -2.65$ $s = 0.833$ (Gives $\mu = 0.1$ m $\sigma = 0.1$ m)
Friction factor of the surface (-) (CRUNCH only)	Normal	$\mu = 0.01$ $\sigma = 0.002$
Tank inventory (kg)	Uniform	$x_1 = 0.0$ kg $x_2 = 5\times10^4$ kg
Toxicity tolerance distribution mean (-)	Normal	$\mu_\mu = 14.45$ $\sigma_\mu = 1.5$
Toxicity tolerance distribution standard deviation (-)	Normal	$\mu_\sigma = 1.087$ $\sigma_\sigma = 0.15$
Humidity (%) (GASTAR only)	Uniform	$x_1 = 10$ $x_2 = 90$

* See Appendix A, Section A4.1

5. RESULTS FOR THE COMMON CRUNCH/GASTAR TEST CASE

The first case which was run was intended to exploit as much common ground as possible between the two consequence models. The case was therefore selected to be based on a continuous release of chlorine, but with an upper limit for the release duration, as described in Section 4. The case uses the probability distributions as summarised in Table 3, with a fixed stability category of 'F' (see Section 4.5).

5.1 GASTAR

With GASTAR (Version 3.05) as the consequence model, FARSIDE was run for Q set as the fraction of deaths, P_d, with Q_{lim} set respectively as 0.005, 0.1, 0.5 and 0.9 (Table 1).

The results are summarised as follows. Figure 5 to Figure 8 show the variation with distance of the importances (I) calculated for each variable (see Section 2), with values of $I \leq 10^{-4}$ set to 10^{-4}. Table 5 presents the values of β and P_{fl} for each of the cases discussed in Section 5, and Appendix E summarises the MLFP values of each of the uncertain variables, for each case discussed in this report.

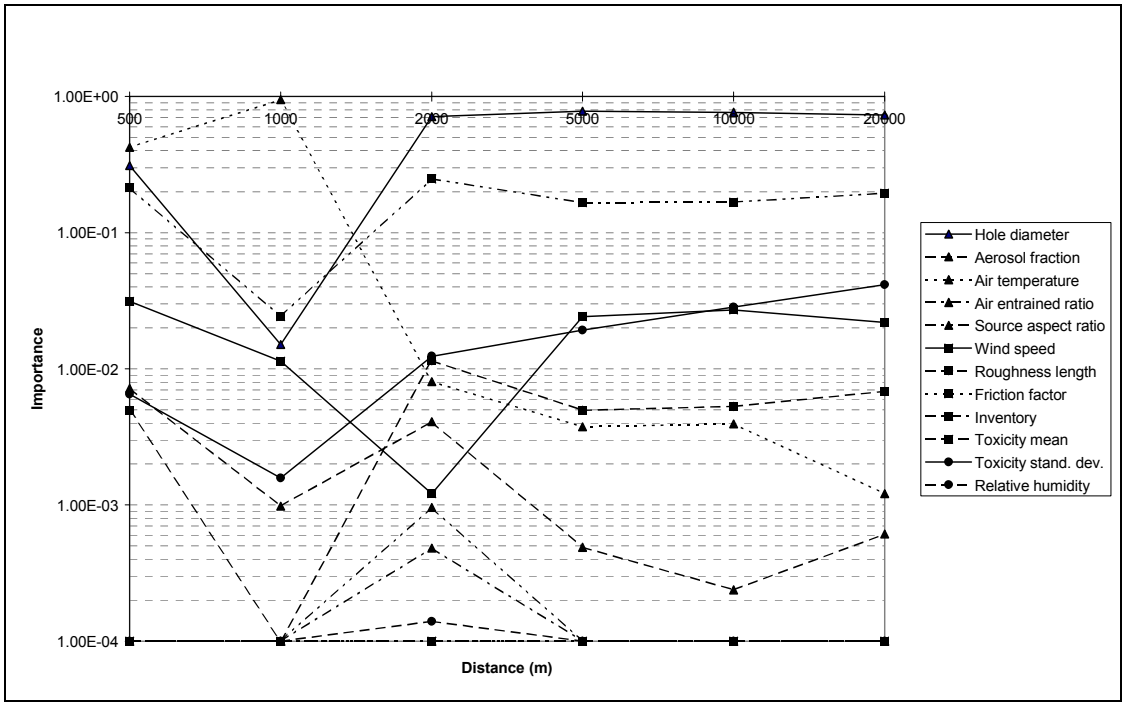

Figure 5
GASTAR/CRUNCH common case. Importances derived from GASTAR results for Q_{lim} (fractional number of deaths) = 0.005

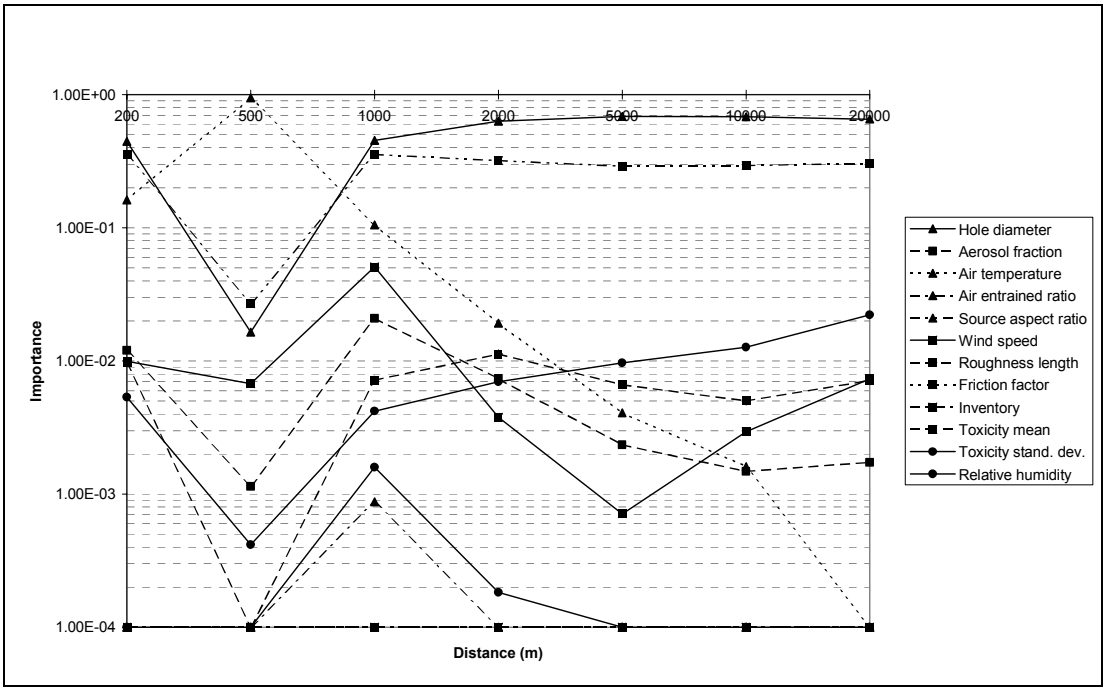

Figure 6
GASTAR/CRUNCH common case. Importances derived from GASTAR results for Q_{lim} (fractional number of deaths) = 0.1

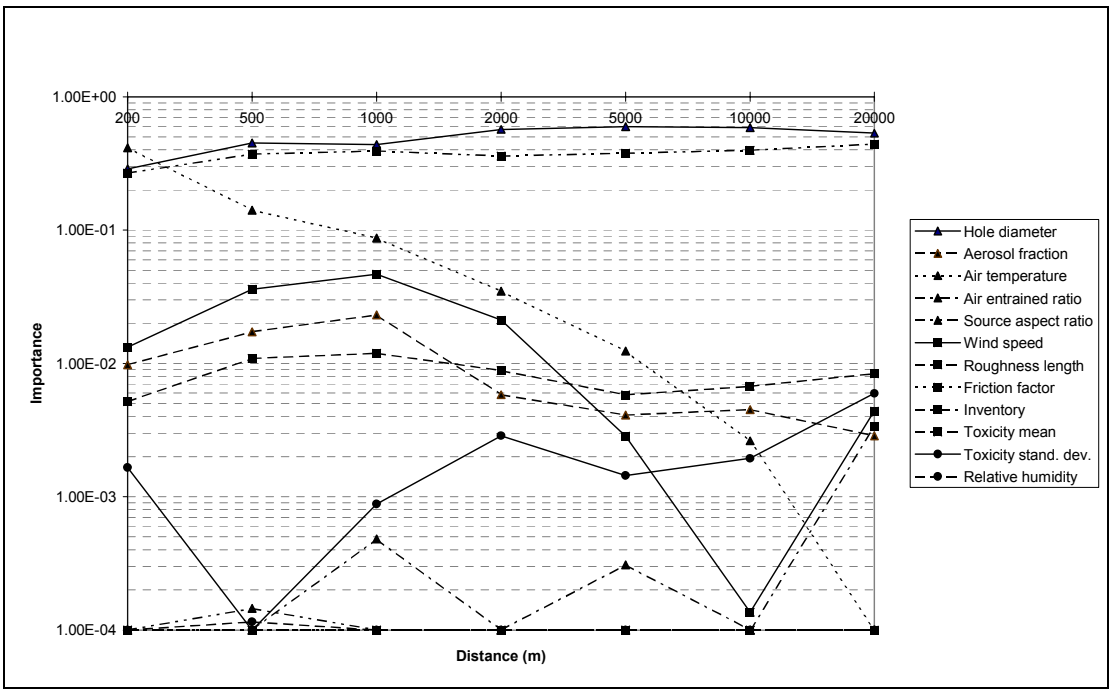

Figure 7
GASTAR/CRUNCH common case. Importances derived from GASTAR results for Q_{lim} (fractional number of deaths) = 0.5

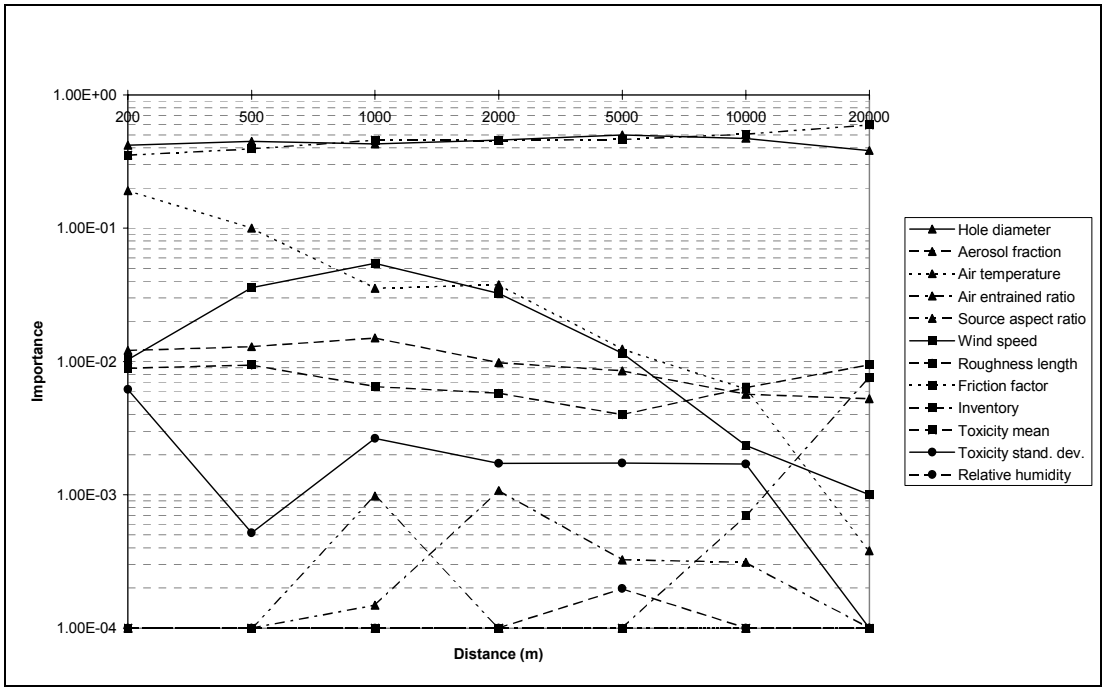

Figure 8
GASTAR/CRUNCH common case. Importances derived from GASTAR results for Q_{lim} (fractional number of deaths) = 0.9

A number of features of the results are evident:

- the importances are dominated by those of the hole diameter and the mean of the toxicity distribution

- with odd exceptions (discussed below) these importances are not particularly sensitive to Q_{lim} or to d, the distance from the source.

- the importance of the standard deviation of the toxicity distribution is small compared with that of the mean.

- the importances of the air entrained ratio, the source aspect ratio, the relative humidity and the friction factor are insignificant (that of the latter is zero, implying that the variable is irrelevant in the model). These quantities can be set to reasonable fixed values without affecting the calculated risk.

- the importances of the aerosol fraction, the wind speed and the roughness length are in the middle range of importance, where further investigation might be considered desirable. The typical values are not particularly sensitive to Q_{lim}, though some variation with d is evident.

- the importance of air temperature is (surprisingly) high at short distances from the source, particularly for $Q_{lim} = 0.005$ (1000 m), $Q_{lim} = 0.1$ (500 m), and $Q_{lim} = 0.5$ (200 m), but falls rapidly at longer distances.

- the importance of the inventory begins to increase sharply for high Q_{lim} / large d (Figures 5.3, 5.4).

- there are large fluctuations in some of the smaller importances.

It should be noted that a negative value of β implies that the best estimate point lies in the fail region. Also, it should again be emphasised that P_{fl} is an estimate of the conditional risk. The value should be multiplied by the leak frequency (i.e. $60x10^{-6}$/y : Table 2) to get the actual risk per year.

Before proceeding with a detailed analysis of the results, it is worth discussing the significance of the fluctuations which can be seen in some of the smaller importances (for example, see the values corresponding to the air entrained ratio in Figure 5 to Figure 8). As discussed in Section 2 and described in detail in Appendix A, the FARSIDE search algorithm calculates the local gradient of the 'quantity of interest' Q by altering each input quantity by a small amount and recalculating Q (re-running the consequence model). In cases where the change in a variable should have little or no effect, the value of Q may nevertheless change due to numerical noise or discontinuities, such as those discussed in Section 3.3. It is therefore to be anticipated that very small importances (which are of no significance anyway) may be swamped by random fluctuations. In addition, because of the normalisation to 1.0, importance values for all variables will be affected if one displays erratic behaviour (see particularly Section 5.1.3).

Table 5
Values of β and conditional risk (as estimated by P_{f1}) for cases discussed in Section 5

Case		Results at distance (m) of:						
		200	**500**	**1000**	**2000**	**5000**	**10000**	**20000**
Continuous / GASTAR / $P_d = 0.005$ (Figure 5)	β	-	-0.65	-0.43	-0.21	0.48	1.00	1.53
	P_{fl}	-	7.4×10^{-1}	6.7×10^{-1}	5.8×10^{-1}	3.2×10^{-1}	1.6×10^{-1}	6.4×10^{-2}
Continuous / GASTAR / $P_d = 0.1$ (Figure 6)	β	-0.919	-0.399	0.146	0.303	0.977	1.520	2.076
	P_{fl}	8.2×10^{-1}	6.6×10^{-1}	4.4×10^{-1}	3.8×10^{-1}	1.6×10^{-1}	6.4×10^{-2}	1.9×10^{-2}
Continuous / GASTAR / $P_d = 0.5$ (Figure 7)	β	-0.462	0.007	0.442	0.886	1.55	2.10	2.69
	P_{fl}	6.9×10^{-1}	5.0×10^{-1}	3.3×10^{-1}	1.9×10^{-1}	6.1×10^{-2}	1.8×10^{-2}	3.6×10^{-3}
Continuous / GASTAR / $P_d = 0.9$ (Figure 8)	β	-0.03	0.595	1.06	1.52	2.17	2.73	3.36
	P_{fl}	5.1×10^{-1}	2.8×10^{-1}	1.5×10^{-1}	6.4×10^{-2}	1.5×10^{-2}	3.1×10^{-3}	4.0×10^{-4}
Continuous / CRUNCH / $P_d = 0.005$ (Figure 12)	β	-1.84	-0.82	-0.204	0.336	1.16	1.85	2.30
	P_{fl}	9.7×10^{-1}	8.0×10^{-1}	5.8×10^{-1}	3.7×10^{-1}	1.2×10^{-1}	3.2×10^{-2}	1.1×10^{-2}
Continuous / CRUNCH / $P_d = 0.5$ (Figure 13)	β	-0.871	-0.025	0.527	1.23	2.19	2.91	3.67
	P_{fl}	8.1×10^{-1}	5.1×10^{-1}	3.0×10^{-1}	1.1×10^{-1}	1.4×10^{-2}	1.8×10^{-3}	1.2×10^{-4}

The above observations concerning the importances will now be discussed.

5.1.1 Hole diameter

The FARSIDE (MLFP) algorithm (usually) starts with the best-estimate values of the uncertain input variables and then attempts to adjust the values of these variables within their ranges in order to increase the value of Q until it reaches Q_{lim}. The fact that the leak hole diameter has been calculated as having a high importance indicates that, whatever the required increase in Q to reach Q_{lim}, this can be achieved relatively easily (i.e. by a change which is small when expressed as a number of standard deviations) by increasing the hole size.

The MLFP calculation in this example demonstrates formally what would be expected intuitively: that our uncertainty as to the threat posed by the installation is due mainly to a lack of knowledge about what kind of failure of the containment will actually occur. However, the results in Section 5.1.2 indicate that other uncertainties may (in principle) have a significant influence. This will be discussed further in Section 7.

5.1.2 Toxicity

The uncertainty in the value of (particularly) the mean of the population tolerance to chlorine poisoning was derived in a fairly approximate manner from published estimates of the variability. The MLFP calculation indicates that the calculated risk is rather sensitive to this quantity, and that this sensitivity is likely to apply in general.

The calculation also implies that the use of the mean values μ_x and σ_x of the tolerance distribution would be likely to underestimate the risk significantly, since it ignores the 'risk' that the population may actually be more vulnerable than these values would imply. For the GASTAR case of $P_d = 0.5$ / d = 10000 m, we have (Table 5 and Figure 7):

$$\beta = 2.1$$

$$P_{fl} = 1.78 \times 10^{-2}$$

$$I_T \approx 0.4$$

where the subscript T (toxicity) indicates that the value includes the distributions of both μ and σ. From Equations (2.21) and (2.24)

$$\frac{\Delta P_{fl}}{P_{fl}} \approx \frac{\beta^2}{2} I_T \approx 0.88 \qquad (5.1)$$

i.e. one would expect that ignoring the uncertainty in μ would strongly affect the calculated risk. In order to investigate this prediction, the above case was run again with the standard deviations of both μ_x and σ_x set to very small values. The resulting decrease in risk was even greater than (5.1) suggests:

P_{fl} (with distribution of toxicity tolerance) = 1.78×10^{-2}

P_{fl} (with mean/median values of toxicity tolerance) = 1.07×10^{-5}

The omission of the uncertainty in toxicity tolerance has therefore given rise to a large decrease in the calculated risk. The reason for this very large change can be identified easily

from the FARSIDE results: since the algorithm cannot now increase Q by adjusting the toxicity tolerance, it must do so primarily by changing the values of the other variables. The following table (Table 6) compares the MLFP values between the two cases.

Table 6
Conditions at MLFP for cases with and without toxicity uncertainty

Input quantity	MLFP values for cases with:			
	Uncertain Toxicity		Best estimate toxicity	
	Value	I	Value	I
Hole size (mm)	32.5	5.87×10^{-1}	55.6	3.91×10^{-1}
Aerosol fraction (-)	0.444	4.50×10^{-3}	0.052	1.46×10^{-1}
Air temperature (K)	284	2.63×10^{-3}	276	3.93×10^{-2}
Air entrained ratio (-)	9.6	9.59×10^{-6}	11.2	1.54×10^{-2}
Source aspect ratio (-)	1.0	0	1.0	0
Wind speed (m s^{-1})	2.31	1.36×10^{-4}	2.68	1.77×10^{-3}
Roughness length (m)	0.061	6.73×10^{-3}	0.011	2.78×10^{-1}
Friction factor (-)	0.01	0	0.01	0
Inventory (kg)	2.5×10^{4}	0	4.66×10^{4}	1.24×10^{-1}
Toxicity mean	12.5	3.97×10^{-1}	14.4	2.31×10^{-3}
Toxicity std. dev.	1.10	1.94×10^{-3}	1.09	1.26×10^{-6}
Rel. humidity (%)	50.2	1.07×10^{-5}	50.0	0

The first feature to notice is that the hole diameter has been increased to a value close to the upper end of its range (55.6 mm) such that only a fraction of about 0.004 of the probability distribution lies between this value and the upper limit of 66 mm. At this hole diameter, two other effects occur:

– it becomes 'easier' for the algorithm to alter parameters other than the hole size.

– the total amount of material available (i.e. the inventory) becomes significant

For hole diameters below about 38.7 mm, the dose cannot be increased by increasing the inventory, since the tank contents (assuming a mean inventory of 2.5×10^{4} kg) will not be exhausted by the end of the period of 20 minutes assumed (Section 4.3) as the upper limit for the leak duration. However, above 38.7 mm, the release is curtailed by the emptying of the tank, and hence the quantity present at the start of the leak becomes significant.

This effect is also seen over the range covered by Figure 5 to Figure 8, where the importance of the inventory, which is zero for most of the entries, begins to increase for the highest Q_{lim} (0.9) and the furthest distances (Figure 8). Inspection of the FARSIDE results indicate that, at a dose point distance of 10000 m, the MLFP hole diameter is 38.2 mm, consistent with the above analysis.

Returning to the case with best estimate toxicity: Table 6 shows (as one would expect) increased importances for some of the 'transport' variables which exhibited only minor significance in the original case. In particular, the importances of the aerosol fraction and the roughness length have increased substantially.

The importance of the roughness length is reasonably consistent with the results of Ref. 3 (where the toxicity tolerance was not included) for large values of d. However, the calculated importance may partly be due to the relatively large standard deviation assumed: although the roughness length is a fitted parameter without a well-defined value, the uncertainty may be greater than that normally considered for known terrain.

The reason why a reduction in aerosol fraction should lead to a significant increase in dose at this distance (i.e. 10000 m) is not entirely clear. The lower liquid fraction should give rise to a warmer plume, which may inhibit the lateral spread, but this is a rather speculative explanation. In fact, because of the rather large spread of the distribution (Section 4.4.1), the aerosol fraction has in this case been reduced to a value which can be viewed as unrealistic. The high importance of this quantity calculated in this particular case may therefore be considered to be spurious. The effect on the importance of a reduction in the standard deviation of the distribution may be estimated, at least approximately, by the use of Equation (2.30).

5.1.3 Air temperature

As described above, the high and erratic importance exhibited by the air temperature at short distances was unexpected and has been investigated further. Figure 9 shows a plot of concentration χ (ppm) as predicted by GASTAR at a dose point 1000 m from the source, as a function of air temperature. The figure was produced by multiple runs, generated by a calling program which increments the selected x-axis variable over a chosen range. Other input variables were set to values corresponding to the MLFP values for 1000 m in Figure 5 (see Appendix E).

The predicted concentrations exhibit a sudden fall as the air temperature T_a approaches the ground temperature T_g, which was fixed at 290 K in all the cases carried out.

In the case corresponding to d = 1000 m in Figure 5, the MLFP value of air temperature was predicted to be 290 K, so the large value of I is therefore due to the large (negative) value of $d\chi/dT$ (corresponding to the term dQ/dx in Equation (2.30)) on the low-temperature side of this value.

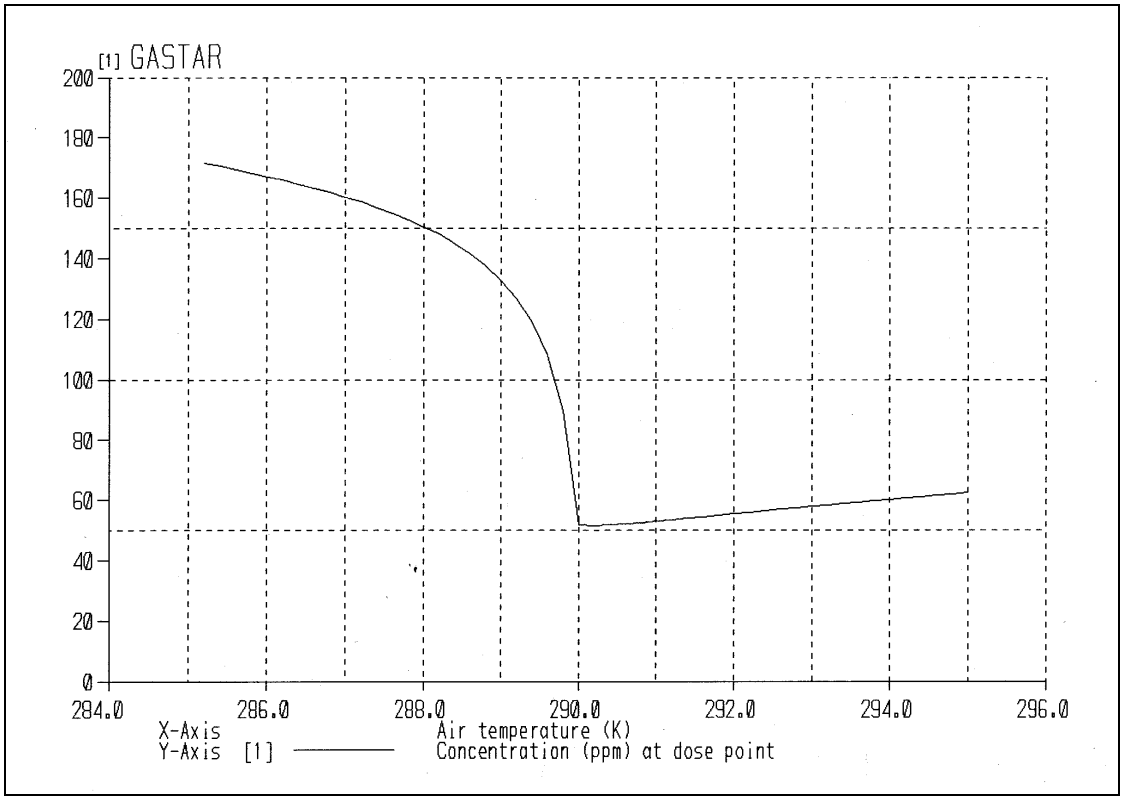

Figure 9
GASTAR: variation with air temperature of concentration at 1000 m. Ground temperature = 290 K. Other inputs as Figure 5 1000 m MLFP

It should be noted that the shape of the curve shown in Figure 9 is sensitive to the selected wind speed. Figure 9 shows the relationship for a wind speed v of 2.27 ms^{-1}. At lower wind speeds, the variation in the left hand branch of the curve reduces and that in the right hand branch increases, so that the relationship resembles a mirror image (about T_a = 290 K) of that illustrated (see Section 6.1).

The behaviour shown in Figure 9 is difficult to explain on physical grounds without further investigation, since it seems unlikely at first sight that the ΔT between ground and air (or ground and cloud) could have such a large effect over such a short distance. However, from the evidence provided by Figure 9 and by subsequent runs (see Section 0) it is possible that the observed behaviour is simply a feature of the way in which this particular model has been constructed and may only be evident for a small range around $\Delta T = 0$. In general, it would be expected that the importance of this quantity would be relatively low (as it proves to be for larger values of d. Further calculations relating to the effect of ΔT are reported in Section 6.1.

5.1.4 Other transport quantities

The nature of the plume as it forms in the vicinity of the source is subject to considerable uncertainty but, as the calculations show, the impact of this uncertainty (as expressed by the distributions of air entrainment and source aspect ratio) is relatively minor. The reason for this is first, that the mass of air entrained at the source, though large compared with the mass of material released, is relatively small compared with the mass entrained during the transport of the plume. The second reason is that the aspect ratio that the plume adopts a short distance downstream of the source due to slumping is almost independent of the aspect ratio that it possesses when released.

5.1.5 Wind speed and roughness length

The calculations indicate a relatively low importance for wind speed: at least compared with the importances of leak hole size and toxicity discussed above, even though this quantity had appeared to be quite important in the calculations reported in Ref. 3. Because of this, the variation of concentration with wind speed was investigated for two dose point distances: 500 m and 10000 m. The other variables were set to the MLFP values for the appropriate dose point distances in Figure 7 and the stability was set to 'F'.

The calculated variation is shown in Figure 10 for 500 m and in Figure 11 for 10000 m. The curves are markedly different, with a peak concentration at 6.0 m s^{-1} for d = 500, reducing to about 2.5 m s^{-1} for the d = 10000 m case. The figures therefore indicate a fairly modest variation in concentration with wind speed, leading to the relatively low importance referred to above. The variation of concentration with changes in wind speed is investigated further in Section 6.1, where the behaviour at low wind speeds at least in the GASTAR model is shown to be strongly determined by the ground-air temperature difference.

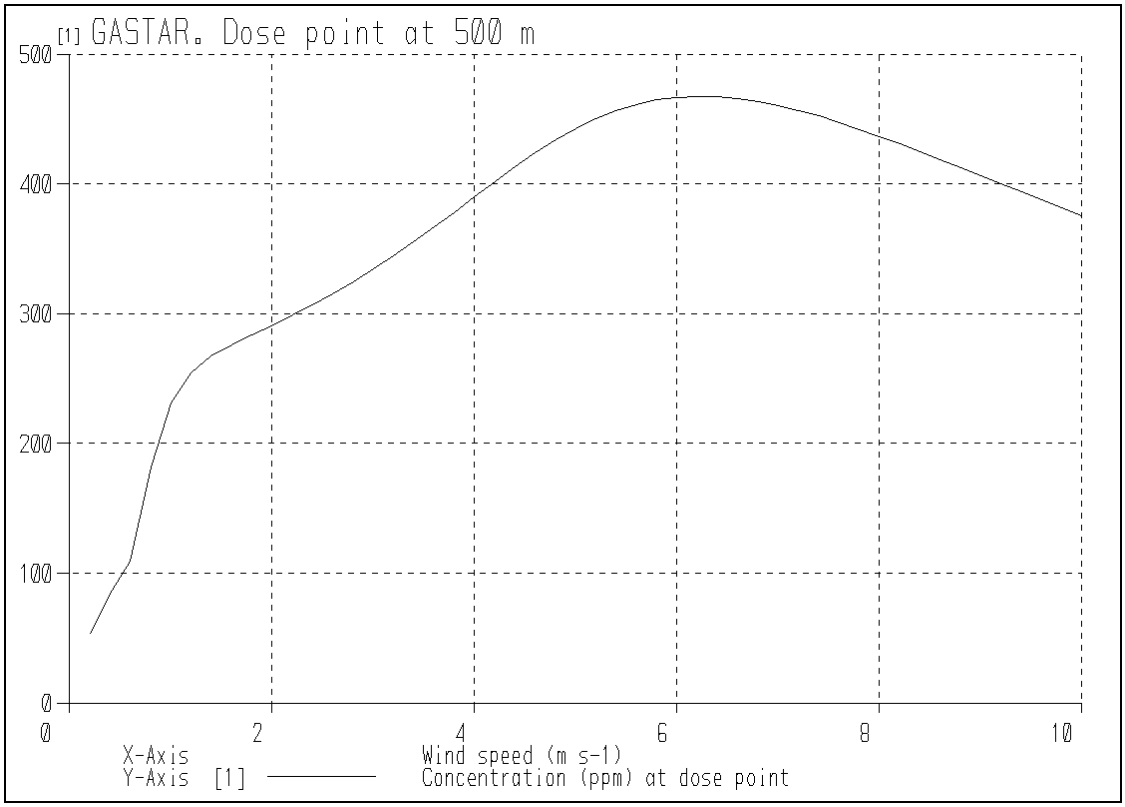

Figure 10
GASTAR: variation with wind speed of concentration at 500 m. Other inputs as Figure 7 500 m MLFP

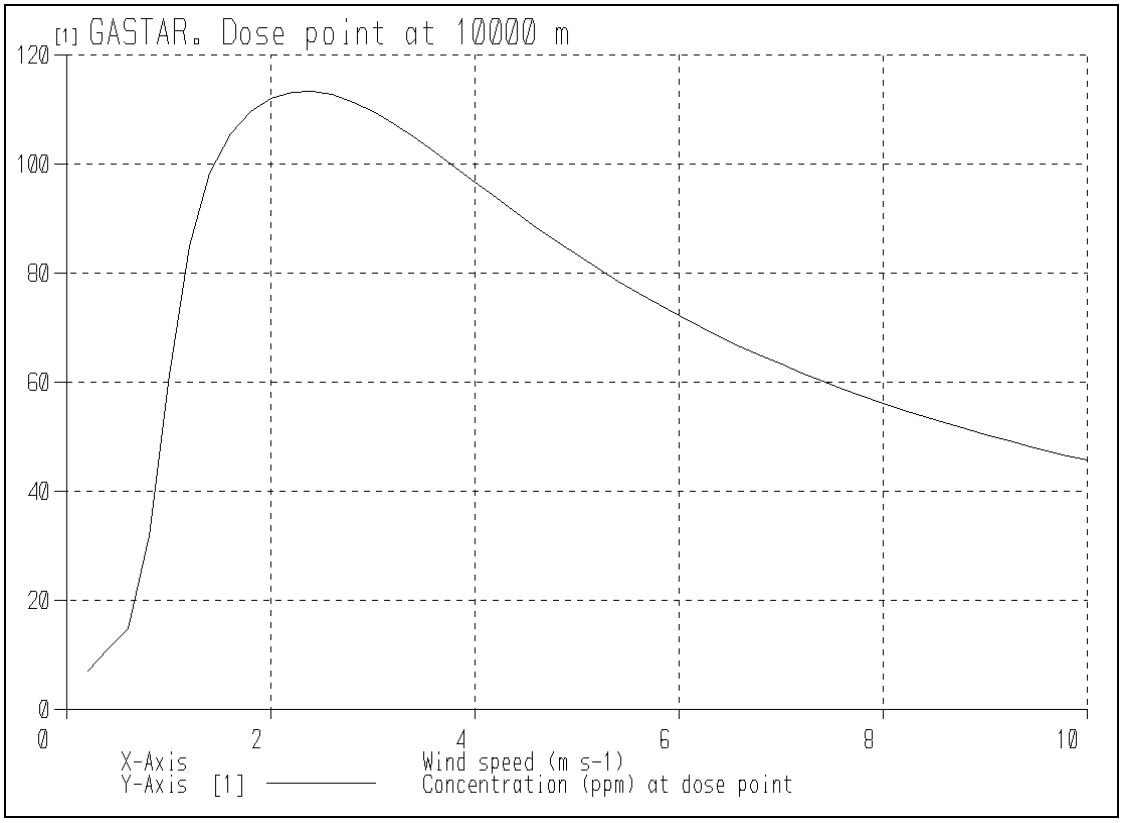

Figure 11
GASTAR: variation with wind speed of concentration at 10000 m. Other inputs as Figure 7 10000 m MLFP

5.2 CRUNCH

The CRUNCH version of the common GASTAR/CRUNCH test case used identical probability distributions to the GASTAR runs, with the exception of the exclusion of the relative humidity, which is not used in CRUNCH (and is shown to be of low importance in GASTAR). As was discussed in Section 4, CRUNCH contains some inherent features which make this type of calculation difficult, and another characteristic of the model (see below) was discovered during the course of the calculation. However, it was found to be possible to extract meaningful and consistent results, as described below.

CRUNCH was run for only two of the cases analysed by GASTAR: for Q_{lim} = 0.005 and 0.5. Figure 12 and Figure 13 therefore present the results (see also Table 5).

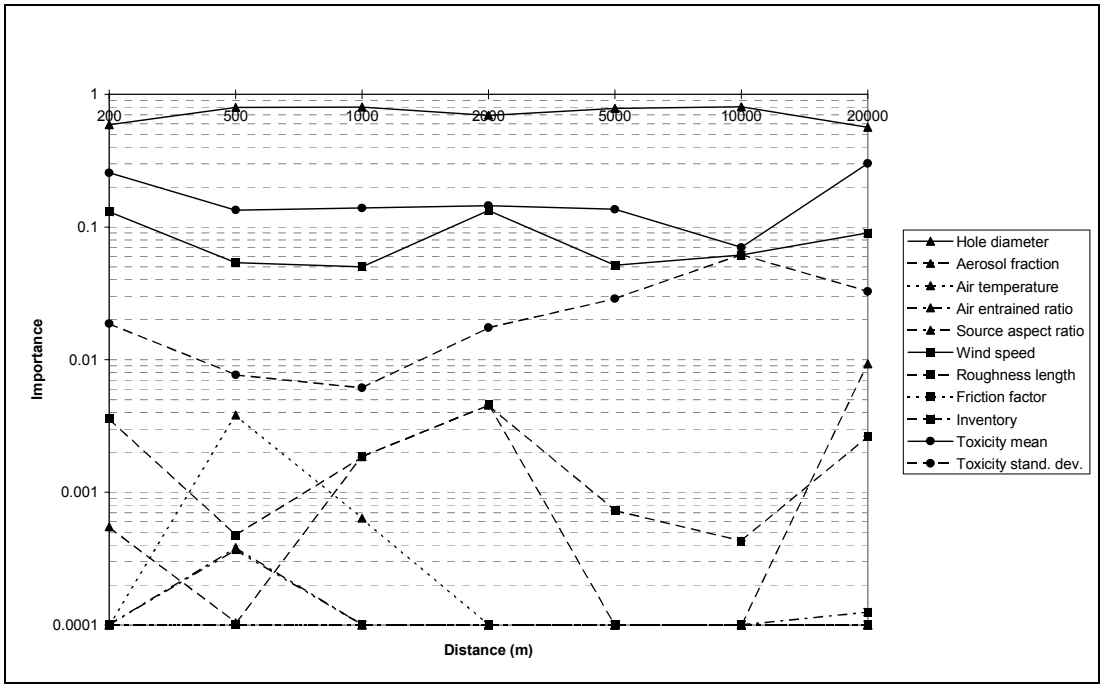

Figure 12
GASTAR/CRUNCH common case. Importances derived from CRUNCH results for Q_{lim} (fractional number of deaths) = 0.005

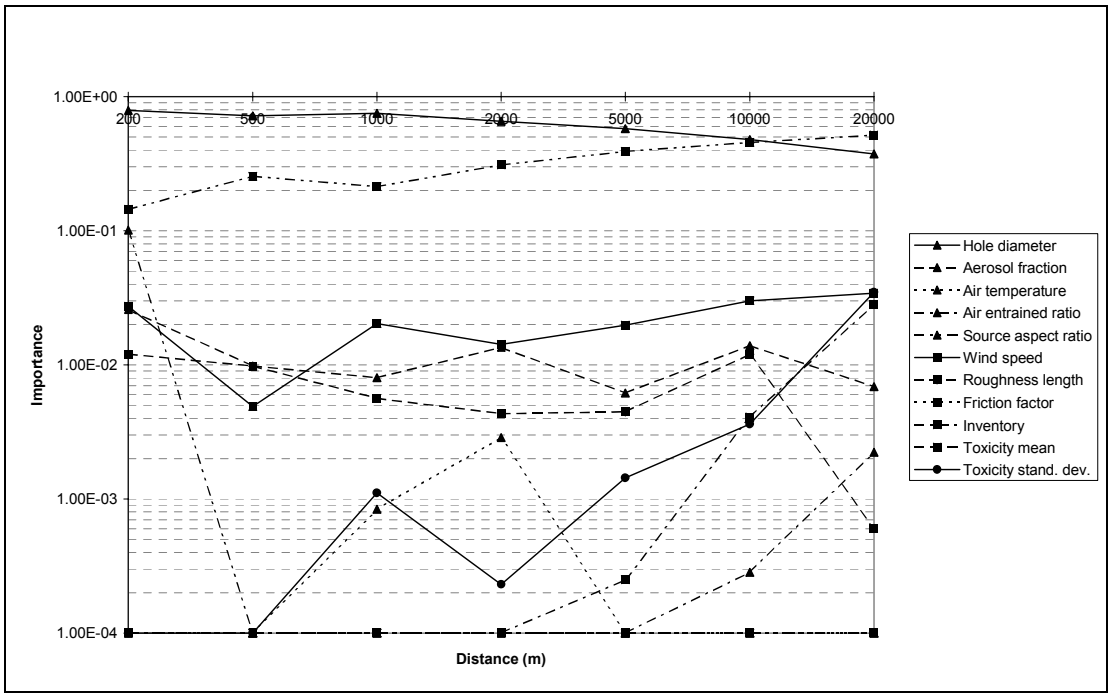

Figure 13
GASTAR/CRUNCH common case. Importances derived from CRUNCH results for Q_{lim} (fractional number of deaths) = 0.5

Although the magnitudes of some of the smaller importances differ between the two cases, those of the larger importances (i.e. hole diameter and toxicity) are similar and, of most direct relevance, the conclusions as to which category each variable falls within are unchanged.

One noticeable difference between the GASTAR and CRUNCH results is that, for distances greater than about 1000 m, the risks (expressed by P_{fl}; Table 5) as calculated using CRUNCH begin to fall below those predicted on the basis of the GASTAR model. Inspection of the GASTAR and CRUNCH outputs for a comparable case indicate that the predictions agree quite well for the dense phase, but differ (with GASTAR giving the higher concentration) over the passive phase of the plume. The reason for this difference has not been investigated: if either the models being executed or the calculations being specified are slightly different, this will promote, rather than hinder, the generality of the conclusions drawn from this study.

A number of convergence problems were noticed during the CRUNCH calculations and inspection of the FARSIDE outputs indicated a problem with the variation of concentration with roughness length z_o. A survey of this relationship was therefore carried out. The CRUNCH input parameters were set to the MLFP values corresponding to d = 500 m in Figure 13, and the concentration in ppm was then calculated for this distance, for a range of z_o from 10^{-4} m to $5x10^{-2}$ m, as shown in Figure 14.

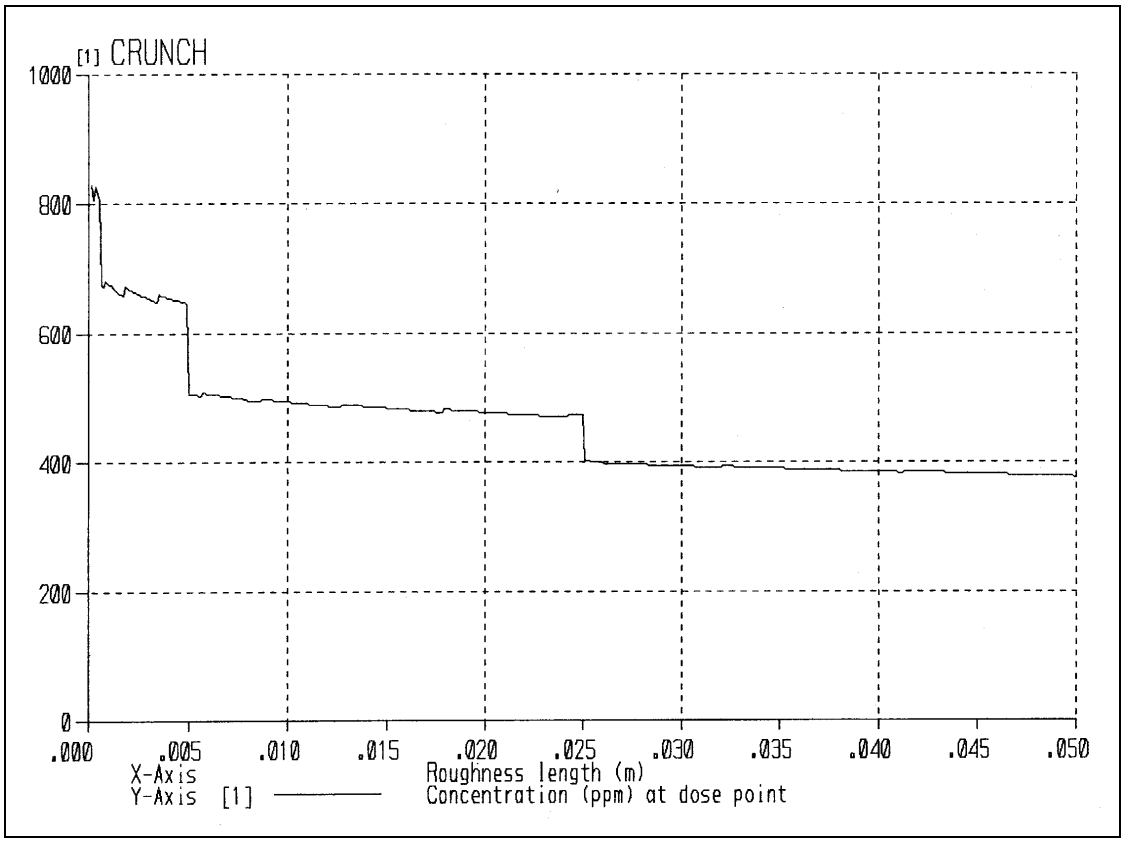

Figure 14
CRUNCH: variation with roughness length of concentration at 500 m. Other inputs as Figure 13 500 m MLFP

In addition to some slightly erratic behaviour at very low values, there are two clear steps at z_o = 0.005 m and z_o = 0.025 m. This is likely to be caused by the use of different formulas for the dispersion coefficients applying to different ranges of z_o; a relatively common assumption. The MLFP value of z_o in the above case is 0.071 m: well away from these features. However, in other cases, if z_o were close to the above values, it would be likely that the FARSIDE convergence would be impaired.

As implied by the analysis in Section 2, the MLFP method requires that the quantities calculated by the consequence model should be continuous functions of the input quantities. This is discussed further in Sections 6 and 7.

6. RESULTS FOR VARIANT CASES

6.1 'D' Stability (continuous release) (Variant Case 1)

The CRUNCH/GASTAR test case was re-run using GASTAR version 3.05 for Q_{lim}=0.5, but with the stability category set to D (i.e. neutral). The results are displayed in Figure 15. It is evident that the observations made in Section 5.1 also hold in broad terms for this case with the notable exception of the importance of wind speed, which dominates at 5000 m and above.

The greatly increased importance of wind speed shown in Figure 15 (compared with the analogous case in Figure 7) prompted an investigation of the response of the GASTAR model to changes in this quantity.

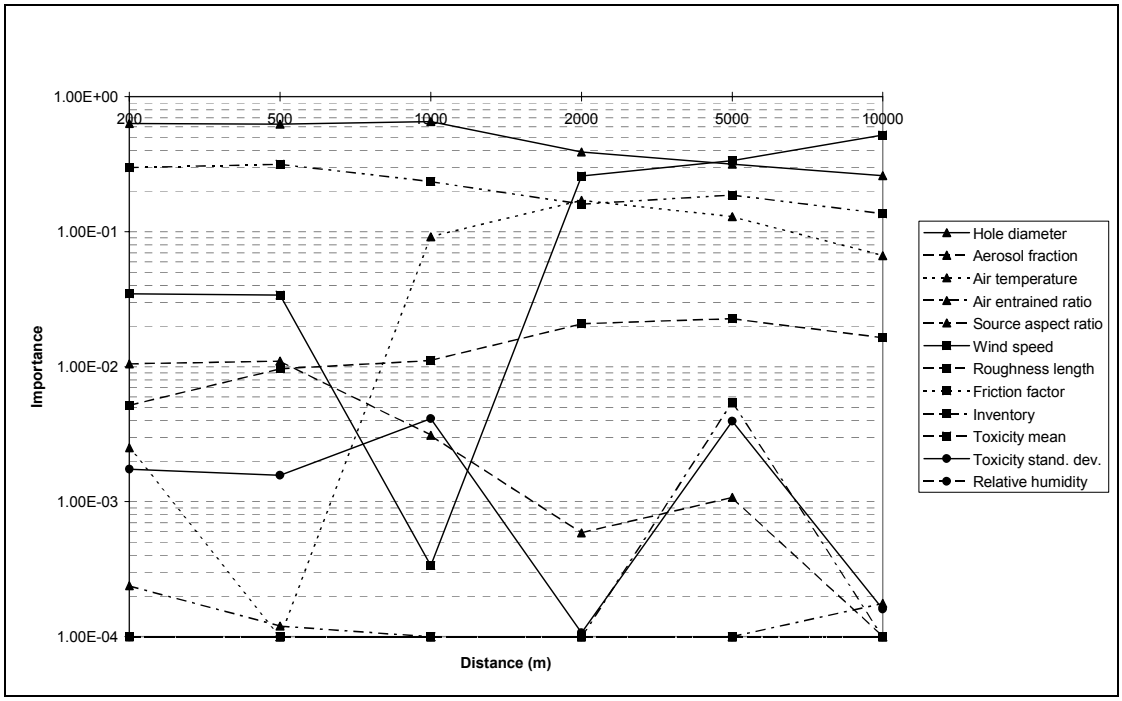

Figure 15
Variant Case 1. Importances derived from GASTAR results for Q_{lim} (fractional number of deaths) = 0.5. 'D' stability

The concentration at a dose point 2000 m from the source was calculated for wind speeds ranging from 0.1 m s^{-1} to 10 m s^{-1} and for a small range in air temperature around the fixed ground temperature of 290 K. The results are shown in Figure 16. Although the responses are reasonably consistent for v>2 m s^{-1}, they diverge below this range. The importance of the wind speed variable is therefore a very sensitive function of $\Delta T = T_g - T_a$.

It should be pointed out that dense gas dispersion models tend to be inaccurate at low wind speeds [9]. Nevertheless, it may be the case that the region of low wind speed contributes significantly to the calculated value of risk. The question of whether the consequence model can be considered valid under such circumstances is discussed in Section 7.3.

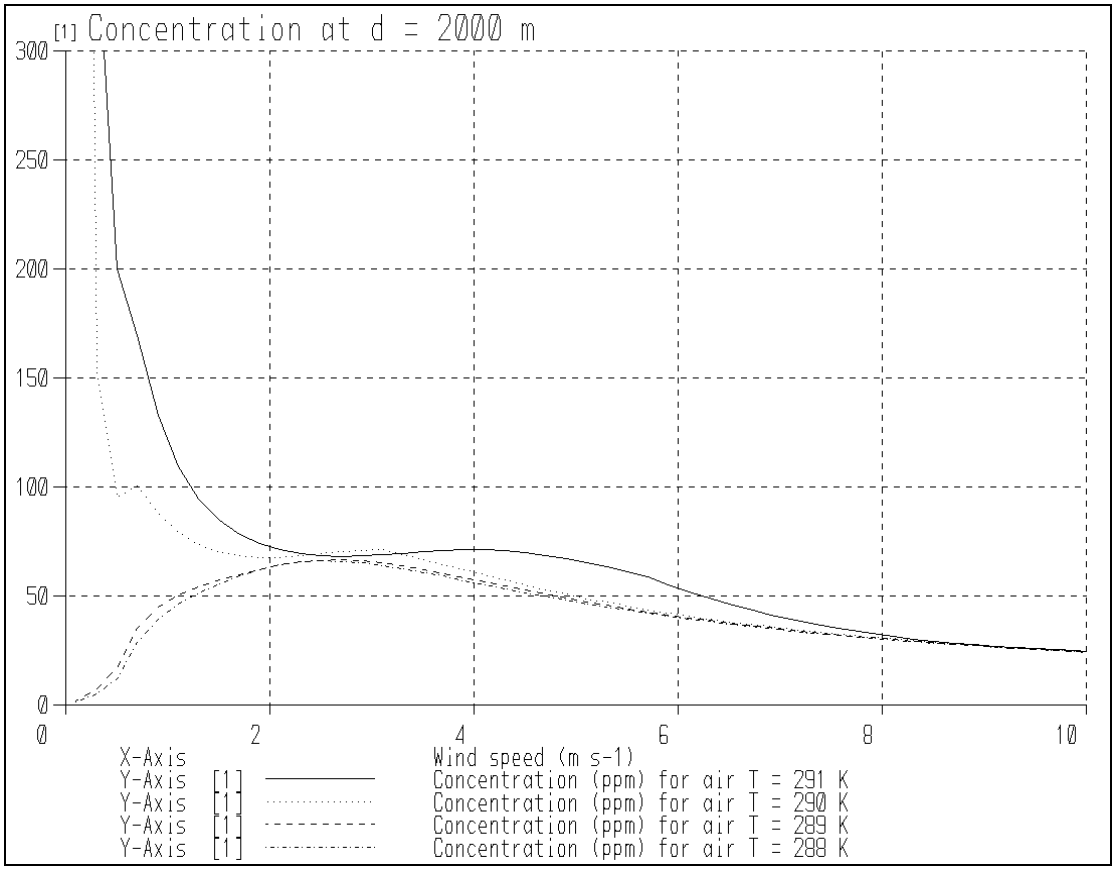

Figure 16
GASTAR: variation with wind speed and air temperature of concentration at 2000 m. Ground temperature = 290 K. Other inputs as Figure 15 2000 m MLFP

For $\Delta T > 0$, (ground warmer than air) the concentration is predicted to fall with decreasing wind speed. The FARSIDE algorithm will therefore decide that it cannot achieve any significant increase in concentration by adjustments to this variable, and therefore assign it a low importance. For $\Delta T < 0$, (air warmer than ground) on the other hand, small adjustments below about 2 m s^{-1} yield large increases in concentration, with consequent high importances for wind speed.

This behaviour accounts for the apparently inconsistent values obtained for the importances of both air temperature (Section 5.1.3) and wind speed (Section 5.1.5). For air temperatures within a degree or two of the surface temperature, and for wind speeds below about 2 m s^{-1}, the calculated concentration is highly sensitive to both the quantities, leading to a large value of dQ/dx_i and high importances. Outside this range (whose extent may vary with the other variables), the degree of variation is small.

The calculated importance of each of these quantities will therefore depend upon whether the MLFP is located inside or outside this region.

This result shows that (with GASTAR at least) the calculated dose may depend strongly upon the assumptions made regarding the relative temperatures of the air and ground. Further work would be required in order to investigate the reason for this sensitivity.

6.2 Sudden release (Variant Case 2)

Initial calculations carried out with GASTAR, Version 3.05, under the assumption of a sudden (i.e. catastrophic) release of chlorine, encountered a problem with the code which necessitated a delay until a subsequent version (3.08) was issued. This version was then used for the current calculations.

The assumptions concerning the probability distributions of the uncertain variables in the sudden release case are summarised in Table 7. They are explained in more detail below.

Table 7
Variant Case 2 (Sudden release / 'E' stability).
Assumed probability distributions

Quantity	Distribution	Parameters
Mass released (kg)	Uniform	$x_1 = 0.0$ kg $x_2 = 5\times10^4$ kg
Mass fraction of release as an aerosol (-)	Uniform	$x_1 = 0.0$ $x_2 = 1.0$
Air temperature (K)	Uniform	$x_1 = 295.0$ K $x_2 = 300.0$ K
Multiple by mass of air entrained at source (-)	Lognormal	m = 3.89 s = 0.198 (Gives $\mu = 50.0$ $\sigma = 10.0$)
Aspect ratio of the source (-)	Normal	$\mu = 1.0$ $\sigma = 0.4$
Wind speed (m s^{-1})	Weibull	k = 1.5 w = 6.0 (Gives $\mu = 5.42$ m s^{-1} $\sigma = 3.68$ m s^{-1})
Roughness length (m)	Lognormal	m = -2.65 s = 0.833 (Gives $\mu = 0.05$ m $\sigma = 0.1$ m)
Friction factor of the surface (-)	Normal	$\mu = 0.01$ $\sigma = 0.002$
Toxicity tolerance distribution mean (-)	Normal	$\mu_\mu = 14.44$ $\sigma_\mu = 1.5$
Toxicity tolerance distribution standard deviation (-)	Normal	$\mu_\sigma = 1.087$ $\sigma_\sigma = 0.15$

The mass released was assumed to be equal to the entire inventory of the tank, which was represented by the same distribution as before (i.e. uniform, between 0 and 50000 kg). The distribution of the aerosol mass fraction was kept as before. These assumptions are not entirely consistent with a real catastrophic release, where one would expect that the size of the

release and the amount of droplets carried aloft would be linked (since the vapour fraction would be more or less constant [10, 11]), but this should not be of relevance to the calculation of the importances.

Opportunity was also taken to alter some of the other probability distributions, in order to widen the scope of the survey. The distribution of the mass of air entrained was changed to have a mean of 50x and a standard deviation of 10x the mass of chlorine released, this being roughly in line with the sort of ratios typical of a catastrophic release [10, 11]. The spread of the source aspect ratio was also increased, from $\sigma = 0.2$ to $\sigma = 0.4$.

Of the atmospheric variables, the range of air temperature was shifted and narrowed in order to avoid the condition where ground and air temperatures are equal, thus avoiding the condition which gives rise to the behaviour shown in Figures 5.5 and 6.2. The wind speed distribution, which for the continuous release was typical of a site having relatively light winds, was altered to represent a slightly higher wind speed, with a mean of about 5.4 m s^{-1}. The distribution of atmospheric humidity was kept as before and the Pasquill stability class was set to 'E'.

Of the two quantities which represent the mechanical interaction between the atmosphere and the ground, the roughness length and the friction factor, the mean of the former was reduced to represent a smoother surface. The latter, which appears to have no effect, was not altered.

The distributions of the toxicity variables represent 'best estimates' and were not altered.

It is noted that, in Version 3.05, the facility existed to improve the time resolution by means of manually altering the quantity flagged in the input file as 'OutEpsilon'. Unfortunately, this facility appears to have been removed upon release of Version 3.08. Originally, it had been the intention to perform the sudden release calculations for a similar range of values of Q_{lim} and d as was used for the continuous release case. However, it was found that, for the lower values of Q_{lim} and the larger values of d, the default time points used for the integration of concentration (Equation 4.6) were too widely spaced for the integration to be accurate. An example is shown in Figure 17, which shows the dose calculated at d = 500 m as a function of wind speed, for a release of 50000 kg (since $T_a < T_g$, the dose reduces with reducing wind speed). It is understood [12] that additional time points may be specified explicitly. However, this facility was not used in the cases analysed.

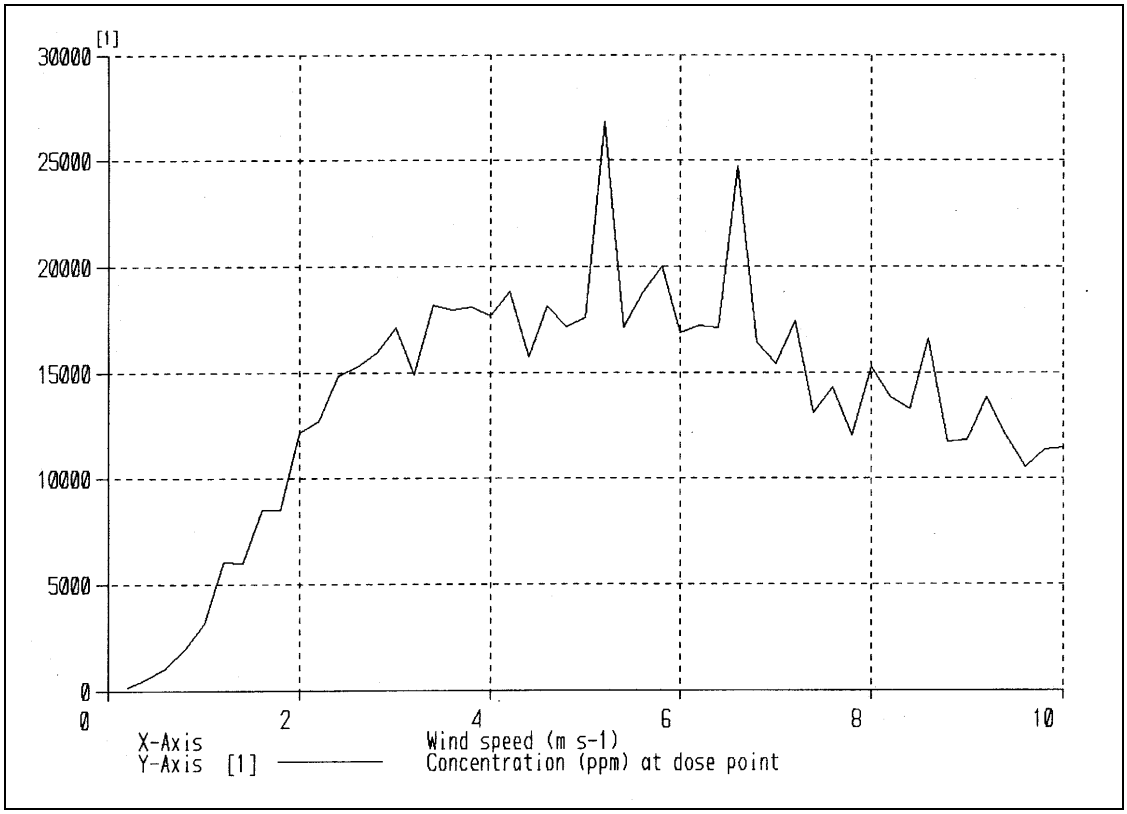

Figure 17
GASTAR: illustration of erratic calculation of dose at 500 m for sudden release

The FARSIDE calculation was therefore restricted to the case $Q_{lim} = 0.5$, with $d \leq 2000$ m. However, the results of the calculation reported in Section 5 indicate that the results of this rather limited survey can be generalised with reasonable confidence.

The results of the sudden release calculations are summarised in Figure 18. The most important input (by a significant margin) is the toxicity. The inventory (i.e. the mass released) is of relatively low importance for the near field cases ($d \leq 500$ m) since (broadly speaking) any size of release tends to give the required Q_{lim}. The importance tends to increase as the release size increases, although the fall between 1000 m and 2000 m may be due to the erratic nature of the dose prediction, as described above.

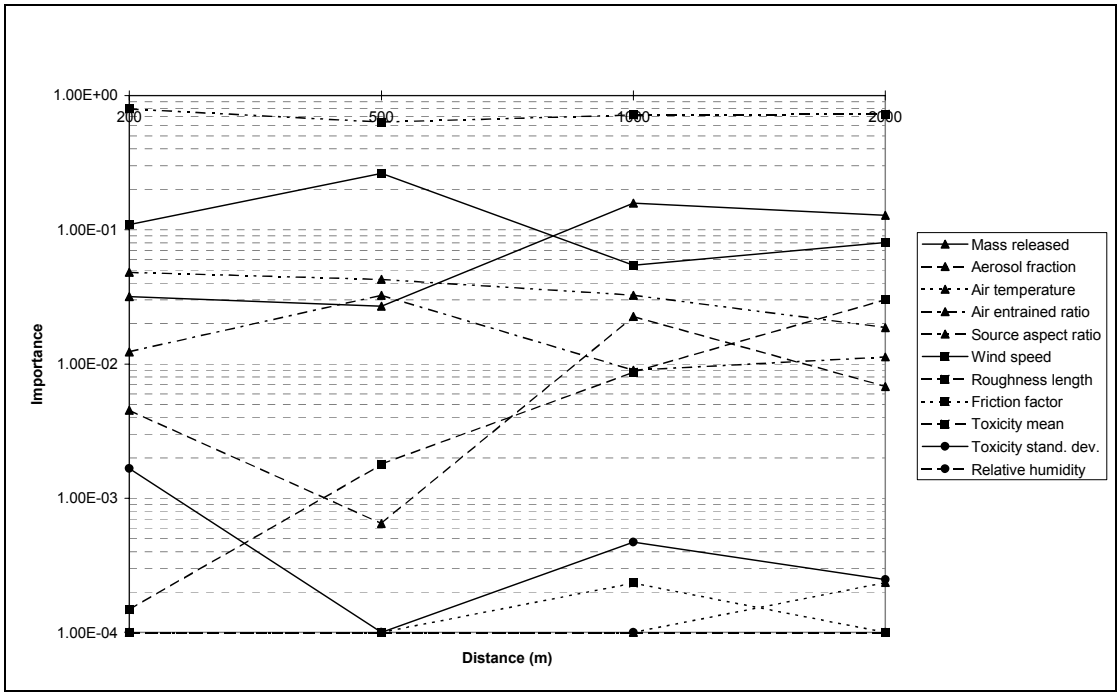

Figure 18
Variant Case 2. Importances derived from GASTAR results for Q_{lim} = 0.5

The other main variables describing the initial state of the source (aerosol fraction, air entrainment ratio and aspect ratio) are predicted as having intermediate importance. The importance of the source aspect ratio falls with distance (as one might expect), but no obvious trend is seen for the other two quantities. It should be pointed out that the standard deviations adopted for these quantities were fairly large, so it might be judged that the importances are thereby somewhat exaggerated and can therefore be ignored.

The importance of roughness length is seen to increase significantly with distance (see Ref. 3) and it is likely that this trend would continue for larger distances.

The wind speed importance increases from 200 m to 500 m, but falls thereafter. Because of previous questions raised regarding the effect of this quantity (Section 5), the variation of dose with wind speed was calculated for a 25000 kg release at d = 1000 m, with the results being shown in Figure 19. Two features of this curve are of immediate interest:

– since for this case $T_a > T_g$, the dose rises sharply for wind speeds ≤2 ms^{-1}.

– there are a number of irregularities between about 3 m s^{-1} and 8 m s^{-1}.

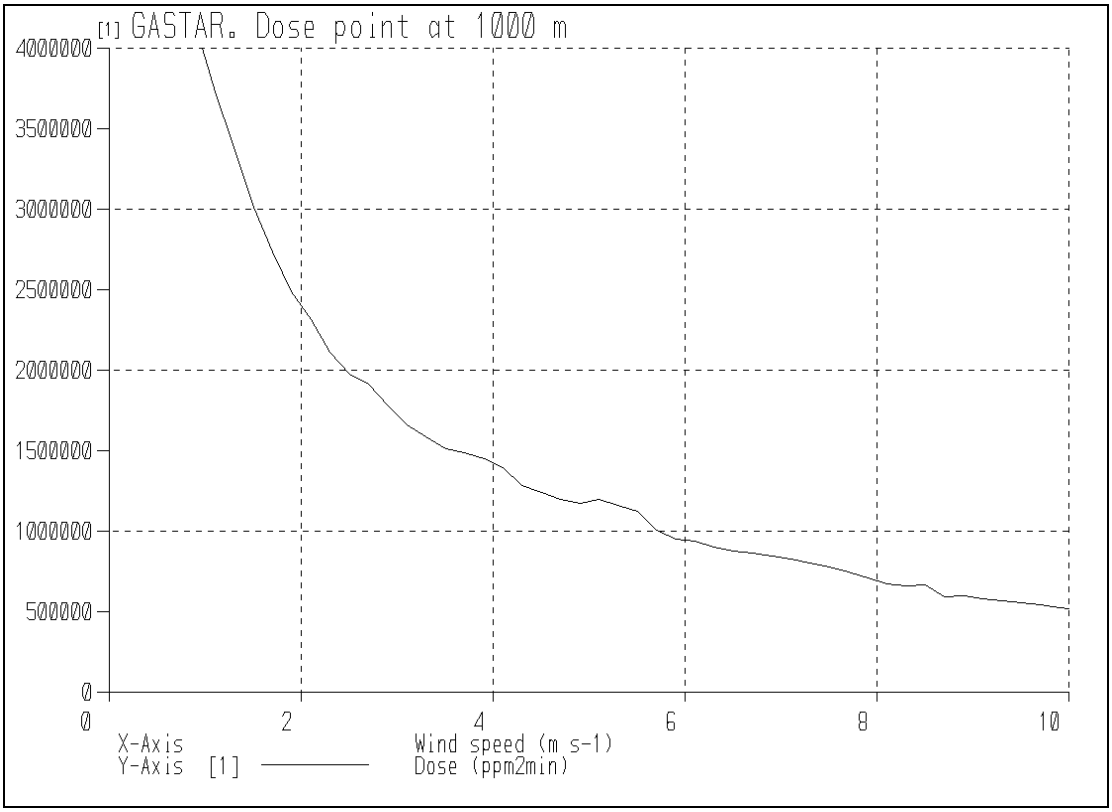

Figure 19
GASTAR: variation with wind speed of dose at 1000 m (sudden release). Other inputs as Figure 18 1000 m MLFP

The fluctuating importance of wind speed shown in Figure 18 is probably due to the variability of the gradient in the vicinity of these humps, with the average value reflecting the mean gradient. For the largest distance in Figure 18 (i.e. 2000 m), the MLFP value of wind speed is 3.87 m s^{-1}, indicating that FARSIDE has not 'seen' the rapid rise below 2 m s^{-1}, referred to above. This suggests that the use of the wind speed distribution used for the continuous release case (Section 4.5) would exhibit a higher wind speed importance, since it has a smaller mean value and would therefore tend to produce points on the steep part of the curve. The d = 2000 m case was therefore re-run with this distribution, with the results being shown in the last column of Table 8. The wind speed importance (corresponding to an MLFP value of 1.7 m s^{-1}) has increased substantially consistent with the calculations reported in Refs. 2 and 3.

Table 8
Importances for Variant Case 2 (Sudden release / 'E' stability). GASTAR results (Q_{lim} = 0.5). Comparison of results for different wind speed distributions

	Importance at distance from source (m):	
	2000	***2000****
β	**0.96**	***0.56***
P_{fl}	**1.69×10^{-1}**	***2.89×10^{-1}***
Mass released	1.28×10^{-1}	*8.06×10^{-2}*
Release temperature	1.70×10^{-5}	*2.62×10^{-6}*
Aerosol fraction	6.79×10^{-3}	*2.51×10^{-3}*
Air temperature	2.36×10^{-4}	*4.99×10^{-4}*
Air entrained ratio	1.13×10^{-2}	*1.99×10^{-2}*
Source aspect ratio	1.87×10^{-2}	*2.82×10^{-3}*
Wind speed	8.05×10^{-2}	*4.95×10^{-1}*
Roughness length	3.02×10^{-2}	*4.69×10^{-2}*
Friction factor	0	*0*
Toxicity mean	7.24×10^{-1}	*3.50×10^{-1}*
Toxicity stand. dev.	2.47×10^{-4}	*9.34×10^{-4}*
Relative humidity	1.89×10^{-5}	*1.60×10^{-5}*

* as previous column, but with 'light wind' distribution

6.3 Revised air temperature range

The sensitivity of the calculated concentration to ground-air temperature difference ΔT (defined as $\Delta T = T_a - T_g$) close to the condition $\Delta T = 0$ can dominate the calculated importances, as discussed in Section 5.1.3. Because of this, an additional set of GASTAR cases was analysed with the ΔT range chosen to avoid the region around $\Delta T = 0$.

These cases were based on the GASTAR/CRUNCH common case with the following differences:

– (with T_g fixed at 293 K) the air temperature was represented by a uniform distribution between 295 K and 300 K.

– the stability category was set to 'E'.

The first of these changes would be expected to produce a variation of concentration with air temperature typical of the right hand branch of the curve in Figure 9 (i.e. increasing smoothly with increasing T_a) and a variation of concentration with wind speed typical of the upper curve in Figure 16. The second of the changes was selected simply to increase the range of stability categories represented in the continuous release calculations (see Section 4.5). As discussed at the end of Section 4, ΔT and the stability category are closely related. The above assumptions are reasonably consistent.

The variations with distance of the importances of each of the uncertain inputs are shown in Figure 20 and Figure 21, which represent the results for $Q_{lim} = P_d = 0.005$, and 0.5 respectively.

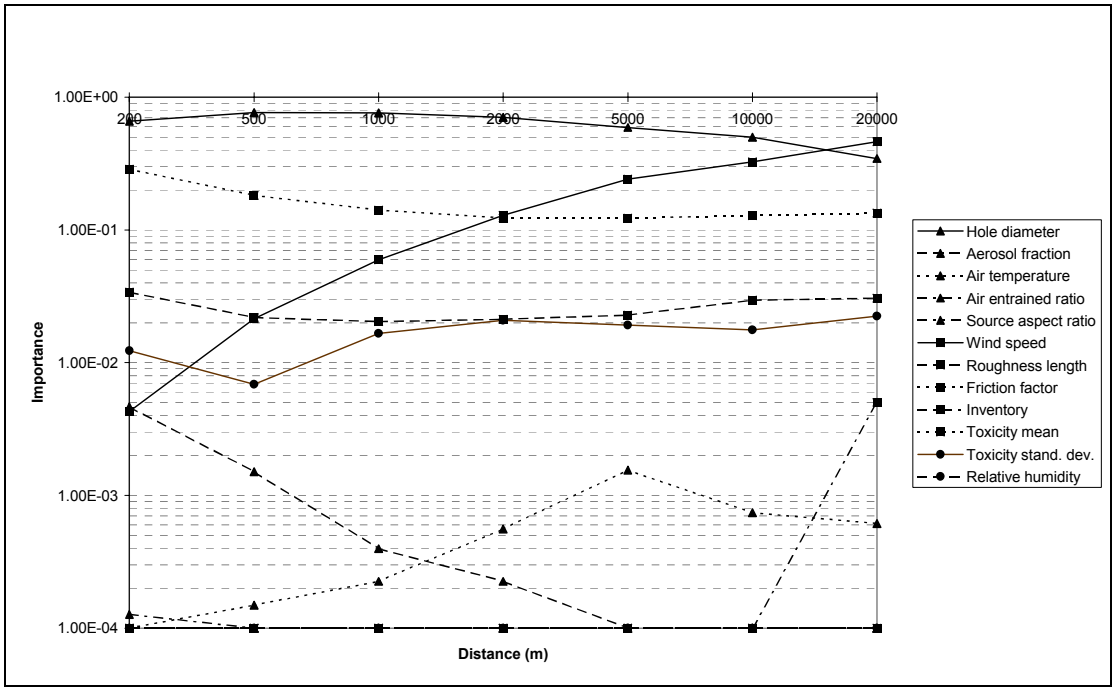

Figure 20
Variant Case 3: Importances derived from GASTAR results for Q_{lim} = 0.005

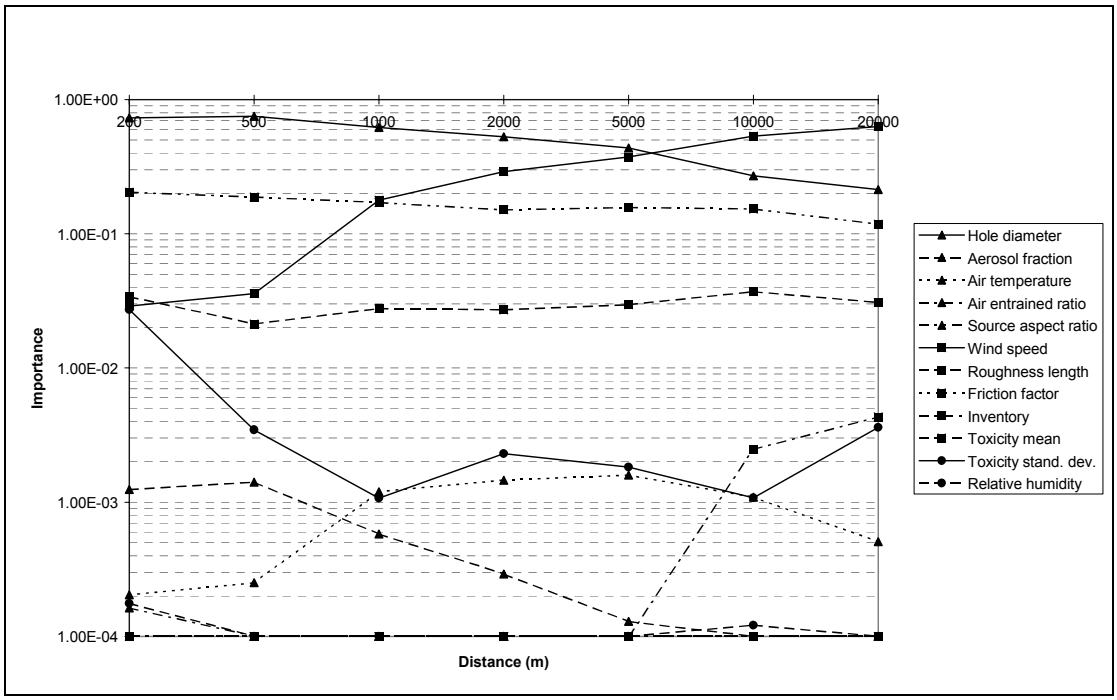

Figure 21
Variant Case 3: Importances derived from GASTAR results for Q_{lim} = 0.5

The results present a much more coherent picture of the relative importances of the various input quantities. The large and somewhat erratic values of the air temperature importance observed in Figure 5 to Figure 8 are absent, with the result that the other importances vary much more smoothly.

In both Figure 20 ($Q_{lim} = 0.005$) and Figure 21 ($Q_{lim} = 0.5$) the importances behave in similar ways. At short distances from the source, only the leak hole diameter and the toxicity mean are clearly important, with the roughness length, wind speed and the toxicity standard deviation being possibly of some significance (see Section 2.4). The importances of the other variables are clearly negligible.

As distance is increased, two effects of interest occur:

- the wind speed importance grows, eventually becoming dominant for far field dose points;

- at very long distances the importance of the inventory increases.

The latter was observed previously and has been discussed in Section 5.1.2. The former is related to the behaviour discussed in relation to Table 8. As shown in Figure 16, for $T_a > T_g$, the dose begins to rise for winds speeds below about (in the case illustrated) 2 m s^{-1} and this increase accelerates as the wind speed falls. The importance of the wind speed (which is related to the gradient of this curve) is therefore relatively low for near field dose points, where the MLFP algorithm has to perform only small adjustments to the input quantities from their median values in order to reach the condition $Q = Q_{lim}$. However, as distance increases, the algorithm finds that this condition is increasingly more easily achieved by reducing the wind speed rather than by increasing the hole diameter or reducing the mean toxicity. The MLFP values for these two cases are summarised in Tables E9 and E10 of Appendix E.

6.3.1 Discussion

The identification of the most likely failure point as occurring at very low wind speeds raises questions about the validity of the consequence model, since the type of model used in GASTAR or CRUNCH (and many other codes) is unlikely to produce accurate predictions at wind speeds below about 1 m s^{-1}. For example, assume that a QRA is carried out using conventional methods (as discussed in Section 2.2.1) and assume that most or all of the cases where toxic dose D is predicted to exceed the dangerous dose D_d occur outside the range of validity of the consequence model (e.g. at low wind speeds, as above). Should the results of the QRA be accepted? It is beyond the scope of the current work to provide a complete answer to this question, but it is possible at least to suggest that the whole QRA can be considered valid if the consequence model is valid at the MLFP, since the model will then correctly identify the position of the border between the 'pass' and 'fail' regions discussed in Section 2. If the MLFP is outside the valid range of the model, then it should still be possible to argue that the QRA is pessimistic provided that the result predicted at the MLFP is itself pessimistic. The question of the validity of the whole QRA may therefore (it is suggested) be reduced to the question of the validity of a single case.

6.4 Continuous release of ammonia (Variant Case 4)

Although most of the cases carried out during the course of this work assumed a release of chlorine, a variant case was carried out in order to investigate data importances for an ammonia release. The case was a repeat of that for $P_d = 0.005$ described in Section 5.1 (i.e. using GASTAR) but with ammonia specified instead of chlorine. This involved replacing chlorine properties with those of ammonia in the GASTAR.GPL file (Section 3.4), and altering the relevant physical properties in the routine which calculates source width (Section 4.4.3). In addition, the ammonia toxicity was represented by the same model as used for chlorine (Section 4.6). Derivation of appropriate parameters for ammonia is described in Section 6.4.1 below.

6.4.1 Ammonia toxicity

Table 18.28 of [7] (the source of the values for chlorine in Equation (4.16)) gives the following toxicity parameters for ammonia:

$$a = -35.9 \tag{6.1a}$$

$$b = 1.85 \tag{6.1b}$$

$$n = 2 \tag{6.1c}$$

which yield (see Section 4.6.2)

$$\mu_x = 22.1 \tag{6.2a}$$

$$\sigma_x = 0.54 \tag{6.2b}$$

The uncertainties in each of these quantities are set equal to the values assumed for chlorine in Section 4.6.3, i.e.

$$\sigma_\mu = 1.5 \tag{6.3a}$$

$$\sigma_\sigma = 0.15 \tag{6.3b}$$

The best-estimate values of toxicity give the following relationship (Table 9) between fraction of deaths P_d, dose D and concentration χ for a 20-minute exposure (compare with Table 1):

Table 9
Correspondence between fraction of deaths, dose and concentration for ammonia

Fraction of deaths P_d	Dose D (ppm^2 min) (Note 1)	Concentration χ (ppm) (Note 2)
0.005	1.0×10^9	7050
0.1	2.0×10^9	10000
0.5	4.0×10^9	14130
0.9	8.0×10^9	20000

Notes: 1. Assuming best-estimate toxicity
2. Assuming 20 minutes exposure

6.4.2 Results

The results for the $P_d = 0.005$ case are summarised in Figure 22. An unidentified code or system failure occurred, apparently during one of the GASTAR runs, for the case d = 10000 m, so the results were run out to a distance of 5000 m only.

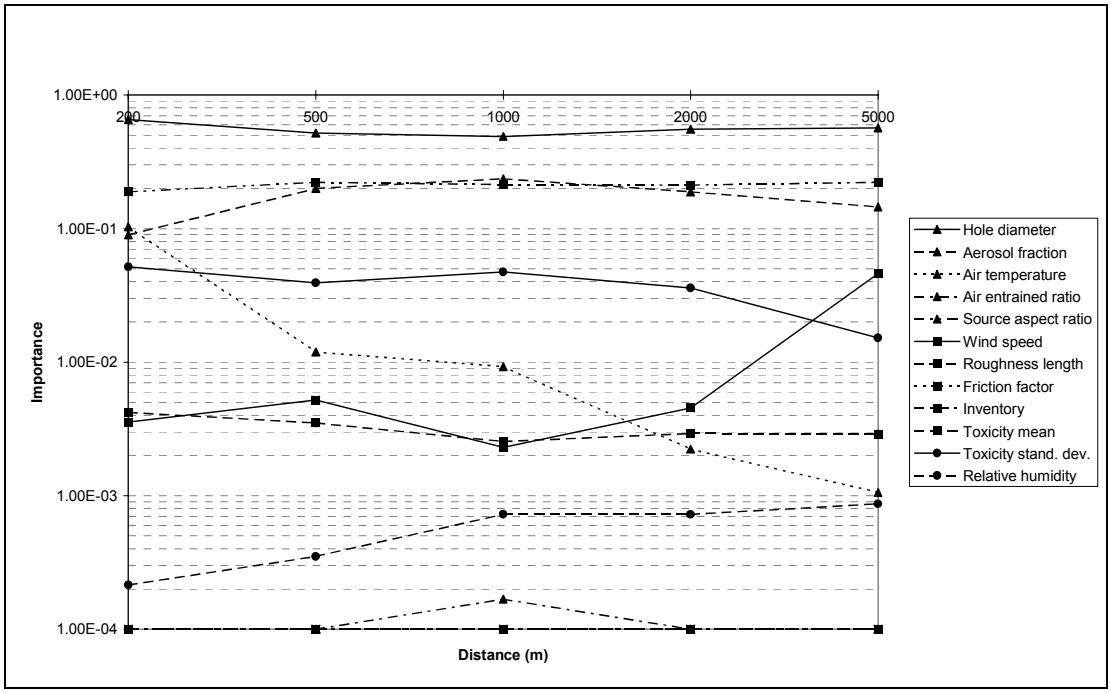

Figure 22
Variant Case 4 (continuous ammonia release). Importances derived from GASTAR results for Q_{lim} = 0.005

The results shown in Figure 22 can be compared with the corresponding chlorine case in Figure 5. It is initially evident (see Appendix E) that the lower toxicity of ammonia gives rise to a smaller value of risk (i.e. P_{fl}) for the same distance from the source. The importances, however, (with one exception) show broadly the same features as for chlorine; that is:

– High importances for release rate (hole size) and toxicity

– Low importances for source aspect ratio, air entrained ratio, humidity and inventory at low release rates.

The major difference between the ammonia and chlorine cases (compare with Figure 5) is the much higher importance for aerosol fraction, which for ammonia is comparable with that of the toxicity. The reason for this is probably due to the significantly different physical properties of ammonia, which has

– a much lower density at ambient temperature and pressure

– a higher vapour specific heat capacity

– a higher liquid-vapour latent heat

than chlorine, and tends to form persistent aerosols. It should again be pointed out that the spread assumed for the distribution of aerosol fraction is somewhat in excess of that which might be expected physically (see discussion in Sections 4.4.1 and 5.1.2).

7. DISCUSSION AND CONCLUSIONS

7.1 General comments

This report is concerned with the calculation of risk arising from an installation containing a toxic volatile material (particularly chlorine). It describes how the most likely failure point (MLFP) method can be used to estimate the degree to which uncertainties in the values input to a consequence model influence the predicted value of risk.

One of the ways in which this can be expressed is by means of the importance I of each variable. The importance is an approximate normalised measure of sensitivity which is proportional to the fractional reduction in the calculated risk which would occur if the variable was replaced by a fixed (best-estimate) value. Importances are produced as a by-product of the MLFP calculation.

Based upon experience gained in this and previous work, the following procedure is proposed for the interpretation of the importance information. For each variable i; i=1..N, where N is the total number of uncertain variables represented in the analysis:

1. If $I_i \leq 5\times10^{-3}$, the uncertainty in the input quantity is of no significance. The uncertain variable may be replaced by a best-estimate value.

2. If $5\times10^{-3} \leq I_i \leq 5\times10^{-2}$, the uncertainty may be of significance. The influence on risk should be checked by an appropriate sensitivity formula (Section 2)

3. If $I_i \geq 5\times10^{-2}$, the uncertainty is having a significant influence on the calculated risk. The user should determine if this is a genuinely variable quantity (e.g. wind speed or release rate). If not, then it may be worthwhile trying to narrow down the uncertainty in this parameter.

7.2 Calculation of importances for two dense gas dispersion models

The procedure outlined above was used to evaluate input variable importances for sample risk assessments using the following dense gas dispersion codes:

– GASTAR

– CRUNCH

The risk assessments were carried out over a range of conditions in order, as far as possible, to identify general behaviour; that is to identify

– variables which are always important

– variables which are always unimportant

– variables which are sometimes important

to the risk estimate, and to determine the circumstances of relevance to the last category. Probability distributions assigned to the model input quantities were intended to represent typical degrees of uncertainty for each variable.

7.2.1 Quantities which are always important

Under the assumptions made during the course of the calculations, the following uncertain variables were found to be important in all the test cases:

– the release rate distribution (for continuous releases)

– the toxicity

The results indicate that the uncertainty in the mean of the population tolerance to toxic dose can have a significant effect on the calculated risk.

The importances of the distributions of wind direction and atmospheric stability were not covered during the work, the latter because the consequence models could not fully represent stability as a continuous variable (see discussion in Section 4.5). The former may be regarded as important under most circumstances. The latter requires further investigation.

7.2.2 Quantities which are unimportant

The following quantities were found to be unimportant in the cases studied:

– the ground friction factor

– the relative humidity

7.2.3 Quantities which are sometimes important

The test cases indicated that the majority of the input variables tended to be important (or, at least, not unimportant) under some circumstances. The following lists these variables and then attempts to identify the circumstances under which each may be significant. The variables are:

– wind speed

– air temperature (for fixed ground temperature)

– ground roughness length

– source aerosol fraction

– amount of air entrained at source

– source aspect ratio

– tank inventory

The unexpected and very strong influence of the ground-air ΔT upon GASTAR predictions at low windspeed has been discussed in Sections 5.1 and 6.1 and, until this issue is investigated further, it is difficult to draw conclusions as to the importances of wind speed and air temperature. A variant case (Variant Case 3; Section 0) was run with the ΔT range set to avoid this problem. The case demonstrated much more coherent behaviour of the importances. Based on this and the results of previous work [2,3], it would be anticipated that the wind speed distribution would be important, with the importance increasing for dose

points located a significant distance from the source. These circumstances would also be expected to give rise to a higher importance of ground roughness length, although, in the cases studied here, this was only evident in Variant Cases 2 and 3 (Sections 6.2 and 0).

The source aerosol fraction was only seen to be significant in the case of the flashing ammonia release, possibly because the relatively large latent and vapour specific heats (compared with chlorine) give rise to a persistent aerosol and hence a relatively cool and dense cloud, compensating for the lower vapour density of ammonia at ambient temperature.

Both the amount of air entrained at source and the source aspect ratio tend to be unimportant for the continuous release cases. CRUNCH predicts a relatively high importance value (0.1) for the source aspect ratio in one instance (Figure 13 at 200 m), but the results at other distances, taken together with those produced by GASTAR, suggest that this is spurious. The general question of the identification of such rogue results is clearly an important one. With the current evidence, it is reasonable to suspect that rogue results are caused mainly by discontinuous behaviour within the consequence model, as discussed elsewhere (see Section 5.1.3) and can be identified because of their inconsistency with the results of (seemingly) comparable cases. As discussed in Section 7.1, if the evidence nevertheless remains ambiguous (i.e. if the user, after consideration of the calculated importances, remains unsure as to the significance of any particular input quantity), the matter should be investigated further by means of a sensitivity case (Section 2).

The somewhat higher importances for the source properties predicted for the sudden release case (Section 6.2) refer to the large uncertainties assumed (compare Table 7 with Table 4). Although these importances, by the criteria of Section 2.4, cannot be completely ignored, the indication is that it should be possible to use best-estimate values for these quantities without significantly affecting the risk calculated.

The importance of the tank inventory depends upon the type and duration of release. For continuous releases, the release was assumed to end after an intervention time of 20 minutes, if it had not ended already due to depletion of the tank contents. Therefore, the importance of the initial inventory is only nonzero if the release rate at the most likely failure point (MLFP) is such that the latter occurs. Since the MLFP value of release rate will rise with increasing distances from the source, this implies that (as observed) the inventory importance will remain zero up to a threshold distance, and then increase. For sudden releases, the importance of the inventory (corresponding in this case to the mass released) increases with increasing distance of the dose point from the source without displaying any threshold behaviour.

7.3 Comments on the consequence models

During the calculations performed for this study, the following features of the consequence models were noted:

CRUNCH:

- the output table (.FCO file) results contained very few significant figures (Section 3.3)

- the results for the start of the passive part of the plume vary discontinuously with (for example) changes in release rate (Section 3.3)

- the variation of the predictions of concentration with ground roughness length exhibits discontinuities at roughness lengths of 0.005 m and 0.025 m (Section 5.2)

GASTAR:

- there is a strong interaction between the ground-air temperature difference and the predictions for low windspeed (Sections 5.1.3, 6.1). This was not expected and the explanation was not discovered during the course of the work.

- in Version 3.08, the default time points used in the integration of the cloud concentration at a fixed point are too widely spaced for the integration to be accurate (Section 6.2), though further points may be specified explicitly if required.

Most of the above were discovered because of their effect on the convergence of the FARSIDE search algorithm (Appendix A).

The sensitivity of the MLFP method to such characteristics of the consequence model can be seen as a disadvantage, compared with the more robust behaviour of the other methods of calculating risk (Section 2). However, an alternative view could be that this property of the MLFP method can be used to detect undesirable features of the consequence model (such as discontinuous behaviour) which might otherwise escape notice. Although it is not (of course) suggested that consequence models should be tailored to suit the requirements of the MLFP method, it is considered that the desirable features of a model from the point of view of the MLFP method coincide with those which are desirable from the point of view of a user. In particular, it is considered that any consequence model should possess the property that its predictions are a continuous function of its inputs, unless:

- there is some underlying physical reason why this should not be so, or

- there is a good practical reason why this cannot be achieved.

It is therefore suggested that the MLFP method (together with figures of the type shown in Figure 8) could be used as part of the commissioning process for a consequence model.

7.4 Model validity

In addition to the requirements discussed above, there is the general question of whether any consequence model is valid (i.e. a suitably close representation of reality) for the use to which it is put. This is a particularly difficult question to answer in the case of a full risk assessment, since the assessment may include cases (such as those for low wind speed) where the model is not fully validated.

It is proposed that the following rule can be applied: a model is valid for a risk calculation if it is valid at the most likely failure point (whose location will depend on the distance from the source). This rule may be applied even if the actual risk analysis is performed by one of the other methods. If the model is *not* valid at the MLFP (for example, if the MLFP value of the wind speed is below the valid range), then further justification may be required to show that the risk analysis is sound (Section 6.3.1).

7.5 Additional work

The work carried out under this contract also included the following:

- restructuring and preliminary documentation of the FARSIDE software, which was developed for MLFP analysis in the previous projects funded by HSE,

– investigation of suitable methods for processing raw atmospheric data measurements to produce smooth multidimensional frequency distributions.

The results of the FARSIDE task are reported in Appendix A. The work described in Appendix B indicates that a suitably smooth distribution can be obtained from a scatter plot of simultaneous wind speed and stability measurements by means of a technique known as the adaptive kernel method. This method represents each point in the plot by a region of finite size, with the overall distribution being made up of summed contribution from all such regions. The method as developed in Appendix B proposes the use of transformed variables based on the wind speed and the Monin-Obukhov length in order to best represent the shape of the scatter plot and the derived distribution. However, further work is required in order to provide a reliable means for optimising the smoothing.

8. REFERENCES

1 Mitchell, B. Comparative assessment of advanced techniques for the evaluation of confidence levels in calculated safety margins. HSE Contract Research Report No. 99/1996. 1996.

2 Mitchell, B. Use of the FORM/SORM (most likely failure point) method for uncertainty analysis. HSE Contract Research Report No. 142/1997. 1997.

3 Mitchell, B. Use of the FORM/SORM (most likely failure point) method for quantitative risk assessment. WSA Report AM5123-R1. 1998.

4 Fitzpatrick, R.D. Modification of CRUNCH to meet the needs of HSE casework. HSE Report IR/L/HA/83/5. 13 October 1983.

5 GASTAR Version 3. User Manual. Cambridge Environmental Research Consultants Ltd. 18 June 1998.

6 Digital Visual Fortran. Digital Equipment Corporation. 1998.

7 Lees, F. Loss prevention in the Process Industries. 2nd Edition. Butterworth. 1996.

8 Guidelines for chemical process quantitative risk analysis. AIChE Center for Chemical Process Safety. 1989.

9 Lines, I.G. and Daycock, J. Guidelines for the inclusion of low windspeed conditions into risk assessments. WSA Report No.AM5128-R1. March 2000.

10 Gilham, S. et al. Modelling of catastrophic flashing releases of liquid. WSA Report AM5210-R1. June 1999.

11 Shepherd, A.M. Source term calculation for flashing releases. Background to the development of ACE. WSA Report No. AM5233-R1. June 2000.

12 Private communication. P. Appleton (HSE) to B. Mitchell (WSA). 14 July 2000.

APPENDIX A

THE FARSIDE PROGRAM

A1. INTRODUCTION

FARSIDE was written by WS Atkins during the course of work funded by HSE. It performs a most-likely failure point (MLFP) calculation using an algorithm developed during the course of the original work. The basic goal of the algorithm is simple, as described in Section 2 of the main report. However, as with many numerical methods, the bulk of the method is taken up with add-ons designed to cope with extreme situations which may cause difficulties with the basic approach.

A2. BASIC APPROACH

In order to locate the MLFP, it is necessary to find the point on the failure surface which is nearest to (i.e. is the smallest number of standard deviations from) the best estimate point (BEP); that is: the point defined by the median values of all the uncertain variables. This process is known as a *constrained minimisation* procedure. It is exactly analogous to the problem illustrated in Figure A1, which represents a contour map of a hilly region, but showing only a single contour C (representing the failure surface) (note however that the actual method works in any number of dimensions).

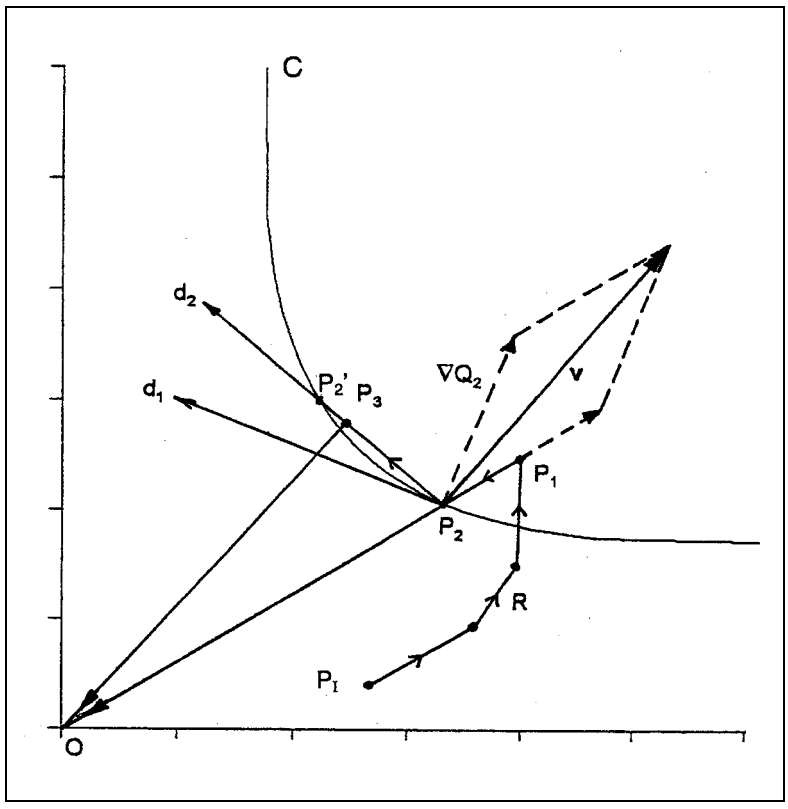

Figure A1
Illustration of search sequence in the FARSIDE constrained minimisation scheme

The problem illustrated by Figure A1 is to find the point of closest approach of the contour C to the point O, where the only information available to the method is the location (i.e. the

coordinates) and elevation of each point visited. It is further assumed that the determination of the elevation (equivalent to the evaluation of the consequence model) has a cost penalty associated with it, so that the number of points visited should be kept to a minimum.

A2.1 Location of the failure region

Assume, for the sake of illustration that the height of C is above that of O. The method used by FARSIDE in this case is as follows. The first goal is to find *any* point P_1 such that C lies between P_1 and O. Therefore, starting from an initial point P_I (which may be at O), the method steps a chosen distance in the first coordinate direction and records the coordinates and elevation of the new point. It then returns to P_I and steps in the other coordinate direction and repeats the process, returning to P_I afterwards (for N dimensions, this process occurs N times). From these measurements, the direction in which the gradient is steepest is calculated, and the method steps a selected distance in that direction. The height is then checked to see if the new point lies above the contour. If not, the process is repeated until this goal is achieved.

It should be clear that even this initial procedure contains a number of potential difficulties:

1. What is a suitable distance to step in each coordinate direction in order to get a good estimate of the gradient ?

2. What is a suitable distance to move along the direction of steepest gradient to look for the next point ?

FARSIDE currently uses 0.02 (standard deviations) for the former, though this can be altered (in Routine PROB in Module USER; Section A3) if the user believes that 'noise' in the consequence model is causing problems (see Sections 5 and 6 of the main report). For the latter, the distance is first calculated such that a gain in height of the required amount will be achieved. Where the gradient is locally small, this can lead to excessively large steps, which can throw the search point into the wrong region. In order to avoid this, the distance is limited to $0.5N^{1/2}$, where N is the number of dimensions. In the above example, N=2, so the distance limit is about 0.71.

With the above constraints on distance, it has been found in all problems studied so far that the algorithm finds the failure region quickly and reliably. In rare cases, the gradient at the default start point (the BEP) may be zero, meaning that the search cannot get underway. In this case, the user can specify a new start point (in Routine START in Module USER) which should preferably be some (very rough) guess as to the location of the MLFP.

A2.2 Location of the failure surface

Returning to the example, once a point on the far side of the contour has been found, the next goal is to find any point on the contour itself. It would be possible to search along the direction just travelled from the last point, but a better approach is to calculate the direction back to the point O and search along that. It may be noted that, since the location is known to be between P_1 and O, the eventual success of the search is assured. The only question is that of its efficiency.

The search for the contour has to cope with the possibility of a range of topographies along the search direction. It may be that there is a constant gradient between P_1 and O. Alternatively, P_1 may be on a plateau, O may be in a valley with a flat floor and C may run along the side of a cliff. The search method first assumes the former. The method steps an appropriate distance (again 0.02 at present) towards O from P_1, calculates the elevation and

estimates the position of P_2 from the gradient (a quasi-Newton iteration scheme). This scheme is continued until P_2 is found (which will occur rapidly if the scheme works) or until either:

– the limit to the number of iterations (10) is reached, or

– the scheme produces an estimate which is outside the current estimate of the region where P_2 is located (P_2 is known initially to be somewhere between O and P_1. This uncertainty should reduce with each new estimate).

If this scheme fails, then the method is switched to one where the next estimate of the position of P_2 is obtained by interpolation between the current best estimates on either side of the path. Although this scheme always produces an improved estimate, it can be very slow to converge in the case (see above) where C runs around a cliff. If the number of iterations is again exceeded, the third method: that of repeated bisection of the current range in which P_2 is known to lie, is used.

The third method above will eventually be successful, but the formulation of the measure of success is not as straightforward as it may appear. The default method is to test if the elevations of the two points which straddle the contour are sufficiently close. However, if the consequence model contains discontinuities (see Sections 5 and 6 of the main report) this may never be achieved. If convergence as defined by this first method fails (i.e. if the number of iterations is exceeded) the search is nevertheless considered to have succeeded as long as the two points are sufficiently close together.

A2.3 Location of the most likely failure point

The contour (failure surface) has now been located, and its point of closest approach to the point O must now be found. Again only the positions and the elevations of the points visited are known. The contour cannot be followed directly.

The next step, once this point is reached, is to calculate the gradient again, using the same method as before. Once this has been carried out, the class of constrained minimisation techniques known as projected gradient methods move off in a direction d_1 (Figure A1) tangent to the contour (which is normal to the direction of steepest ascent: ∇Q_2) for a suitable distance, then return to the path, and repeat this procedure until the closest approach point is found. FARSIDE uses a variation of this procedure which (it is believed) is more efficient.

Instead of using ∇Q_2 as the normal direction, FARSIDE takes v, the mean of ∇Q_2 and the direction through P_2 from the origin. These two directions define a plane (whatever the total number of dimensions) and the next search direction (d_2) represents the line within this plane which is normal to v and which passes through P_2. If the shape of the failure surface is convex towards the origin, then this line will form a chord which cuts the failure surface at two points (denoted P_2 - P_2'). If the surface is concave towards the origin, point P_2' may not exist.

The next search point P_3 is initially calculated as the closest approach that the line d_2 makes to O. The coordinates of this point can be calculated analytically once the equation of the line d_2 has been determined. The new point P_3 should lie in the fail region. However, if

(a) the failure surface is strongly curved near P_2, or

(b) the consequence function is particularly noisy, so that an accurate estimate of the local gradient cannot be made

then it is possible that P_3 may be located either (a) beyond P_2' or (b) in the wrong direction, such that it is back in the pass region. If P_3 is found not to be in the fail region (as it should be), the following recovery procedure is adopted: the distance from P_2 to P_3 is halved and a new point P_3' is generated. A quadratic is then fitted through the values of elevation at these three points and the point of highest elevation found analytically. If this point is within the fail region, and if it is closer to P_2 than to P_3, then it is used as the new P_3. If not, the point is used as the new end point, the distance is halved again and the process repeated. Once a suitable P_3 has been found, the search continues as from P_1.

In the case where the above recovery procedure is caused solely by a strongly curved, but otherwise well-behaved failure surface (Case (a)), then the procedure works fairly well. However, there are other occasions, when P_2 is close to the MLFP, where, due to rounding or other difficulties in calculating the direction ∇Q or d_2, the direction d_2 can be tangent to the failure surface or may even form a chord in the wrong direction (Case (b)). In this case, the recovery procedure will not work since all points lying closer to O than P_2 are in the pass region and have an elevation less than that of the contour. If this is the case then it is likely that no significant improvement can be achieved, and the search may as well consider the current point to be the MLFP (though the user has the option of carrying on).

In cases where the consequence function is reasonably well behaved, then the minimum point (or rather, *a* minimum point) has been found when the line from the origin to the point P on the failure surface lines up precisely with the gradient ∇Q at P (this is the condition contained in the Lagrange multiplier method of constrained minimisation). FARSIDE therefore considers the minimum to be found when the cosine of the angle between these two vectors is sufficiently close to 1.0 (an error of 10^{-4} has been used as default, though this tolerance can be increased substantially in difficult cases without causing any significant error in the results)

When convergence has been achieved, there remains the problem of deciding whether the minimum that has been found is a global minimum, or just a local one. Experience so far indicates that local minima are much rarer than might be expected. However, it is desirable that the user should have some means for testing whether the minimum is global or not. FARSIDE has therefore been modified to allow the user either to accept the calculated MLFP or to try the search again from a different start point. The user is given three options:

- Accept the MLFP
- Restart the search from a specified distance away in a randomly chosen direction
- Restart the search at a user-specified point

Table A1 below summarises the details of the iteration schemes described in the text above.

Table A1
Summary of search methods in FARSIDE

Process	Method	Criterion	Iteration limit (L)	Tolerance
Finding failure region	Steepest ascent/descent	Define $s = Sign(Q_{lim}-Q_m)$; Failure region found when $Q \geq sQ_{lim}$	L=20	Not applicable Maximum jump $= n^{1/2}/2$ Minimum jump = 0.1
Finding failure surface	1. Quasi Newton 2. Interpolation 3. Bisection	Define a typical size for the function Q, as $Q_{typ}=Max(Abs(Q_{lim}), Abs(Q_{lim}-Q_m))$ and let d_f = Furthest point from origin on pass (origin) side; d_n = Nearest point to origin on failure region side. Then if: 1. $d_f \leq d \leq d_n$ and 2. $Abs(d-d_{last}) < \varepsilon_1$ and 3. $Abs(Q-Q_{lim}) < \varepsilon_2 Q_{typ}$ the failure surface has been found. If L is exceeded, it is first assumed that condition 3. has not been satisfied due to a discontinuous Q. If $(d_n-d_f) < \varepsilon_1$, it is assumed that the failure surface has nevertheless been found.	1. $L_1=7$ 2. $L_2=10$ 3. $L_3=10$ Combined: L=22	$\varepsilon_1 = 10^{-3}$ $\varepsilon_2 = 10^{-3}$
Finding MLFP	FARSIDE variant on projected gradient method	If ∇Q_n is the normalised gradient at the current failure surface point and $\underline{u}_n$ is the normalised position of the point, then the MLFP is considered to be found when $\lvert(\nabla Q_n \bullet \underline{u}_n)\rvert > 1 - \varepsilon_1$ Where this convergence does not occur, the MLFP is assumed to be found when $\Sigma\delta d_i < \varepsilon_2$, where δd_i is the change in d between two successive failure surface points. The sum is taken over N steps.		$\varepsilon_1 = 10^{-3}$ $\varepsilon_2 = 3\times10^{-3}$ N = 3
Calculating ∇Q	Finite difference	$(\partial Q/\partial u_i) \approx (Q(u_i + \delta)-Q(u_i))/\delta$		$\delta_g = 0.02$
Calculating $\partial Q/\partial L$ along direction $\underline{d}$	Finite difference	$\partial Q/\partial L \approx Q(\underline{u} + \delta\underline{d}_n)/\delta$ where $\underline{d}_n$ is a normalised version of $\underline{d}$		$\delta_g = 0.02$

A3. STRUCTURE OF FARSIDE

FARSIDE is written mainly in standard Fortran 77, with additional coding specific to Digital Fortran to allow 'calls' to stand-alone consequence models (see Section 3 of the main report). Table A2 gives the names of the modules which make up FARSIDE and Tables A3 - A10 summarise the roles of the individual routines within each module.

Table A2
FARSIDE modules

Module name	Task
FARSIDE	Main control and output routines
FINDERS	Routines controlling search for most likely failure point
PRBDIS	Evaluation of probability distributions
PPREPRO	Preprocessing of user specifications for probability distributions
FORM	Calculates importances and sensitivities for FORM approximation
SORM	Calculates failure surface curvatures and SORM integral
INTER	Provide default search control parameters and call user routines
USER	User-written routines to (a) specify probability distributions for uncertain variables and (b) provide consequence model

Table A3
Summary of routines in Module FARSIDE

Module Name	Routines	Description	Calls	Called by
FARSIDE				
Part 1		**Main routine**		
	FARSIDE	Controls logic of calculation	PROBX DISTRB CONX LINCOM STARTX DIST HIST FINDFR SEARCH AMIN RECOV SUMMH SUMMC SUMMS JUMP JOIPAR FSENS GRAMS CURVE SUMMK SORM SORM1 SORM2 SORM3 SORM4 SCROLL	None
Part 2		**Restart point routines**		
	JUMP	Gets the user's choice as to whether to accept MLFP or start the search again from a new point	RANDM SCROLL CONX	FARSIDE
	RANDM	Random number generator	None	JUMP
	SCROLL	Scrolls the screen	None	Various
Part 3		**Output routines**		
	SUMMC	Writes a summary of the calculation cost	None	FARSIDE
	SUMMS	Writes a summary of the MLFP position and the level of convergence obtained	CONX GANCF DOTP	FARSIDE
	SUMMK	Writes a summary of the curvatures	None	FARSIDE
	PTABLE	Summarises the conditions at the current point in the search	None	FINDFR FINDFS
	GTABLE	Writes the gradient at the current point	None	FINDFR FINDFS
	HIST	Stores the information about the current point in the search	None	FARSIDE FINDFR
	SUMMH	Writes a table of the search history	None	FARSIDE

Table A4
Summary of routines in Module FINDERS

Module Name	Routines	Description	Calls	Called by
FINDERS		**Routines controlling searches**		
Part 1		**Main control routines**		
	FINDFR	Finds failure region from start point	PTABLE GTABLE LINCOM CONX DIST HIST GRAD	FARSIDE
	FINDFS	Finds failure surface	PTABLE GRAD LINCOM NEWTON LINTER BISECT CONX GTABLE	FARSIDE
	SEARCH	Calculates the new search direction from a point on the failure surface	DOTP	FARSIDE
	RECOV	Finds failure region if closest approach calculation overshoots	CONX LINCOM	FARSIDE
Part 2		**Point finders**		
	NEWTON	Performs a quasi-Newton step along a selected direction to try and find the failure surface	GRADLN	FINDFS
	LINTER	Performs a linear interpolation step along a selected direction to try and find the failure surface	None	FINDFS
	BISECT	Performs a bisection step along a selected direction to try and find the failure surface	None	FINDFS
	AMIN	Calculates the point of closest approach of a line to the origin	None	FARSIDE
Part 3		**Gradient calculation**		
	GRAD	Calculates ∇C numerically, where C is the constraint function	GRAD1 or GRAD2	FINDFR FINDFS
	GRAD1	Calculates one-sided gradient of consequence function	CONX	GRAD
	GRAD2	Calculates two-sided gradient of consequence function	CONX	GRAD
	GRADLN	Calculates dC/dL numerically (one-sided) along a specified direction (L=length)	CONX	NEWTON
Part 4		**Utility routines**		
	DOTP	Calculates dot product of two vectors	None	FARSIDE FINDFR FINDFS SEARCH AMIN
	DIST	Calculates the distance between two points	None	Various
	LINCOM	Forms the linear combination of two vectors	None	Various

Table A5
Summary of routines in Module FORM

Module Name	Routines	Description	Calls	Called by
FORM		**First-order calculation routines**		
	FSENS	Calculates sensitivities of β and P_{fi} to μ and σ	GAND	FARSIDE

Table A6
Summary of routines in Module SORM

Group Name	Routines	Group/Routine Description	Calls	Called by
SORM		**Second-order failure probability routines**		
Part 1		**Curvature routines**		
	GRAMS	Produces an orthogonal axis set at MLFP using Gram-Schmidt procedure	None	FARSIDE
	CURVE	Calculates the curvatures of the failure surface	CONX	FARSIDE
Part 2		**Exact SORM integral**		
	SORM	Exact numerical integral	Y SINMIN	FARSIDE
	SINMIN	Finds saddle point of SORM integrand surface	Y	SORM
	Y	Calculates the real part of the SORM integrand	None	SORM SINMIN
Part 3		**Approximate SORM integrals**		
	SORM1	Breitung ψ formula	GANCF GAND	FARSIDE
	SORM2	Breitung β formula	GANCF	FARSIDE
	SORM3	Exponential ψ formula	GANCF GAND	FARSIDE
	SORM4	Exponential β formula	GANCF	FARSIDE

Table A7
Summary of routines in Module PRBDIS

Group Name	Routines	Group/Routine Description		Calls	Called by
PRBDIS		**Calculation of probabilities**			
Part 1		**Univariate distributions**			
		Routines are in groups of 4, with each group corresponding to a type of distribution. The first 3 letters (prefix) of the routine name signifies the distribution type and the following letters (suffix) give the quantity calculated:			
		Distribution types:			
		Prefix	Distribution Type		
		GAN	Gaussian (Normal)		
		GAL	Log-normal		
		UNI	Uniform		
		T2L	Type II largest value		
		WEZ	Weibull (Minimum, $\tau = 0$)		
		TRP	Trapezoidal		
		ELL	Elliptical		
		XPO	Exponential		
		XPC	Truncated exponential		
		H1D	Histogram		
		Quantity calculated			
		Suffix	Quantity Calculated		
		PAR	Mean and standard deviation		
		D	Probability density p=f(x)		
		CF	Cumulative probability P=F(x)		
		CI	Inverse probability $x=F^{-1}(P)$		
Part 2		**Multivariate (histogram) distributions**			
		The naming convention is analogous, as follows:			
		Distribution types:			
		Prefix	Disttribution Type		
		H2D	2-D histogram		
		H3D	3-D histogram		
		Quantity calculated			
		Prefix	Distribution Type		
		PAR	Mean and standard deviation		
		CF	Cumulative probability		
		CI	Inverse probability (for n>1 dimension, this assumes that the first n-1 coordinates are specified		
Part 3		**Utility routines**			
	ERRFNC	Calculates error function		None	GANCF
	EXPECT	Calculates moments of elliptical distribution		None	ELLPAR
	GAMMA	Calculates gamma function		None	T2LPAR WEZPAR

Table A8
Summary of routines in Module PPREPRO

Group Name	Routines	Group/Routine Description	Calls	Called by
PPREPRO		**Pre-processor for probability distributions**		
Part 1		**Control routine**		
	DISTRB	Controls pre-processing of distributions	GANPAR GALPAR UNIPAR T2LPAR WEZPAR TRPPAR ELLPARX XPOPAR XPCPAR PRECOR OUTJNT	FARSIDE
Part 2		**Routines for discrete joint distributions**		
	OUTJNT	Outputs the joint discrete distributions	MDISP	DISTRB
	MDISP	Organises the layout of an array of numbers for output		OUTJNT
	JOIPAR	Calculates means and standard deviations for discrete distributions	H1DPAR H2DPAR H3DPAR	FARSIDE
Part 3		**Routines for Nataf distribution**		
	PRECOR	Calculates de-correlation coefficients for Nataf distribution	RHOINT/ RHOIN1	DISTRB
	RHOINT/ RHOIN1	Performs numerical integral for Nataf distribution	ETATOZ UTOX GANCI	PRECOR
	ETATOZ	Transforms variables	GANCI	RHOINT/ RHOIN1

Table A9
Summary of routines in Module INTER

Group Name	Routines	Group/Routine Description	Calls	Called by
INTER		**Standard interface with model routines**		
	PROBX	Defaults search control and optional output parameters and calls user-written routine PROB	PROB	FARSIDE
	CONX	Converts standard normal to model variables and calls user-written routine CON	UTOX CON	Various
	STARTX	Sets start point to origin and calls user-written routine START	START	FARSIDE
	UTOX	Converts standard normal to model variables	Inverse cumulative probability routines	Various

Table A10
Summary of routines in Module USER

Group Name	Routines	Group/Routine Description	Calls	Called by
USER		**User written routines to define probability distributions and to provide consequence model**		
	PROB	Sets user choices for search control and optional output parameters and specifies probability distributions for uncertain variables.	As required by user	PROBX
	START	Specifies start point for search (may be null routine since STARTX provides default)	As required by user	STARTX
	CON	Returns value of consequence function	As required by user	CONX

A4. PROBABILITY DISTRIBUTIONS

The current version of FARSIDE provides the user with a (small) library of 1-dimensional analytic probability distributions, plus the capability to define discrete distributions in 1, 2 and 3 dimensions. The facility also exists to allow the user to correlate any two of the 1 dimensional distributions.

Though all the facilities have been used successfully, some further work is necessary in order to integrate the discrete distributions and correlation specification fully into the FARSIDE scheme.

A4.1 Analytic 1 dimensional distributions

Table A11 provides a brief description of the 1 dimensional distributions available within FARSIDE.

Table A11
One dimensional analytic distributions available within FARSIDE

No.	Type	Probability density
1	Normal (Gaussian)	$(1/\sigma\sqrt{(2\pi)})\exp(-0.5((x-\mu)/\sigma)^2)$
2	Lognormal	$(1/s\sqrt{(2\pi)})\exp(-0.5((y-m)/s)^2)$; $y=\mathrm{Ln}(x)$
3	Uniform	$1/(x_2-x_1)$: $x_1 \le x \le x_2$; 0 elsewhere
4	Type II largest value	$(kw^k/x^{k+1})\exp(-(w/x)^k)$
5	Weibull (2-parameter)	$(kx^{k-1}/w^k)\exp(-(x/w)^k)$
6	Trapezoidal	$= A((x-a)/(b-a))$ for $a \le x \le b$ $= A$ for $b \le x \le c$ $= A((d-x)/(d-c))$ for $c \le x \le d$ where $A = 2/((c+d)-(a+b))$
7	Elliptical	Not yet complete
8	Exponential	$\lambda\exp(-\lambda(x-x_0))$
9	Truncated exponential	$= A\exp(-\lambda x)$ for $x_1 \le x \le x_2$ = 0 elsewhere where $A = \lambda/(\exp(-\lambda x_1)-\exp(-\lambda x_2))$

A4.2 Discrete distributions

Tables A12 - A14 summarise the arrays which hold the discrete distributions within FARSIDE. The contents of the arrays should be specified by the user in subroutine PROB.

The generation of data for the arrays, from the starting point of a set of raw measurements, is described in Appendix B.

Table A12
One dimensional discrete distributions available within FARSIDE

Variable	Description
NDIST1	Total number of 1-D distributions
NW1(m)	Output instructions for distribution m: = 0: Don't write any details of this distribution = 1: Write raw data distribution only = 2: Write all the derived distributions
FD1(i,m)	Frequency value in interval i for raw data distribution m
NX1(m)	Number of cells in x-direction for distribution m
X1(i,m)	Values of cell boundaries for distribution m
FC1(i,m)	Cumulative distribution for distribution m

Table A13
Two dimensional discrete distributions available within FARSIDE

Variable	Description
NDIST2	Total number of 2-D distributions
NW2(m)	Output instructions for distribution m: = 0: Don't write any details of this distribution = 1: Write raw data distribution only = 2: Write all the derived distributions
FD2(i,j,m)	Frequency value in interval (i,j) for raw data distribution m
NX2(m)	Number of cells in x-direction for distribution m
NY2(m)	Number of cells in y-direction for distribution m
X2(i,m)	Values of cell boundaries in x direction for distribution m
Y2(i,m)	Values of cell boundaries in y direction for distribution m
FX2(i,m)	Cumulative marginal distribution $F(X \leq x)$ for distribution m
FYX2(i,j,m)	Cumulative conditional distribution $F(Y \leq y \mid X=x)$ for distribution m
NXREF2(n)	Cross reference table. If n is the global index number of a variable which is the λth variable in a ν-dimensional distribution ($\lambda>1$), then NXREFν(n) is the global index number of the λ-1 th variable in the distribution.

Table A14
Three dimensional discrete distributions available within FARSIDE

Variable	Description
NDIST3	Total number of 3-D distributions
NW3(m)	Output instructions for distribution m: = 0: Don't write any details of this distribution = 1: Write raw data distribution only = 2: Write all the derived distributions
NX3(m)	Number of cells in x-direction for distribution m
NY3(m)	Number of cells in y-direction for distribution m
NZ3(m)	Number of cells in z-direction for distribution m
X3(i,m)	Values of cell boundaries in x direction for distribution m
Y3(j,m)	Values of cell boundaries in y direction for distribution m
Z3(k,m)	Values of cell boundaries in z direction for distribution m
FX3(i,m)	Cumulative marginal distribution $F(X \le x)$ for distribution m
FYX3(i,j,m)	Conditional distribution $F(Y \le y \mid X=x)$ for distribution m
FZYX3(i,j,k,m)	Conditional cumulative distribution $F(Z \le z \mid X=x, Y=y)$ for distribution m
NXREF3(n)	Cross reference table. If n is the global index number of a variable which is the λth variable in a ν-dimensional distribution ($\lambda>1$), then NXREFν(n) is the global index number of the λ-1 th variable in the distribution.

APPENDIX B

REPRESENTATION OF ATMOSPHERIC VARIABLE FREQUENCY DISTRIBUTIONS

B1 INTRODUCTION

The main difficulty in dealing with the frequency distributions of atmospheric variables is that the distributions are interdependent; that is: the frequency distribution of one variable is dependent upon the value(s) of the other(s). For example, the distribution of wind speed is known to vary significantly with atmospheric stability. The frequency distribution has therefore to be represented in general as a function of two (wind speed and atmospheric stability) or three (wind speed, wind direction and atmospheric stability) variables. Furthermore, this function cannot in general be approximated by standard analytic forms, as can the distributions for a single variable.

In the previous phase of the work [B1], emphasis was placed upon the development of methods for the manipulation of atmospheric frequency distributions when these were represented as (multi-dimensional) histograms. The reasons for the choice of this representation were twofold:

– In many cases, the data will have already been processed into this form

– The histogram is particularly easy to store and manipulate

However, the representation of the frequency distribution in this discontinuous form presents problems for the MLFP method, which works better with smooth distributions. Furthermore, the standard method for processing interdependent distributions within the MLFP method (the Rosenblatt transformation: see Section B3.3.1) fails when the distribution is discontinuous. Although an alternative approximate method for treating histograms was developed during the course of the previous work, it became evident that, where good data were available, an improved method for representing these distributions was desirable.

This work considers the case where unprocessed measurements of atmospheric data are available and presents methods for:

– the selection of suitable independent variables;

– the derivation of a smooth frequency distribution from the scatter plot of raw measurements

– the representation of the distribution for use with the MLFP method

These methods were not used in the work described in the main report.

B2 SELECTION OF INDEPENDENT VARIABLES

For the representation of measured frequency distributions, independent variables are required which vary continuously and monotonically over the ranges of interest and which give rise to a reasonable shape of distribution (e.g. when the measurements are represented as a scatter

plot). These criteria may necessitate that the quantities in which the original measurements are recorded should be transformed in some way.

The methods discussed below were developed for the wind speed and stability variables and this Appendix therefore concentrates upon the representation of a distribution over these two quantities only. The methods in Section B3 onwards can be extended readily to include other variables, notably wind direction.

B2.1 Atmospheric stability

Atmospheric stability is related to the vertical heat flux in the atmosphere which is in turn caused by the heating of the ground by the sun during the day, or the cooling of the ground by thermal radiation into space during the night. During heating (unstable atmosphere), additional turbulence (over and above that due to the wind) is generated by convection currents, whereas at night, any existing turbulence is damped by the stable stratification.

None of the common measures of stability produces a suitable shape of scatter plot.

Atmospheric stability was originally expressed in terms of Pasquill stability categories, which are a semi-quantitative method for ranking stability in terms of readily recognisable types of weather. The categories are traditionally denoted by the letters A to F (with an additional extremely stable category G sometimes added). A numerical form of this scheme, referred to in [B2] as Smith numbers, fulfils most of the requirements listed above apart from that of continuity (see Table B1).

Atmospheric stability is nowadays commonly expressed in terms of a quantity known as the Monin-Obukhov length L, which is a ratio of a number of measurable quantities, including the vertical heat flux. Although this quantity is extremely useful in that (among other interpretations) it acts as a scaling length for the vertical variation of a number of quantities related to atmospheric boundary layer behaviour [B3], it is not itself a suitable quantity for the current purposes, since it tends to $-\infty$ or ∞ as the stability tends towards neutral from, respectively, the unstable and stable regimes (see Table B1). An alternative representation is the inverse stability parameter σ, calculated as

$$\sigma = a/L \qquad \text{(B1)}$$

where a is some constant, which we may assume has dimensions of length in order to render σ dimensionless. However, if σ is used directly as the independent variable, the distribution tends to display a large spike about the origin, since the overwhelming majority of the measurements correspond to relatively small values of σ. A logarithmic version of σ would tend to display the right type of behaviour, provided that the formalism can be extended to values of σ of less than or equal to zero. It is found that the definition of a quantity s, where

$$s = \text{Sign}(\sigma)\ \text{Ln}(1+\sigma\ \text{Sign}(\sigma)) \qquad \text{(B2)}$$

where $\text{Sign}(\sigma) = +1$ when $\sigma \geq 0$ and $= -1$ when $\sigma < 0$, solves all the problems reviewed above.

Table B1 compares these different measures of stability. The stability parameter s in the last column is calculated from L using a value for the coefficient a (Equation (B1)) of 24800 m. In this case, the values of σ are numerically equal to the values of the stability parameter p (which have dimensions J m^{-3} K^{-1}) obtained from the Cardington data set (see Section B2.3 below).

Table B1
Comparison of different representations of atmospheric stability

Type of weather	Degree of stability	Pasquill category	Smith No.	Monin-Obukhov length L (m) [B2] (Note 1)	s (a = 24800 m)
Very sunny summer weather	Extremely unstable	A	0.5	-0.4	-11.0
Sunny and warm	Moderately unstable	B	1.5	-8.4	-8.0
Partial cloud during the day	Slightly unstable	C	2.5	-55.4	-6.1
Overcast day or night	Neutral	D	3.5	-/+∞	0
Partial cloud during the night	Slightly stable	E	4.5	476	4.0
Clear night	Moderately stable	F	5.5	107	5.4

Notes: (1) These are typical values. Other sources give different values.

B2.2 Wind speed

In the previous report, the use of the cube root of the windspeed was found to produce a reasonable shape of distribution when the data were represented on a scatter plot. In this work, this representation has been dropped in favour of the logarithm of the windspeed. This change has been carried out simply to give some kind of correspondence between the transformations of the stability parameter and the wind speed. There is otherwise very little to choose between the two types of representation, and no theoretical basis is claimed for either.

The wind speed variable w is therefore given as:

$$w = \ln(bv_k/v_0) \qquad \text{(B3)}$$

where

v_k is the wind speed in the original units

b is a factor to convert to m s^{-1} (0.5144 if v_k is in knots)

v_0 is 1 m s^{-1}, to produce a dimensionless argument

B2.3 Scatter plot

Figure B1 shows a scatter plot of the database of wind speed / stability measurements taken at Cardington between 1988 and 1990 (see [B1]), using the processed variables discussed above. The value of the coefficient a (Equation (B1)) was the same as that used in Table B1.

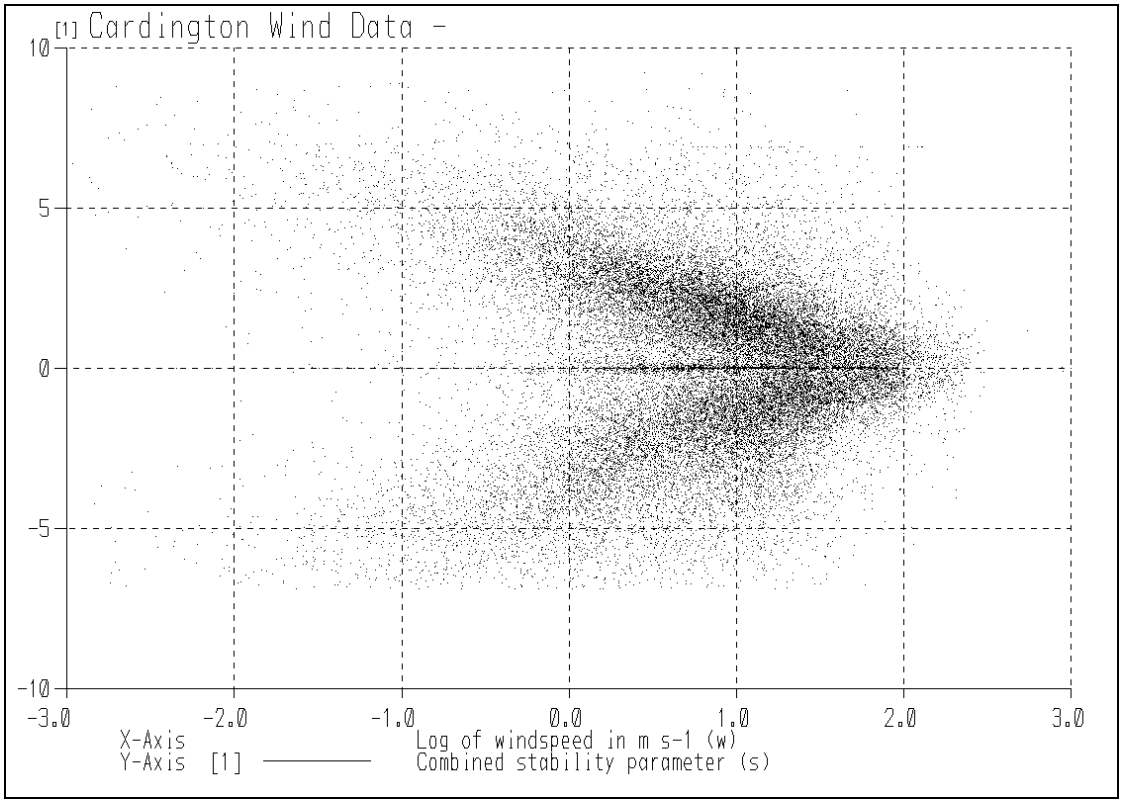

Figure B1
Scatter plot of Cardington wind speed / stability measurements using transformed variables w and s

The transformations described in Sections B2.1 and B2.2 have managed to produce a scatter plot without excessive clustering into small regions. Nevertheless, the peculiar shape of the distribution is evident, being of arrowhead shape with remarkable symmetry about the neutral stability line s = 0. The dark stripe of points along this line is probably composed of spurious readings, but it has been retained since a clear criterion for rejecting these points has not been formulated.

The distribution shown in Figure B1, comprising just over 35000 measurements, is used as the basis for the construction of the probability density estimate, as discussed below.

B3 DERIVATION OF THE FREQUENCY DISTRIBUTION

Within the subject of statistics, the process of the construction of a smooth frequency distribution $p(\underline{x})$ from a set of sampled data is known as *density estimation* [B4, B5]. Historically, density estimation has been almost exclusively parametric. That is: a form of density function (say, a Gaussian) is assumed and the data sample is then used to derive an estimate of the parameters of the distribution; in this example: the mean and the standard deviation.

Since the atmospheric distributions cannot be represented by standard forms, nonparametric density estimation is required. In this process, the density function is constructed directly from the sample with a minimum of prior assumptions regarding the functional form of the distribution. Nevertheless, since the resulting distribution contains more information than the sample from which it is constructed, it is clear that some additional assumptions must be made. It seems reasonable to summarise these assumptions as follows:

– a smooth underlying distribution does exist

– structure within the distribution must be on a larger scale than the typical distance between points on the sampled distribution (which of course varies throughout the distribution)

Both these assumptions are appropriate in the case of the atmospheric distributions being considered.

Although the process of nonparametric density estimation seems intuitively straightforward, there is no method which is clearly superior for the performance of this process in general, and it is remarkable that the histogram stood as the only nonparametric method for density estimation until the middle of the last century [B4].

A review of currently available methods was carried out and, as a result, the *adaptive kernel method* [B4, B5] was selected for use during this work, because:

– it produces a smooth density function;

– it can be integrated analytically;

– its basis is easily understood;

– it is free of arbitrary assumptions such as bin width and bin origin.

The adaptive kernel method (AKM) is based on the kernel method, which is introduced in Section B3.1 below. The AKM itself is described in detail in Section B3.2 and a number of further amendments which allow the AKM to produce distributions which can be used routinely in the MLFP method are discussed in Section B3.3.

B3.1 The kernel method

In the kernel (i.e. smoothing function) method of density estimation, each point on the scatter plot is effectively smeared out, so that a smooth surface representing the density distribution is built up from the sum of overlapping humps, each being of unit 'volume' and centred on a measurement point. It is found (see [B4]) that the shape of the hump is not usually of great importance for accuracy (though it may have significant influence on speed: see below), but it is evident that the width of the humps (the 'smoothing parameter') can have a considerable influence on the features of the surface which results from the application of this method. If the smoothing parameter is too small, then the distribution (particularly the tail region) becomes lumpy. If the parameter is too large, then important structure may be lost.

Because, in the current application, the tail of the frequency distribution (i.e. the stable, low windspeed region) is important, it is clear that this area should be well represented. A variant of the kernel method: the adaptive kernel method (AKM), makes the best use of the kernel approach in all the regions of the distribution by varying the smoothing parameter according to the local density of points in the region of the point in question: small (i.e. a narrow kernel) for the dense regions near the peak of the distribution and relatively large (a broad kernel) for the sparse regions in the tails. The way that this process is carried out is described below.

B3.2 The adaptive kernel method

The procedures for producing an adaptive kernel estimate of the probability density function are as follows:

1. Select suitable scaled versions of each variable.

2. Choose a baseline value h for the smoothing parameter (see below).

3. Produce an initial ('pilot') estimate $p_p(\underline{x})$ of the probability density using any appropriate method (in this case, a straight (i.e. non-adaptive) kernel estimate: Section B3.1).

4. For each data point, calculate local scale factors for the smoothing parameter, based on the pilot estimate of the density.

5. Re-apply the kernel density estimate using the local scale factors.

6. Calculate a measure of the goodness of fit. If this is not optimal, amend the value of h and repeat steps 3 - 6.

The first of the above steps is present since the form of the kernel used below is rotationally symmetric: the value of the kernel function depends only on the distance from the point at which it is centred. It is therefore necessary to normalise each of the coordinates so that the extent of the scatter in each direction is broadly the same. Formal methods for performing this process do exist, but these tend to apply to scatter plots which are broadly elliptical (thereby converting them to a roughly circular shape). For the shape of distribution shown in Figure B1, these methods were judged to offer no advantage, so the scale factors were simply estimated by eye.

Therefore, for the 2 dimensional case of wind speed and stability, the kernel estimates were carried out using the coordinates (x_1,x_2), where

$$x_1 = a_w w \tag{B4a}$$

$$x_2 = a_s s \tag{B4b}$$

where a_w and a_s are suitable scale factors.

B3.2.1 General forms of kernel

As discussed above, there are a number of possible kernels to choose from. The most appealing in terms of its functional form is the multivariate standard normal (i.e. Gaussian) kernel:

$$\phi_M(r) = (2\pi)^{-M/2} \exp(-r^2/2) \tag{B5}$$

where M is the number of dimensions, and

$$r = (u_1^2 + u_2^2 + \ldots. + u_M^2)^{1/2} \tag{B6}$$

The normalised coordinates u_m, m = 1…M are defined relative to the location of the data point about which the kernel is centred, and are given by

$$u_m = (X_m - x_m)/h \qquad m=1….M \tag{B7}$$

where

x_m represents the scaled global coordinates, described above

X_m is the mth coordinate of the data point about which the kernel is centred

h is the baseline smoothing parameter

However, the presence of the exponential makes the function significantly slower to evaluate than some of the simpler forms. Because of this, the pilot estimate (for which the kernel has to be evaluated many more times: see below) is carried out using a particularly simple form of smoothing function known as the Epanechnikov kernel [B5]:

$$K_{eM}(r) = \tfrac{1}{2} c_M^{-1}(M+2)(1-r^2) \qquad r < 1 \qquad \text{(B8)}$$

$$= 0 \qquad r \geq 1$$

where c_M is the volume of a sphere of unit radius in M dimensions and where the value of h_p, the (fixed) smoothing parameter used for the pilot estimate, is chosen to give a reasonably smooth distribution within the tails (although the overall method is reported [B6] to be relatively insensitive to h_p).

B3.2.2 Construction of the pilot estimate

The pilot estimate $p_p(\underline{x})$of the density has to be evaluated at each of the K (K > 35000) data points and is given by

$$p_p(\underline{x}) = \frac{1}{Kh_p^M} \sum_{k=1}^{K} K_{eM}(r_x) \qquad \text{(B9)}$$

where

$$r_k = (u_{k1}^2 + u_{k2}^2 + \ldots + u_{kM}^2)^{1/2} \qquad \text{(B10)}$$

where

$$u_{km} = (X_{km} - x_m)/h_k \qquad \text{(B11)}$$

and where h_p is the smoothing parameter for the pilot estimate (which is not necessarily equal to h).

It should be noted that (in principle), calculation of the pilot estimate for every one of the K points requires evaluation of the kernel at all the other points, leading to K^2 calculations. For K large, this is clearly a time-consuming task. Fortunately, the finite range of the kernel (B8) means that most points can be screened out without the need to evaluate any complex expressions. Furthermore, more efficient methods [B7] exist by which the nearest neighbours to any point can be identified. These methods have not been used during this work, but could be employed if improvements in calculation speed are required.

Once the pilot estimate has been constructed, the local smoothing parameter scale factors λ_k, k=1…K for each of the K data points are calculated as

$$\lambda_k = [p_{pk} / g]^{-1/2} \qquad \text{(B12)}$$

where

p_{pk} is the pilot density at point k

and

g is the geometric mean of the p_{pk}, i.e.

$$\mathrm{Ln}(g) = (1/K)\Sigma\ \mathrm{Ln}(p_{pk}) \tag{B13}$$

The adaptive kernel estimate is then calculated as

$$p(\underline{x}) = \frac{1}{K}\sum_{k=1}^{K}\frac{\phi_M(r_k)}{h_k^M} \tag{B14}$$

where the kernel is now the Gaussian kernel, described above (Equation (B5)), and

$$h_k = \lambda_k h \tag{B15}$$

and it may be observed that (B4)

$$p(w,s) = a_w a_s p(\underline{x}) \tag{B16}$$

B3.3 Use of the AKM to derive distributions for the most likely failure point method

B3.3.3 Requirements of the MLFP method

It may be recalled [B1], that the MLFP method requires that interdependent distributions are:

1. expressed in their cumulative form $\left(\text{i.e.}P(\underline{x}) = \int p(\underline{x})d\underline{x}\right)$, and

2. factorised into the product of M distributions, where M is the number of interdependent variables.

In the case of 2 dimensions (say, for 'wind speed' w and 'stability' s), the MLFP method generates values of the corresponding standard normal variables (u_w and u_s) and derives values of w and s by solving the equation

$$\Phi_1(u_w) = P_w(w) \tag{B17}$$

for w and the equation

$$\Phi_1(u_s) = P_s(s \mid w) \tag{B18}$$

for s, where

$P_w(w)$ is the cumulative marginal distribution for w

and the value of w found from the first equation is passed through to the second. As discussed in [B1], there are two possible forms for $P_s(s \mid w)$. The first is

$$P_s(s \mid w) = P(s \mid W=w) \tag{B19}$$

which is referred to in the literature as the Rosenblatt transformation, and the second is

$$P_s(s \mid w) = P(s \mid W \leq w) \tag{B20}$$

The latter (B20) is obtained from the identity

$$P(w,s) = P(w)P(s \mid W \leq w) \tag{B21}$$

where P(w,s) is the joint cumulative distribution of w and s. It has the advantage that it may be used when the density distribution p(w,s) is discontinuous (e.g. a histogram).

The origin of (B19) may be illustrated by considering a Monte Carlo calculation, rather than an MLFP calculation, in which the objective is to generate points in the (w,s) plane with a density proportional to p(w,s). Random points r are generated between the values 0 and 1. A value of windspeed can first be produced by generating a random point r_w and then solving the equation:

$$r_w = P(w) \tag{B22}$$

This process generates the correct density of points in each interval of width dw. The total number of points generated in this interval have then to be shared out correctly in the s direction. It is evident that this can be achieved by generating another random point r_s and solving

$$r_s = P(s \mid W=w) \tag{B23}$$

where P(s | W=w) is the cumulative distribution in the s direction, with the value of variable W fixed at w (i.e. ignoring the rest of the distribution).

This process generates points with the correct density, and is therefore 'exact' for use with the Monte Carlo method. It is evident therefore that the method will produce the same result if the variables are taken in a different order.

B3.3.4 Calculation of the probability estimates

According to (B12), the adaptive kernel expression requires a summation over K data points each time the expression is evaluated. Although this is quite feasible for a pre-processing routine, it would be too slow for use within the MLFP method, where the calculation may have to be performed many times.

In order to overcome this problem, the method selected here is as follows: the full distribution is calculated in advance at points on a Cartesian grid of M dimensions. This array of values is then used within the MLFP calculation, where the appropriate value of the distribution can rapidly be interpolated.

The use of this method has a further advantage. From (B5):

$$\phi_M(r) = \prod_{m=1}^{M} \phi_1(u_m) \tag{B24}$$

where ϕ_1 is the one-dimensional normal distribution. For the 2 dimensional case of 'wind speed' w and 'stability' s:

$$\phi_2(r_k) = \phi_1(u_{wk})\phi_1(u_{sk}) \tag{B25}$$

and, for the cumulative distribution:

$$\Phi_2(r_k) = \Phi_1(u_{wk})\Phi_1(u_{sk}) \tag{B25}$$

Therefore, assume that the density function and the cumulative distributions (above) are to be evaluated at a grid of points (w_i, s_j); i=1…I, j=1…J.

For a point (w_i, s_j), it is straightforward to show that:

$$p(w_i, s_j) = \frac{a_w a_s}{K} \sum_{k=1}^{K} \phi_1'(u_{wki})\phi_1(u_{skj}) \tag{B27}$$

$$P_w(w_i) = \frac{1}{K} \sum_{k=1}^{K} \Phi(u_{wki}) \tag{B28}$$

and

$$P_s(s_j | W = w_i) = \frac{\sum_{k=1}^{K} \phi_1'(u_{wki})\Phi_1(u_{skj})}{\sum_{k=1}^{K} \phi_1'(u_{wki})} \tag{B29}$$

where

$$\phi_1' = \phi_1 / h_k \tag{B30}$$

Because of the property that the Gaussian kernel in M dimensions can be factorised into the product of M one-dimensional kernels (B24), it is possible to evaluate the AKM probability (density, cumulative or conditional) by evaluating the kernels separately along each grid direction, so requiring (I+J)K calculations of the kernel, rather than the IJK evaluations which would be required otherwise. It is therefore possible to generate quite a fine grid for interpolation without paying too high a penalty in terms of the required number of evaluations of the Gaussian kernel.

B3.4 Optimisation of the smoothing parameter

Although the above method scales the quantity h_k to be proportional to the local probability density function, (or, at least, to the pilot estimate of this quantity) it provides no information as to whether the original choice of the smoothing parameter h is optimal. This question might seem to be incapable of resolution, since there is no existing probability density function against which the estimate (B14) can be compared to determine the best fit. However, Ref. B6 uses the following argument:

Assume that, in addition to calculating the density p_k at each data point k; k=1…K, the nearest-neighbour volume V_{k1} (the volume of a circle (or sphere or hypersphere) of radius equal to the distance from k to its nearest neighbour) is calculated. The product $p_k V_{k1}$ should be constant over the whole distribution but, because of the finite number of data points, this quantity will correspond to a random variable with a certain distribution. Ref. B6 states that, if p_t is the true density, then the values

$$z_k = \exp(-Kp_{tk}V_{k1}) \qquad k = 1...K \tag{B31}$$

have a distribution that is approximately uniform over the interval [0,1]. This means that for the k'th largest value of z_k:

$$z_k^{(k')} \approx \frac{k'}{K} \tag{B32}$$

Therefore, as h is varied, the closest approach to the true density should occur when the quantity S where:

$$S = \sum_{k'=1}^{K} \left| z_k^{(k')} - k'/K \right| \tag{B33}$$

is at its smallest value. This therefore should provide an objective way of choosing the baseline smoothing parameter h.

B4 EXAMPLE CALCULATIONS

B4.1 Range scaling factors

The AKM as described above was used to derive an estimate of the probability distribution underlying the Cardington data, shown in Figure B1. The figure suggests that over the main region of the distribution, the range of the variable s is roughly five times that of w. Therefore, the scaled variables:

$x_1 = w$

$x_2 = 0.2s$

were defined in order to allow the use of a symmetric kernel.

B4.2 Optimisation of the baseline smoothing parameter (h)

The value of h was optimised using the method described in Section B3.4. Although the pilot estimate smoothing parameter h_p could also have been included in the optimisation, it was decided to use a fixed value of 0.02, since the final result would be expected to be reasonably insensitive to this quantity.

Figure B2 shows the variation of the optimisation parameter S (Section B3.4) with the change in h. The minimum value of S occurs at a value of h of 0.024.

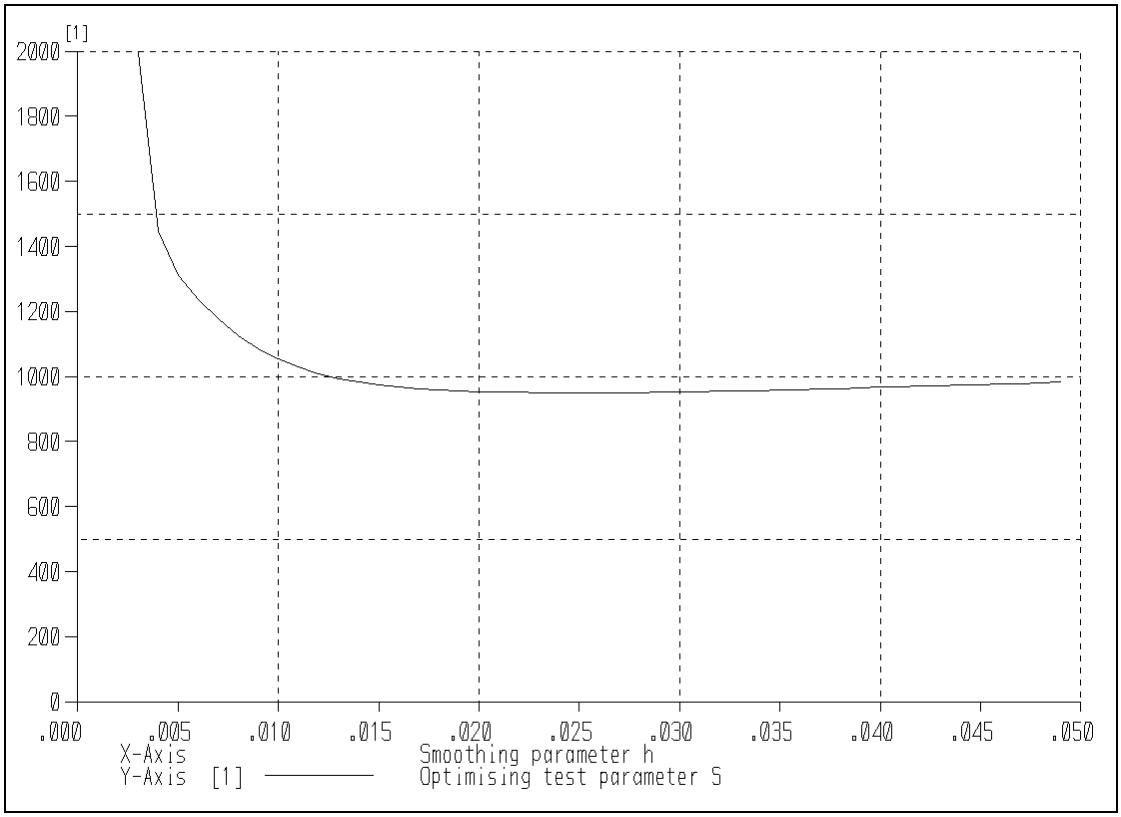

Figure B2
Variation of optimising parameter S with baseline smoothing parameter h

B4.3 Grid points

The ranges of the parameters w and s were taken as follows:

w from -3.0 to 2.64

s from -10 to +10

where the former corresponds to wind speeds v ranging from 0.05 m s^{-1} to 14.0 m s^{-1}.

20 equally-spaced points (i.e. 19 equal intervals) were taken over each of these ranges.

B4.4 Results

For the value of h calculated in Section B4.2 (i.e. 0.024), Figure B3 shows a surface plot of the probability density $p(\underline{x})$, for values generated at the 20 x 20 mesh described in Section B4.3.

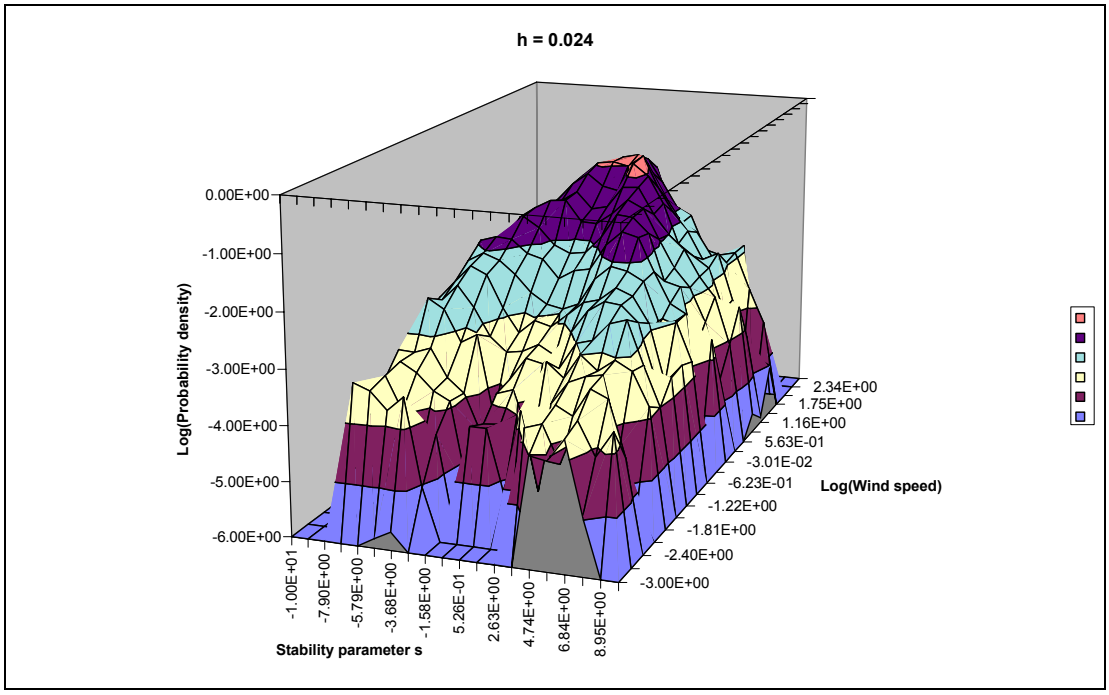

Figure B3
Probability density derived by adaptive kernel method using smoothing parameter of 0.024

The density is represented on a logarithmic scale in order to emphasise the representation of the probability density in the tails of the distribution. Very small (or zero) values of p are represented by a cutoff value of Ln(p) = -6.0.

Although Figure B3 was derived with the 'optimised' value of h, it is clear that the distribution thereby obtained is too uneven, particularly in the tails. As Figure B2 shows, the quantity S varies only slowly as h is increased, suggesting that some flexibility should be allowed in setting the location of the best value. The calculation was therefore repeated with h doubled to 0.048, with the resulting distribution being shown in Figure B4. The appearance of the surface is seen to be considerably improved, with the two lobes at low windspeed and the spurious stripe of readings along the s = 0 axis being clearly visible.

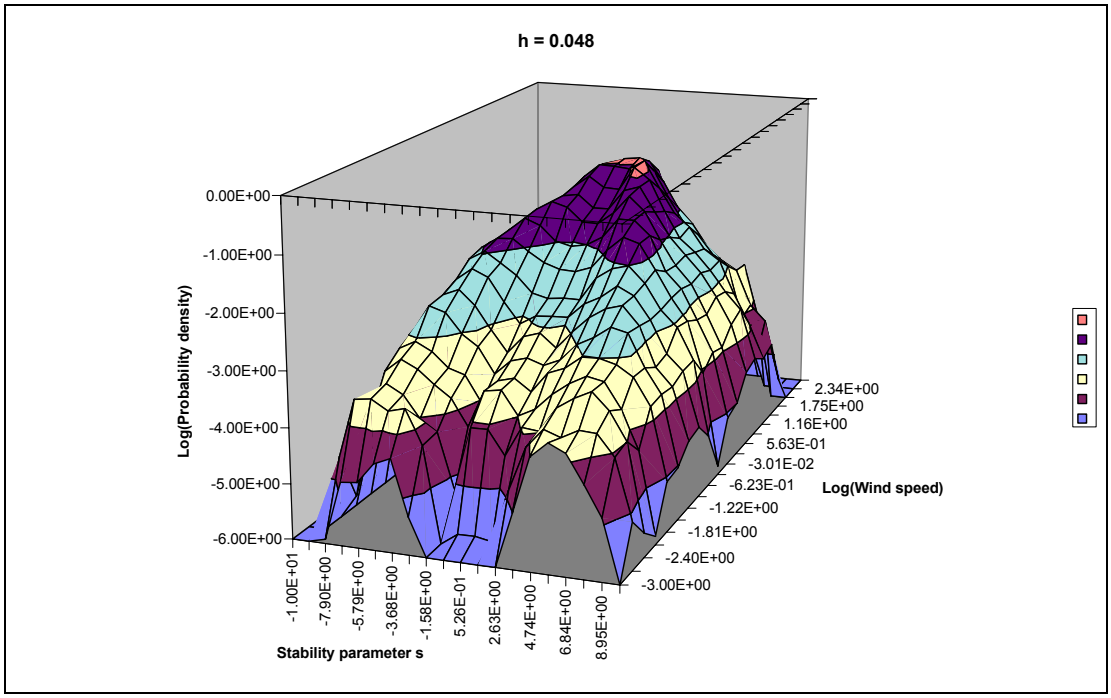

Figure B4
Probability density derived by adaptive kernel method using smoothing parameter of 0.048

B5 CONCLUSIONS

The use of the variables described in Section B2.1 allow the direct transformation of measured Monin-Obukhov length into a measure of atmospheric stability which bears a reasonable resemblance to the more familiar quantities such as the Pasquill categories or Smith numbers, their numerical equivalent. This, together with the logarithmic transformation of wind speed described in Section 0, allow an improved representation of collections of simultaneous windspeed/stability measurements as a scatter plot.

This type of plot covers the full ranges of the variables without concentrating the majority of the measurements into small regions.

The adaptive kernel method (AKM) provides the potential of allowing the representation of the scatter plot described above, as a smooth surface, though further work is necessary in order to produce an automatic method for optimising the smoothing parameter.

A number of possible areas for further investigation would be:

– development of an alternative criterion for optimisation;

– increasing the exponent in Equation (B12), thereby increasing the local smoothing factors for regions of low density;

– investigation of the use of different forms for the pilot density used in the derivation of the local scale factors.

Interpolation of the surface generated by the AKM over a reasonably fine Cartesian grid then allows this surface to be used as an interdependent probability distribution within the most likely failure point method.

B6 REFERENCES

B1. Mitchell, B. Use of the FORM/SORM (most likely failure point) method for quantitative risk assessment. WSA Report AM5123-R1. 1998.

B2. Hunt, J.C.R. and others. Developments in modelling air pollution for regulatory uses. 18th NATO-CCMS International Conference on Air Pollution Modelling and its Applications. Vancouver, Canada. 1990

B3. Kaimal, J.C. and Finnigan, J.J. Atmospheric Boundary layer Flows. Oxford University Press. 1994.

B4. Scott, D.W. Multivariate Density Estimation. Wiley-Interscience. 1992.

B5. Silverman, B.W. Density Estimation for Statistics and Data Analysis. Chapman and Hall. 1986.

B6. Breiman, L., Meisel, W. and Purcell, E. Variable kernel estimates of multivariate densities. Technometrics, Vol 19, No. 2, May 1977. Pages 135-144

B7. Friedman, J.H., Bentley, J.L. and Finkel, R.A. An algorithm for finding best matches in logarithmic expected time. ACM Transactions on Mathematical Software, Vol. 3, No. 3, September 1977. Pages 209-226

APPENDIX C

REPRESENTATION OF CRUNCH INPUT QUANTITIES

The following table has been derived mainly from information in Ref. 4 of the main report, supplemented by inspection of the correspondence between input quantities and their reappearance in the output file. Symbols are as defined in Ref. 4. For the symbols used in the rest of the report, see Appendix D.

Keyword	Symbol (Ref. 4)	Description
(Source definition)		
GAS		
	M_g	Release rate (kg s^{-1})
	T_g	Release temperature (K)
	ρ_g	Gas density at NTP (kg m^{-3})
	C_{pg}	Gas specific heat capacity (J kg^{-1} K^{-1})
	T_b	Boiling temperature (K)
	L	Latent heat of vaporisation (J kg^{-1})
	K_1	Mass fraction of release as an aerosol (-)
AIR		
	T_a	Air temperature (K)
	K_a	Multiple by mass of air entrained at source (this is increased automatically if necessary to evaporate any aerosol)
SOURCE		
	a_r	Shape of the source (see n)
	n	Switch for source shape: =0: a_r is the aspect ratio; =1: a_r is half width (m) =2: a_r is height (m)
PULSE		(Seems to have no effect)
	T_1	Duration of transient release (s)
MULT		
	n	Number of sources in array
	d	Separation (m)
ATMOS		
		Keyword signifies that the gas is to be assumed passive (neutral buoyancy) at release from source

Keyword	Symbol (Ref. 4)	Description
(Transport conditions)		
WEATHER		
	n_1	Number of stability categories to be considered. Followed by n_1 lines with:
	S_P	Pasquill stability class (A - F)
	v	Wind speed (m s^{-1})
GROUND		
	z_0	Roughness length (m)
	z_w	Reference height for wind speed (m)
	a_1	Switch for heat exchange between air and plume
	c_f	Friction factor of the surface (undefined)
SPRAY		
	X	Distance of spray from source (m)
	P	Strength of spray (kg s^{-1} m^{-1})
	S	Aspect ratio of the plume after the spray
	W	Width of the spray (m)
(Plume behaviour)		
VAN ULDEN		
	K	Constant governing growth in plume width (-)
ALPHA		
	α*	Edge entrainment coefficient (-)
	α^1	Top entrainment coefficient (-)
	α	Ground heating coefficient (presumably W m^{-2} K^{-1})
(Output control)		
TERP		(Seems to have no effect)
	$x_1 \ldots x_n$	Distances (m) at which interpolated plume parameters are to be evaluated
TOXIC		
	t_1	Hazard level 1 (ppm)
	t_2	Hazard level 2 (ppm)
	t_3	Hazard level 3 (ppm)

APPENDIX D

NOMENCLATURE

Symbol	Meaning
a	Normalising constant in truncated exponential distribution (Section 4.3)
a	Constant in stability parameter definition (m) (Appendix B)
a	Constant in definition of Probit y (Section 4.6.2)
a_s	Scaling factor for atmospheric stability variable
a_w	Scaling factor for wind speed variable
b	Constant in definition of Probit y
c	(Chlorine) release rate (kg s^{-1})
c_M	Volume of unit sphere in M dimensions
d	Distance (m) from source to dose point
D	Toxic dose ($ppm^n min$) (see definition of n below)
D_d	Value of 'dangerous' toxic dose
f_T	Frequency (y^{-1}) of a release (of any size)
g	Geometric mean of pilot estimate of probability density
h	Smoothing parameter in adaptive kernel method (Appendix B)
h	Leak discharge parameter (kg s^{-1} mm^{-2}) (Section 4.3)
h	Plume/cloud height (m) (Section 4.4.3)
h_k	Smoothing parameter for point k in adaptive kernel method
h_p	Smoothing parameter in 'pilot' estimate of probability density in adaptive kernel method
i	Index for I
I	Number of points in x-direction in joint frequency distribution grid
I_i	Importance of the i^{th} uncertain input quantity to the consequence model
j	Index for J
J	Number of points in y-direction in joint frequency distribution grid
k	Parameter in Weibull and Type II largest value probability distributions
K	Number of data points in scatter plot
K_{eM}	Epanechnikov kernel (smoothing) function
L	Monin-Obukhov length (m)
m	'Mean' in Lognormal probability distribution
M	Number of dimensions in interdependent probability distribution
M_g	Tank inventory (kg)
n	Exponent in expression for toxic dose D
N	Number of uncertain input variables to the consequence model
p	Atmospheric stability parameter in Cardington database (J m^{-3} K^{-1}) (Section B2.1)
p	Probability density function
p_p	'Pilot' estimate of probability density in adaptive kernel method

Symbol	Meaning
P	Cumulative probability distribution
P_d	Fraction of population killed by toxic dose
P_f	The probability that $Q \geq Q_{lim}$
P_{fl}	First-order estimate of P_f
Q	'Quantity of interest' predicted by the consequence model (here: the quantity being used to measure the impact of the toxic release)
Q_{lim}	The value of Q determining the 'failure surface' in the MLFP method
r	Normalised distance
r_a	Aspect ratio of cloud (= 2h/w)
R	Risk
R_t	Risk per unit time
R_c	Conditional risk
s	Atmospheric stability parameter (-) (Section B2.1)
s	'Standard deviation' in Lognormal probability distribution (Appendix A)
S	Smoothing parameter optimising variable
t	Time
T	Temperature
T_a	Air temperature (K)
T_g	Ground temperature (K)
u	Any variable having a standard normal probability distribution
u_c	Standard normal variable corresponding to chlorine release rate
u_i	Standard normal variable corresponding to the i^{th} uncertain input quantity to the consequence model
u_{im}	Value of u_i at the most likely failure point
u_s	Normalised variable for stability (-)
u_v	Standard normal variable corresponding to wind speed
u_w	Normalised variable for wind speed (-)
u_x	The standard normal variable corresponding to the variable x
v	Wind speed (m s^{-1})
v_k	Wind speed in expressed original units (usually knots)
V	Variable quantifying physical cause of harm or injury
V_{kl}	Spherical volume surrounding data point k, as far as its nearest neighbour
w	Processed variable for windspeed (-) (Section B2.2)
w	Plume/cloud width (m) (Section 4.4.3)
w	Parameter in Weibull and Type II largest value probability distributions (Appendix A)
x	An input quantity to the consequence model
x	Variable in Eq (4.11)
x_i	The ith input quantity to the consequence model
X	Fortran array containing particular values of uncertain variables (Section 3.2)

Symbol	Meaning
X	Coordinate of data point in scatter plot (Appendix B)
y	Probit: harm or injury variable expressed as y=a+bLn(V)
z_k	Random variable for data point k in scatter plot
z_0	Ground roughness length (m)
α	Constant in Equation (4.9)
β	Distance (number of standard deviations) from the best estimate point to the most likely failure point
β	Constant in Equation (4.9)
δ	Leak hole diameter (mm) (Section 4.3)
δ	Incremental value of u in FARSIDE search algorithm (Appendix A)
ε	Tolerance in FARSIDE search algorithm
θ	Angle relative to downwind direction (Section 4.5)
λ	Constant in Equation (2.25)
λ	Constant in truncated exponential distribution
λ_k	Smoothing parameter local to data point
μ	Mean value
μ_x	Mean of population toxic tolerance (as expressed by the variable x = Ln(D))
σ	Atmospheric stability parameter (Section B2.1)
σ	Standard deviation
σ_x	Standard deviation of population toxic tolerance (as expressed by the variable x = Ln(D))
τ	Release duration
τ_m	Release duration (min)
τ_s	Release duration (s)
ϕ_n	Gaussian probability density function in n dimensions
Φ_n	Gaussian cumulative probability function in n dimensions
χ	Concentration of pollutant in air (usually ppm)
Ψ	$\Psi(x) = \phi(x) / \Phi(-x)$

APPENDIX E

MLFP VALUES FROM TEST CASES

Table E1
MLFP values for GASTAR/CRUNCH common case:
GASTAR / Q_{lim} = 0.005 (see Figure 5)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.34	-	6.68	8.87	7.93	13.4	19.4	26.6
Aerosol fraction (-)	0.5	-	0.522	0.505	0.505	0.496	0.494	0.485
Air temperature (K)	285	-	290	290	285	285	286	286
Air entrained ratio (-)	9.58	-	9.58	9.58	9.59	9.59	9.6	9.56
Source aspect ratio (-)	1.0	-	1.0	1.0	1.0	1.0	1.0	1.0
Wind speed (m s^{-1})	2.35	-	2.15	2.27	2.36	2.22	2.06	1.96
Roughness length (m)	0.0701	-	0.0735	0.0708	0.072	0.0688	0.0666	0.0637
Friction factor (-)	0.01	-	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	-	25000	25000	25000	25000	25000	25000
Toxicity mean (-)	14.44	-	14.9	14.5	14.6	14.2	13.8	13.4
Toxicity stand. dev. (-)	1.087	-	1.08	1.08	1.08	1.10	1.11	1.13
Relative humidity (%)	50	-	50.1	50.0	50.1	50.0	49.9	50.0

Table E2
MLFP values for GASTAR/CRUNCH common case:
GASTAR / Q_{lim} = 0.1 (see Figure 6)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.34	5.32	8.88	8.51	11.5	18.5	25.7	34.0
Aerosol fraction (-)	0.5	0.54	0.51	0.508	0.49	0.481	0.477	0.466
Air temperature (K)	285	289	290	286	284	284	284	285
Air entrained ratio (-)	9.58	9.56	9.58	9.59	9.57	9.58	9.59	9.58
Source aspect ratio (-)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Wind speed (m s^{-1})	2.35	2.19	2.29	2.29	2.38	2.30	2.2	2.04
Roughness length (m)	0.0701	0.0763	0.0708	0.0714	0.0688	0.0662	0.0647	0.0611
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	25000	25000	25000	25000	25000	25000	25000
Toxicity mean (-)	14.44	15.3	14.5	14.6	14.2	13.7	13.2	12.7
Toxicity stand. dev. (-)	1.087	1.1	1.09	1.09	1.09	1.10	1.11	1.13
Relative humidity (%)	50	50.0	50.0	50.2	49.9	49.9	50.1	50.0

Table E3
MLFP values for GASTAR/CRUNCH common case:
GASTAR / Q_{lim} = 0.5 (see Figure 7)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.34	7.42	9.34	12.1	16.5	24.6	32.5	40.1
Aerosol fraction (-)	0.5	0.518	0.50	0.473	0.473	0.46	0.44	0.43
Air temperature (K)	285	289	285	283	283	283	284	285
Air entrained ratio (-)	9.58	9.57	9.58	9.55	9.59	9.66	9.6	9.63
Source aspect ratio (-)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Wind speed (m s^{-1})	2.35	2.25	2.35	2.53	2.59	2.50	2.31	2.04
Roughness length (m)	0.0701	0.0727	0.0707	0.0679	0.0660	0.0641	0.0613	0.0576
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	25000	25000	25000	25000	25000	25000	28100
Toxicity mean (-)	14.44	14.8	14.4	14.0	13.6	13.0	12.5	11.8
Toxicity stand. dev. (-)	1.087	1.09	1.09	1.09	1.09	1.10	1.10	1.12
Relative humidity (%)	50	50.0	50.0	50.0	49.8	50.2	50.2	50.1

Table E4
MLFP values for GASTAR/CRUNCH common case:
GASTAR / Q_{lim} = 0.9 (see Figure 8)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.34	9.15	13.2	16.8	21.8	30.9	38.2	42.6
Aerosol fraction (-)	0.5	0.501	0.473	0.448	0.44	0.421	0.418	0.404
Air temperature (K)	285	285	283	283	282	282	282	284
Air entrained ratio (-)	9.58	9.58	9.58	9.61	9.72	9.69	9.71	9.65
Source aspect ratio (-)	1.0	1.0	0.999	1.01	1.0	1.0	1.0	1.0
Wind speed (m s^{-1})	2.35	2.34	2.56	2.82	2.87	2.79	2.59	2.16
Roughness length (m)	0.0701	0.0709	0.0674	0.0659	0.0642	0.0631	0.0590	0.0539
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	25000	25000	25000	25000	25400	26400	30700
Toxicity mean (-)	14.44	14.5	13.9	13.4	12.9	12.2	11.5	10.6
Toxicity stand. dev. (-)	1.087	1.09	1.08	1.08	1.08	1.07	1.07	1.09
Relative humidity (%)	50	50.0	49.9	49.7	49.7	51.0	50.5	50.3

Table E5
MLFP values for GASTAR/CRUNCH common case:
CRUNCH / Q_{lim} = 0.005 (see Figure 12)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.34	2.86	4.76	7.88	11.9	21.7	33.5	35.0
Aerosol fraction (-)	0.5	0.517	0.503	0.504	0.491	0.5	0.499	0.412
Air temperature (K)	285	285	284	285	285	285	285	285
Air entrained ratio (-)	9.58	9.55	9.53	9.58	9.58	9.58	9.58	9.6
Source aspect ratio (-)	1.0	1.0	0.997	1.0	1.0	1.0	1.0	1.0
Wind speed (m s^{-1})	2.35	3.71	2.71	2.43	2.13	1.9	1.6	1.29
Roughness length (m)	0.0701	0.0775	0.718	0.0712	0.0694	0.0689	0.0685	0.0641
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	25000	25000	25000	25000	25000	25000	25500
Toxicity mean (-)	14.44	15.8	14.9	14.6	14.3	13.8	13.7	12.6
Toxicity stand. dev. (-)	1.087	1.05	1.08	1.08	1.09	1.12	1.16	1.15
Relative humidity (%)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	

Table E6
MLFP values for GASTAR/CRUNCH common case:
CRUNCH / Q_{lim} = 0.5 (see Figure 13)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.34	4.61	9.13	13.9	21.3	33.6	41.3	46.7
Aerosol fraction (-)	0.5	0.556	0.501	0.481	0.443	0.432	0.366	0.381
Air temperature (K)	285	285	285	285	286	285	285	285
Air entrained ratio (-)	9.58	9.56	9.58	9.58	9.58	9.58	9.72	10.1
Source aspect ratio (-)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Wind speed (m s^{-1})	2.35	2.62	2.35	2.22	2.09	1.83	1.54	1.30
Roughness length (m)	0.0701	0.0766	0.0709	0.0684	0.0661	0.0626	0.0542	0.0656
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	25000	25000	25000	25000	25700	28700	36600
Toxicity mean (-)	14.44	14.9	14.5	14.1	13.4	12.4	11.5	10.5
Toxicity stand. dev. (-)	1.087	1.09	1.09	1.09	1.09	1.10	1.11	1.19
Relative humidity (%)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Table E7
MLFP values for Variant Case 1 ('D' stability):
GASTAR / Q_{lim} = 0.5 (see Figure 15)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.34	7.43	14.2	23.8	24.1	28.2	29.3	-
Aerosol fraction (-)	0.5	0.513	0.474	0.468	0.482	0.468	0.495	-
Air temperature (K)	285	285	285	290	293	294	293	-
Air entrained ratio (-)	9.58	9.56	9.60	9.60	9.60	9.63	9.47	-
Source aspect ratio (-)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	-
Wind speed (m s^{-1})	2.35	2.25	2.56	2.40	0.981	0.553	0.224	-
Roughness length (m)	0.0701	0.0733	0.0672	0.0624	0.0565	0.0518	0.0522	-
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-
Inventory (kg)	25000	25000	25000	25000	25000	28600	25000	-
Toxicity mean (-)	14.44	14.7	13.9	13.4	13.3	12.8	12.9	-
Toxicity stand. dev. (-)	1.087	1.09	1.09	1.10	1.09	1.11	1.09	-
Relative humidity (%)	50	50.0	50.0	50.0	50.0	50.0	50.0	-

Table E8
MLFP values for Variant Case 2 (Sudden release):
GASTAR / Q_{lim} = 0.5 (see Figure 18)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	*2000*	10000	20000
Release size (kg)	25000	22800	24800	27200	31700	*28100*	-	-
Aerosol fraction (-)	0.5	0.517	0.499	0.483	0.469	*0.511*	-	-
Air temperature (K)	297.5	297.5	297.5	297.5	297.5	*297.5*	-	-
Air entrained ratio (-)	49.0	49.7	49.1	48.8	48.0	*48.3*	-	-
Source aspect ratio (-)	0.928	0.979	0.932	0.910	0.883	*0.918*	-	-
Wind speed (m s^{-1})	4.7 *(2.35)*	5.48	4.80	4.46	3.78	*1.70*	-	-
Roughness length (m)	0.022	0.0226	0.0223	0.0216	0.0182	*0.0192*	-	-
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	*0.01*	-	-
Toxicity mean (-)	14.44	15.3	14.5	14.1	13.2	*13.9*	-	-
Toxicity stand. dev. (-)	1.087	1.09	1.09	1.09	1.09	*1.09*	-	-
Relative humidity (%)	50	50.0	50.0	50.0	49.9	*50.1*	-	-

Note: The column in italics gives the results of a repeat run for d = 2000 m in which the windspeed distribution was altered from that shown in Table 7 of the main report to the lighter wind distribution used for the continuous release cases (Table 4).

Table E9
MLFP values for Variant Case 3. (Continuous release with $T_a > T_g$). GASTAR / Q_{lim} = 0.005 (see Figure 20)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.30	3.40	6.74	10.9	16.0	22.9	27.7	28.8
Aerosol fraction (-)	0.50	0.539	0.506	0.498	0.495	0.497	0.497	0.497
Air Temperature (K)	297.5	298	297	298	298	298	298	298
Air entrained ratio (-)	9.58	9.53	9.57	9.58	9.58	9.59	9.58	9.57
Source aspect ratio (-)	1.00	0.999	1.00	1.00	1.00	1.00	1.00	1.00
Wind speed (m s^{-1})	2.35	2.18	2.46	2.26	1.89	1.28	0.83	0.49
Roughness length (m)	0.0707	0.0880	0.0743	0.0690	0.0645	0.0591	0.0536	0.0496
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	25000	25000	25000	25000	25000	25000	28400
Toxicity mean (-)	14.44	15.6	14.7	14.3	14.0	13.7	13.4	13.1
Toxicity stand. Dev. (-)	1.087	1.06	1.08	1.09	1.10	1.12	1.13	1.14
Relative humidity (%)	50.0	49.6	49.9	50.0	50.1	50.1	50.4	50.5

Table E10
MLFP values for Variant Case 3. (Continuous release with $T_a > T_g$). GASTAR / Q_{lim} = 0.5 (see Figure 21)

Quantity	BEP value	MLFP values at distance from source (m):						
		200	500	1000	2000	5000	10000	20000
Hole diameter (mm)	9.30	6.88	12.9	17.9	22.3	28.2	27.1	27.8
Aerosol fraction (-)	0.50	0.505	0.494	0.491	0.490	0.490	0.493	0.494
Air temperature (K)	297.5	297	298	298	298	298	298	298
Air entrained ratio (-)	9.58	9.56	9.59	9.60	9.59	9.60	9.51	9.51
Source aspect ratio (-)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Wind speed (m s^{-1})	2.35	2.47	2.20	1.67	1.17	0.663	0.292	0.13
Roughness length (m)	0.0707	0.0750	0.0671	0.0617	0.0579	0.0523	0.0468	0.0457
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	25000	25000	25000	25000	25000	27500	28900
Toxicity mean (-)	14.44	14.7	14.2	13.8	13.6	13.2	12.9	12.9
Toxicity stand. dev. (-)	1.087	1.09	1.09	1.09	1.10	1.10	1.10	1.14
Relative humidity (%)	50.0	49.8	50.1	50.2	50.1	50.6	50.9	50.3

Table E11
MLFP values for Ammonia release case: GASTAR / Q_{lim} = 0.005 (see Figure 22)

Quantity	BEP value	MLFP values at distance from source (m):				
		200	500	1000	2000	5000
Hole diameter (mm)	9.34	12.0	15.9	18.8	24.7	34.7
Aerosol fraction (-)	0.5	0.457	0.350	0.281	0.242	0.193
Air temperature (K)	285	285	284	284	284	286
Air entrained ratio (-)	9.58	9.57	9.56	9.53	9.54	9.54
Source aspect ratio (-)	1.0	1.0	1.0	1.0	1.0	1.0
Wind speed (m s^{-1})	2.35	2.39	2.46	2.45	2.16	1.56
Roughness length (m)	0.0701	0.0694	0.0678	0.0672	0.0658	0.0638
Friction factor (-)	0.01	0.01	0.01	0.01	0.01	0.01
Inventory (kg)	25000	25000	25000	25000	25000	25300
Toxicity mean (-)	22.1	21.9	21.5	21.3	21.0	20.5
Toxicity stand. dev. (-)	0.54	0.552	0.566	0.579	0.586	0.582
Relative humidity (%)	50	50.2	50.5	51.0	51.4	52.1

Printed and published by the Health and Safety Executive
C1 10/01

MAIL ORDER
HSE priced and free
publications are
available from:
HSE Books
PO Box 1999
Sudbury
Suffolk CO10 2WA
Tel: 01787 881165
Fax: 01787 313995
Website: www.hsebooks.co.uk

RETAIL
HSE priced publications
are available from booksellers

HEALTH AND SAFETY INFORMATION
HSE InfoLine
Tel: 08701 545500
Fax: 02920 859260
e-mail: hseinformationservices@natbrit.com
or write to:
HSE Information Services
Caerphilly Business Park
Caerphilly CF83 3GG

HSE website: www.hse.gov.uk

CRR 338

£20.00

ISBN 0-7176-2013-1

9 780717 620135