

Propagation of uncertainty through the hazard chain

Problem presented by

Mike Redwood
Joe Gillard

DSTL

Executive Summary

DSTL wish to explore methods for propagating uncertainty through a succession of linked models. The Study Group have looked at the particular example of casualty estimation from airbourne dispersion and suggested two different potential solutions. If the structure of the models is sufficiently simple, and the number of degrees of freedom relatively small, a semi-analytical approach based on Bayes' theorem can be used. In the more general case, intelligent sampling methods can be used to gradually build a picture of likely outcomes.

Version 1.1
June 15, 2012
iii+13 pages

Report author

Chris Cawthorn (Industrial Mathematics KTN)

Contributors

David Barton (University of Bristol)

Russell Davies (Cardiff University)

Melissa Maczka (University of Oxford)

Zubin Siganporia (University of Oxford)

Andrzej Tolarczyk (Polish Academy of Sciences)

ESGI85 was jointly organised by

The University of East Anglia

The Knowledge Transfer Network for Industrial Mathematics

and was supported by

The Engineering and Physical Sciences Research Council

Contents

1	Introduction	1
1.1	Background and scope	1
2	Uncertainty propagation	2
2.1	Details of the dispersion model	2
2.2	Details of the casualty model	3
2.3	Approach 1: Sampling	4
2.4	Approach 2: Semi-analytical	7
3	Other sources of uncertainty	10
3.1	Variability in the population	10
3.2	Parametric uncertainty	11
4	Further extensions	11
5	Conclusions	12
A	Online resources	12
	Bibliography	13

1 Introduction

Toxic substances released into the environment pose both an immediate and delayed risk to human health. When this release is in the form of a gas or vapour it is necessary to predict where the substance will disperse and deposit in the environment as this will allow a first responder to undertake appropriate mitigation strategies. To this end many organisations have worked to produce models which predict parts of this process. However, in order to produce an estimate of casualties that may result from exposure to the substance these disparate models must be tied together. Importantly, it is not only necessary to predict a casualty estimate but also to have an associated uncertainty with this value. DSTL challenged the Study Group to improve upon their existing methodology to incorporate uncertainty into the resulting casualty estimates.

1.1 Background and scope

- (1.1.1) In order to produce casualty estimates following the release of toxic substances into the environment, DSTL have developed a suite of disparate models that attempt to capture different parts of the system, from release, via dispersion, through to the expected 'dose' received by individuals and an eventual casualty estimate. These models are referred to as the *hazard chain*, and are described in more detail below.
- (1.1.2) **Meteorological models** provide forcing to the dispersal of the toxic substance (e.g. winds). Typically, several different meteorological scenarios will be investigated.
- (1.1.3) An understanding of the **release mechanism**, including the location and volume of release, are assumed to be known, and provide an initial condition to the dispersion model.
- (1.1.4) A **dispersion model** is used to track the motion of the toxic substance in space and time. The model used will vary depending on the local terrain, but could be the Second-order Closure Integrated Puff model (SCIPUFF) for open terrain or the Urban Dispersion Model (UDM) in built-up areas. It is understood that each of these models accounts for some uncertainty, both by performing an ensemble average over several (meteorological) conditions and by parametering intrinsic uncertainty that might arise due to turbulence, for example. The output from the dispersion model is in the form of the mean and variance in substance concentration at each point in space and time used in the numerical grid.
- (1.1.5) A **casualty model** that calculates the chemical dosage received by an individual over time, and estimates the expected casualty rate in the population by way of a probit curve.

- (1.1.6) The principal flaw with the existing approach is that the casualty model formally requires the *actual* concentration at each point in time and space, whereas statistical nature of the dispersion model output means that only the *mean* concentration is available for use in the casualty model. This means that extreme concentrations (which are likely to have a significant effect on casualty rates) are not accounted for by the casualty model.
- (1.1.7) The problem set for the Study Group was to suggest methodologies to allow uncertainty to be propagated correctly between models, focussing on incorporating the statistical output of the dispersion model into the casualty model. In addition, there was to be some discussion regarding how to handle other types of uncertainty in the system, such as parametric uncertainty and variability within the population.

2 Uncertainty propagation

This section addresses the main task set for the Study Group: determining how to calculate the distribution of casualty rates given the distribution of concentrations, as parameterised by its statistical moments. After a brief description of the dispersion and casualty models, we present two methods for uncertainty propagation: Monte-Carlo-like sampling and a semi-analytical approach based on Bayes' Theorem.

2.1 Details of the dispersion model

- (2.1.1) The Urban Dispersion Model (UDM) used by DSTL is a Gaussian puff model that takes as input a set of meteorological conditions and release parameters. It performs a set of simulations with stochastic processes representing the uncertainty in the forcing (e.g. winds). More details of this model can be found in [1], and references therein.
- (2.1.2) The output from the UDM is given in the form of the mean and variance calculated over the ensemble of simulations carried out. Without accounting for the distribution associated with this mean and variance, one cannot capture fluctuations and extremes of concentrations. However, it is precisely these extreme values that can be most significant when estimating casualty rates.
- (2.1.3) It is considered in the established literature [2] that the clipped normal distribution is the most appropriate distribution to represent the concentrations modelled by the UDM and other Gaussian puff models. The clipped normal distribution has the same familiar bell shape as the usual normal distribution, but the probability density of negative concentrations is set to zero. The probabilistic weight removed from the negative concentrations is then applied at zero concentration. The probability density

function is therefore:

$$p(c|\hat{\mu}, \hat{\sigma}) = \delta(c)\mathbb{P}\{C = 0|\hat{\mu}, \hat{\sigma}\} + \frac{H(c)}{\sqrt{2\pi}\hat{\sigma}^2} \exp\left(-\frac{(x - \hat{\mu})^2}{2\hat{\sigma}^2}\right), \quad (1)$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the mean and variance of the un-clipped distribution and $H(c)$ is the Heaviside function (zero for negative c , one otherwise). The finite probability of zero concentration is represented by the delta function term. The probability of zero concentration is calculated from the standard normal cumulative distribution function:

$$\mathbb{P}\{C = 0|\hat{\mu}, \hat{\sigma}\} = \Phi(-\hat{\mu}/\hat{\sigma}). \quad (2)$$

(2.1.4) It is important to note that $\hat{\mu}$ and $\hat{\sigma}^2$ are not the mean and variance of the clipped normal distribution. The relevant transformation is:

$$\mu = \mathbb{E}(C) = \frac{\hat{\sigma}}{\sqrt{2\pi}} + \hat{\mu} \left[1 - \Phi\left(-\frac{\hat{\mu}}{\hat{\sigma}}\right)\right], \quad (3)$$

$$\sigma^2 = \text{Var}(C) = \hat{\sigma}^2 \left[1 - \Phi\left(-\frac{\hat{\mu}}{\hat{\sigma}}\right)\right] - \mu(\mu - \hat{\mu}). \quad (4)$$

We will use this transformation later when working with the clipped normal distribution.

2.2 Details of the casualty model

(2.2.1) The casualty model comprises two stages, described in detail by [5]. First of all, the *inhaled dose*, D_{inhal} (henceforth referred to as the dose, for simplicity) of the toxic chemical is computed as a function of time using the integral expression:

$$D_{inhal}(t) = R_c R_{BR} \int_0^t \left(\frac{BR \max(c(t') - c_{thr}, 0)}{R_c R_{BR}} \right)^n dt', \quad (5)$$

where

- c_{thr} is threshold concentration
- R_c is the reference concentration level in kg/m^3
- BR is the breathing rate in m^3/s
- R_{BR} is the reference breathing rate in m^3/s

and n is the *toxic load exponent*, which reflects the relative severity of inhaling larger concentrations for short times when compared to inhaling smaller concentrations over a longer timeframe. Typically, all of these quantities save the breathing rate will be assumed constant in time and across the population. For now, we also assume a constant breathing rate, and defer a discussion of variable breathing rate to later sections.

- (2.2.2) The concentration used in the dose calculation ought to be the *actual* concentration. However, in the absence of any information other than the statistics of concentration obtained from the dispersion model, the current procedure used by DSTL is to use the *mean* concentration when calculating the dose.
- (2.2.3) Having calculated the dose, the next step is to calculate the effect on the population. To capture the variation of resistance to the chemical, DSTL use a probit curve, which gives the probability of a particular response given the dose. The form of probit curve used is:

$$B(D|D_{50}, \beta) = \frac{1}{2} \left\{ 1 + \operatorname{erf} \left[\frac{\beta}{\sqrt{2}} \log_{10} \left(\frac{D}{D_{50}} \right) \right] \right\}, \quad (6)$$

where D_{50} is the dose for which one expects 50% of the population to suffer a response. Different values of D_{50} apply to different responses, increasing in severity from eye irritation, through incapacitation, to death. The probit slope, β , which is estimated from toxicology studies, will also vary depending on effect.

- (2.2.4) To compute a casualty estimate, DSTL currently use a so-called ‘Lucky Number’ approach, by which it is understood that a number of values equal to the local population size are sampled uniformly on the interval $[0,1]$. If a number lies below the probit value for the observed concentration, a casualty is recorded. It was pointed out by the Study Group that simply multiplying the probit by the population ought to give the same response statistically.
- (2.2.5) During the Study Group, the principal focus was on incorporating the uncertainty associated with the concentrations in the dose calculation, with the understanding that proceeding to a casualty estimate via the probit curve is a trivial extension. Furthermore, we begin by considering the concentration at only a single point in space. The consequences of spatial variation and correlation in concentration will be discussed in Section 4. In what follows, we shall therefore determine a means of estimating the distribution of doses at a single point in space given time series data of the mean and variance of concentration at that point.

2.3 Approach 1: Sampling

- (2.3.1) The first method of estimating the dose distribution suggested by the Study Group was a straightforward sampling approach. Given the distribution of concentrations at each point in time, it is possible to randomly generate a time series of concentrations using the clipped normal distribution described in Section 2.1. Such a time series can be inserted into the dose calculation (5) and the resulting dose computed. This process can

repeated a large number of times and the statistics of the calculated doses can be studied to estimate the distribution of doses. There are many different sampling approaches designed to propagate distributions through models. In this report, we consider a simple version. A much wider and more sophisticated set of tools can be found in the work of the Modelling Uncertainty in Complex Models (MUCM) research group [4].

- (2.3.2) An important observation is that successive concentrations in the time series will be correlated with one another, provided that the timesteps reflected in the output from the dispersion model are sufficiently short. The Study Group chose to model the correlation function as a smoothly decaying Gaussian:

$$\text{Corr}(C_i, C_j) = \exp\left(-\frac{(t_i - t_j)^2}{t_D^2}\right), \quad (7)$$

where t_D is a timescale chosen to reflect the typical time over which concentrations are correlated. This could be chosen based on experience of the dispersion model, or directly calculated and output by the UDM along with the mean and variance of the concentration. Typical correlation times calculated by the dispersion model are of the order of 5 minutes.

- (2.3.3) When sampling concentrations, it is necessary to enforce the correlation between concentrations. For most distributions, including the clipped Gaussian, this is very difficult to achieve while preserving the marginal distribution of each concentration. As the number of degrees of freedom (that is, the number of timesteps) increases, the difficulty and cost of this sampling will increase significantly. Worse still, one ought to also include spatial correlations as well as temporal correlations. Given that a typical simulation may use 10,000 grid points and 500 timesteps, one would need to sample approximately 5M correlated random variables for each dose calculation. Building a good distribution could easily require tens of thousands of individual dose calculations, making this procedure quite computationally intensive. However, it is possible to exploit the reasonably strong correlations expected to vastly reduce the number of degrees of freedom at the cost of a small loss of fidelity in representing the specified correlation. Given that the exact form of the correlation is uncertain, this does not constitute a large drawback. One such method of reducing the dimensionality of the problem is the Karhunen-Loève expansion.
- (2.3.4) The Karhunen-Loève (K-L) expansion [6] is a means of representing an infinite dimensional stochastic process as the linear combination of an infinite number of orthogonal basis functions, where the linear factors are uncorrelated random variables. The K-L expansion has the property that the variance is maximised for the first random variable and the variance monotonically decreases for the subsequent random variables. This means

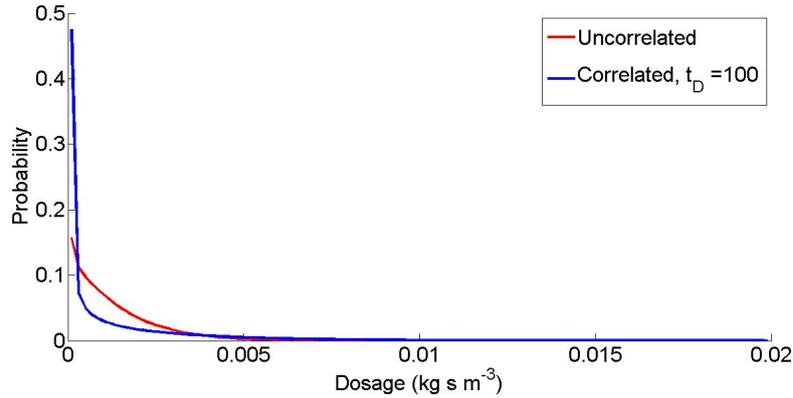


Figure 1: Plots of 10,000 dosages simulated from example mean concentration data with (red) no concentration temporal correlation and (blue) a concentration temporal correlation timescale of 100.

that the K-L expansion is, in some senses, optimal if a finite truncation of the expansion is desired. For a particular stochastic process $X(t)$, the definition of the K-L expansion is

$$X(t) = \sum_{i=0}^{\infty} Z_k e_k(t) \quad (8)$$

where Z_k are uncorrelated random variables and e_k are the eigenfunctions of the covariance function of $X(t)$. The K-L expansion is the continuous-time analogue of proper orthogonal decomposition (POD).

(2.3.5) There are arbitrarily many ways to transform correlated random variables into uncorrelated random variables for sampling purposes. Another method, which does not permit a reduction in the dimensionality of the problem, is to use the Cholesky decomposition of the covariance matrix.

(2.3.6) Cholesky decomposition was used to simulate 10,000 dosage values for a single spatial location using example mean concentration data provided by DSTL. Two types of data were simulated to realise the importance of modelling temporal correlations in concentration (see Fig. 1):

- 1) temporally independent data, and
- 2) data with a temporal correlation timescale of 100 (a value that was arbitrarily chosen for demonstration purposes only).

Fig. 1 shows that the dosage distribution is different when temporal correlations in concentration are modelled. From this it can be concluded that it is important to model concentration temporal concentrations, despite the increased computational cost.

It was also noticed and verified that the simulated dosages are well approximated by the gamma distribution (see Fig. 2).

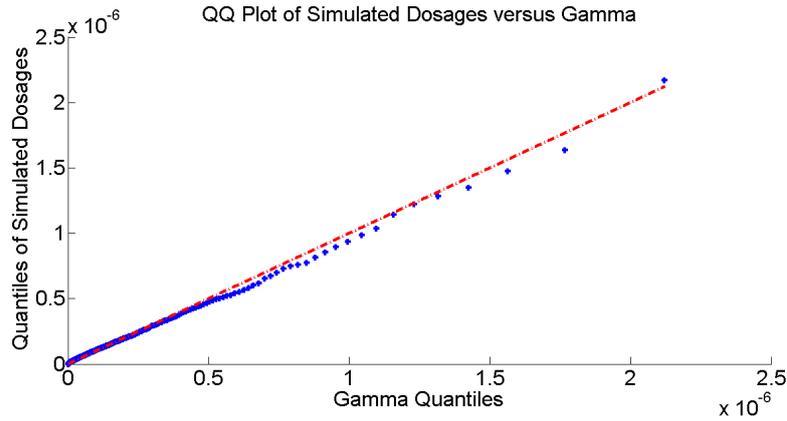


Figure 2: QQplot comparing simulated dosages (with a concentration temporal correlation timescale of 100) to gamma-distributed data. Notice the goodness of fit to $y=x$, indicating that the simulated dosages are well approximated by the gamma distribution.

- (2.3.7) Before leaving this discussion of sampling, it bears repeating that everything discussed above can apply equally to casualty estimates as it does to dose - it is simply a matter of proceeding to calculate the casualty estimate from the dose via the probit function, and building a distribution of casualty estimates from the resulting data.

2.4 Approach 2: Semi-analytical

- (2.4.1) The sampling method described above can be used for any model. All that is needed is knowledge of the distribution of the input variable(s) (concentration in this example). The casualty model itself may be treated as a black box whose output may be used to build a distribution of dose or casualty rate. In this section, we discuss a more direct method that exploits our knowledge of the structure of the casualty model. The idea is to employ Bayes' Theorem to integrate across the concentration distribution, thereby accounting for the variability in concentration when computing casualty estimates.
- (2.4.2) Suppose that $\mathbf{C} = (C_1, C_2, \dots, C_N)$ are random variables representing the concentration at each point in space and in the dispersion model. We know that concentration C_i has mean μ_i and variance σ_i^2 , and that concentrations follow the clipped Gaussian distribution with probability density function $p_C(c|\mu, \sigma)$ given by (1), with $\hat{\mu}$ and $\hat{\sigma}$ found by inverting the relationship given by (3-4). Bayes' theorem then tells us that the dose, D , has probability density function:

$$f_D(d) = p_D(d|\mu, \sigma) = \int p(d|\mathbf{c}) p_C(\mathbf{c}|\mu, \sigma) d\mathbf{c}, \quad (9)$$

where $\mu = (\mu_1, \mu_2, \dots, \mu_N)$ and $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$.

- (2.4.3) The conditional probability density of the dose given the concentration history, $p(d|\mathbf{c})$, is a *known* function described by the casualty model. In the simplest case, described in Section 2.2, the dose is a deterministic function of concentration given by the integral expression (5), thus:

$$p(d|\mathbf{c}) = \delta(d - D_{inhal}(\mathbf{c})). \quad (10)$$

The presence of this delta-function in (9) has the effect of restricting the volume of integration to include only those values of \mathbf{c} such that $D_{inhal}(\mathbf{c}) = d$, hence:

$$f_D(d) = \int_{\mathbf{c}:D(\mathbf{c})=d} p_C(\mathbf{c}|\mu, \sigma) d\mathbf{c}. \quad (11)$$

This new integration domain will typically occupy $N - 1$ dimensions, and may have a complicated shape depending on the value of the toxic load exponent chosen and on any time-dependence in the breathing rate.

- (2.4.4) A slightly simpler approach would be to compute the *cumulative distribution function*, rather than the probability density function:

$$F_D(d) = \mathbb{P}\{D < d\} = \int H(d - D(\mathbf{c}))p_C(\mathbf{c}|\mu, \sigma) d\mathbf{c}. \quad (12)$$

Here we use the Heaviside function:

$$H(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x \leq 0, \end{cases} \quad (13)$$

to ensure that only concentration histories that give rise to doses less than or equal to d are included. As with (11), one could modify the region of integration to remove this function:

$$F_D(d) = \int_{\mathbf{c}:D(\mathbf{c}) \leq d} p_C(\mathbf{c}|\mu, \sigma) d\mathbf{c}. \quad (14)$$

- (2.4.5) It is worth reminding ourselves at this point that concentrations are correlated in space and time and therefore the form of $p_C(c)$ is necessarily complicated by the need to account for these correlations. As with sampling, it is possible to handle correlations by using the K-L expansion (see eq. 8). By carefully choosing a set of *independent* random variables, $\mathbf{Z} = (Z_1, Z_2, \dots, Z_k)$, we represent each concentration as a linear combination of these variables, $\mathbf{C}(\mathbf{Z})$. This not only reduces the dimensionality of the problem, but also results in a simpler probability density function:

$$p_C(\mathbf{c}|\mu, \sigma) = \int_{\mathbf{x}:C(\mathbf{x})=\mathbf{c}} p_Z(\mathbf{z}|\mu, \sigma) d\mathbf{z} = \int_{\mathbf{z}:C(\mathbf{z})=\mathbf{c}} \prod_{i=1}^k p_{Z_i}(z_i) d\mathbf{z}. \quad (15)$$

Alternatively, one could bypass the concentration altogether, write the dose as a function of K-L variables:

$$D_{inhal} = D(\mathbf{C}) = D(\mathbf{C}(\mathbf{Z})), \quad (16)$$

and rewrite the cumulative distribution function (14) in the form:

$$F_D(d) = \int_{\mathbf{z}: D(\mathbf{C}(\mathbf{z})) \leq d} p_Z(\mathbf{z}|\mu, \sigma) d\mathbf{z}, \quad (17)$$

with

$$p_Z(\mathbf{z}|\mu, \sigma) = \prod_{i=1}^k p_{Z_i}(z_i). \quad (18)$$

By exploiting the finding from Section 2.3 that the dosages were well approximated by gamma distributions, it would be possible to dramatically reduce the dimensionality of this problem even further with minimal loss of information. This is because dosage can then be described by only the first two moments of a gamma distribution.

- (2.4.6) The K-L expansion offers a marked reduction in the difficulty of evaluating the joint probability density function for concentration, and thus expedites the computation of the probability density function for dose. However, the integration domain (i.e. those values of \mathbf{z} such that $D(\mathbf{C}(\mathbf{z})) \leq d$) could take on a complicated shape depending on the model parameters (particularly the toxic load exponent). For this reason, it is most likely that this integral will need to be evaluated numerically.
- (2.4.7) Many multi-dimensional numerical quadrature techniques for approximating integrals are based on one-dimensional quadrature rules such as Curtis-Clelland or Gaussian quadrature which use a non-equispaced grid over the domain (see [7] for a review, and [8] for code). Typically, grid points are clustered at the boundaries of the domain. To apply a one-dimensional quadrature rule, the grid over which the integrand is evaluated is defined by a tensor product of the corresponding one-dimensional grid. The use of a tensor product means that the number of grid points required grows with the factorial of the number of dimensions, thus leading to severe limitations in its applicability (on modern computer hardware, six dimensions is often cited as an upper limit). An alternative is to use a Smolyak-type sparse grid which reduces the number of grid points needed dramatically, allowing quadrature techniques to be used in up to 10–20 dimensions on modern computer hardware.
- (2.4.8) To approximate a very high dimensional integral, Monte-Carlo methods may be required.

3 Other sources of uncertainty

In addition to the main discussion regarding uncertainty propagation, the Study Group also considered how to incorporate other sources of uncertainty into the casualty model. In this more speculative analysis, distinction was drawn between two different types of uncertainty: parametric uncertainty, which arises from unknown, but constant parameter values; and natural variation across the population. This section describes both sources of uncertainty, and proposes a methodology for bringing them into the casualty model.

3.1 Variability in the population

(3.1.1) Several sources of uncertainty in the model may be traced back to variation between individuals in the population. An attempt is already made to model the variation in susceptibility by way of the probit curve used to estimate casualty rates from chemical doses. Another example is breathing rate, which varies from one person to the next, but is expected to follow some distribution that can be predicted based on both natural variation and difference due to different activity levels between individuals.

(3.1.2) The uncertainty in breathing rate can be captured in much the same way as concentration, using either the sampling or semi-analytical approaches described in Section 2. Consider as an example the semi-analytical approach of Section 2.4.

(3.1.3) The conditional cumulative distribution function of dose given the breathing rate is:

$$F_{D|BR}(d, BR) = \mathbb{P}\{D \leq d | \mu, \sigma, BR\} = \int_{\mathbf{c}: D(\mathbf{c}, BR) \leq d} p_C(\mathbf{c} | \mu, \sigma) d\mathbf{c}. \quad (19)$$

Note that the dependency on breathing rate has been made explicit in $D(\mathbf{c}, BR)$. The function form is exactly the same as in (5). Applying Bayes' theorem, one can write:

$$F_D(d) = \mathbb{P}\{D \leq d | \mu, \sigma\} = \int F_{D|BR}(d, b) p_{BR}(b) db, \quad (20)$$

where $p_{BR}(b)$ is the probability density function for the breathing rate. This is exactly the same treatment given to the concentration variables, although it is vastly simplified by the fact that there is only one breathing rate per individual, as opposed to a whole time series of concentrations.

(3.1.4) Other variations (e.g. in threshold concentration or metabolism time) may be included in the same manner, with each variable requiring an extra integral to account for its variation.

3.2 Parametric uncertainty

- (3.2.1) Parametric uncertainty arises from choosing model parameters with a fixed value without being certain about that value. For example, a toxicologist might suggest a toxic load exponent, $n = 1.3 \pm 0.05$, for a given chemical. The true value will be fixed for that chemical, but the experimental error must be accounted for. In the probit model used currently, the parameters D_{50} and β have the same properties.
- (3.2.2) There are two main approaches to incorporating this parametric uncertainty into the model, stemming from Bayesian and frequentist statistics respectively.
- (3.2.3) The Bayesian approach involves assuming a probability distribution for the unknown parameter value. The associated probability density function can then be used to extend the Bayesian integral in exactly the same way as for natural variability (see Section 3.1).
- (3.2.4) A frequentist would argue that it is not appropriate to specify a distribution for an unknown parameter. Indeed, there is only one true value of the parameter, so any distribution could only express our relative confidence in different values. If our confidence is misplaced for whatever reason, there is a risk that the conclusions will be spurious. The frequentist approach avoids the choice and fitting of a distribution altogether, favouring instead a sensitivity analysis of the parameter. In its crudest form, the ‘distribution-free’ sensitivity analysis could simply involve computing the distribution of casualty estimates for a range of plausible parameters, and analysing the resulting data to look for a best- and worst- case scenario.
- (3.2.5) A more sophisticated sensitivity analysis might use some form of optimisation algorithm to seek the worst case within the plausible parameter range. The reader is directed to [3] for a discussion of optimisation strategies aimed at maximising a function (in this case some statistic of the casualty rate distribution) efficiently (i.e. testing as few parameter values as possible).

4 Further extensions

A handful of other topics were discussed at the Study Group, but the time constraints prevented them from being pursued in detail. These topics are listed in this section.

The probit curve approach (in particular, the ‘Lucky Number’ method) used in the existing casualty estimate sits slightly at odds with the treatment of uncertainty proposed by the Study Group. It was suggested that a more appropriate means of capturing the variation in susceptibility among the population would be to have a *response curve* for each individual, describing the probability of each effect (i.e.

incapacitation, death, etc.) at a given dose *for that individual*. The response curve would have the same ‘sigmoid’ shape as the probit curves, and could even be parameterised by β and D_{50} . Here, however, D_{50} should be interpreted as the dose at which the probability of suffering the effect is precisely 50%. Where this approach differs from the probit curve approach is that the parameters D_{50} and β may vary from individual to individual, accounting for natural variation in their general state of health, or indeed protective equipment they might have. In exactly the same manner as breathing rate is treated in Section 3.1, distributions of D_{50} and β could be constructed and integrated over in order to reflect the variation in susceptibility in the casualty estimate. This obviates the need for any further sampling (such as the ‘Lucky Number’ method).

The previous sections all deal with breathing rates that are constant in time for each individual. This is a highly unrealistic assumption, as individuals’ breathing rates may change as a result of a change in activity level (e.g. going from sleeping to walking to running), or as a result of increased stress as the effects of the chemical attack become apparent. An interesting extension to the existing model would consider time-dependent breathing rates, perhaps by modelling the statistics of transition from low-activity (low-*BR*) to high-activity (high-*BR*) states. A first model could consider only a daily routine, while a more sophisticated model could include feedback from the estimated casualty rate in order to account for the growing stress of personnel during an attack.

5 Conclusions

The original problem set by DSTL was for the Study Group to propose a methodology for propagating uncertainty from the output of a dispersion model, through a casualty model, resulting in a probability distribution for casualty estimates.

The Study Group have suggested two approaches to uncertainty propagation, one based on large-scale sampling from the distribution of concentrations, and another based on the integral form of Bayes’ theorem. Both of these methods ought to work for the models in question, and some sample code has been written to illustrate the sampling approach.

Other discussions at the Study Group concerned the introduction of other sources of uncertainty, and the important distinction between parametric uncertainty and natural variation. Furthermore, it was suggested that the current calculation involving involving a probit curve may not be the most appropriate means of assessing the response of the population to the chemical dose. Finally, some speculation was made regarding the modelling of time-dependent breathing rates.

A Online resources

As part of the online support of the Study Group, a range of additional materials have been made available on the Study Group section of the `_connect` platform at

<https://connect.innovateuk.org/web/partnership-programmes/articles/-/blogs/propagation-of-uncertainty-through-a-hazard-chain>.

Available resources include:

- Problem brief
- Problem presentation (slides and video, without sound by request)
- Final presentation (slides and video)

Bibliography

- [1] BROOK, D. R., FELTON, N. V., CLEM, C. M., STRICKLAND, D. C. H., GRIFFITHS, I. H., AND KINGDON, R. D. (2003) Validation of the Urban Dispersion Model (UDM), *Int. J. Environment and Pollution*, **20**, pp. 11-21.
- [2] LEWELLYN, W. S., SYKES, R. I. (1986) Analysis of Concentration Fluctuations from Lidar Observations of Atmospheric Plumes, *J. Climate Appl. Meteor.*, **25**, pp. 1145-1154.
- [3] JONES, D. R. (2001) A Taxonomy of Global Optimisation Methods based on Response Surfaces, *J. Global Optim.*, **21**, pp. 345-383.
- [4] See <http://www.mucm.ac.uk>.
- [5] TURNER, M. AND LANNING, O. (2012) *Chemical and Biological Hazard Model Technical Description*, DSTL/TR63611 v1.0.
- [6] VAN TREES, H. L. Detection Estimation and Modulation Theory, Part 1, Chapter 3 (Wiley, New York, 1968).
- [7] BUNGARTZ, H.-J., GRIEBEL, M. (2004) Sparse Grids, *Acta Numerica*, **13**, pp. 147-269.
- [8] See <http://sparse-grids.de>