H13-104
THE IMPORTANCE OF CONCENTRATION FLUCTUATIONS IN HAZARD ASSESSMENT AND SOURCE TERM ESTIMATION

Steven Herring and Peter Robins

Dstl Porton Down, Salisbury, England

Abstract: A recent study has investigated how results from the Dstl MCBDF code are affected by different assumptions regarding the form of the concentration probability density function (pdf) defined by the mean and variance outputs from the dispersion model. A total of five test cases were selected for which the results from executing MCBDF with different pdf assumptions could be compared to high quality trials data. The results showed that using the current variance calculation, the concentration mean and variance values output from the dispersion model should be taken to refer to a standard Gaussian distribution, rather than a clipped-Gaussian distribution. The work also showed that whilst a partial solution may be achieved in the inverse modelling process with a relatively simple concentration variance model, a full solution requires a more complex model that can provide variances appropriate to different types of environment.

***Key words:** Inverse modelling, source-term estimation, Monte-Carlo Bayesian Data Fusion.*

INTRODUCTION
One of the principal problems to be overcome in developing an effective response to a release of hazardous material is to identify where the release occurred, and how much material escaped. Unfortunately, especially during the initial phase of a release (e.g. ~10 minutes, but potentially much longer), the true source characteristics will be unknown, and very little data is likely to be available on which to base hazard and warning areas. This leads to a requirement for some method of inverse modelling. However, inverse modelling (or source-term estimation) is difficult; firstly, because the process is only operationally useful if it produces answers within a few minutes; and secondly, because the uncertainties in the data are generally large. To overcome these problems research at Dstl has focused on a technique based on dynamic Bayesian graphical modelling. This technique is implemented in a prototype software module called the Monte-Carlo Bayesian Data Fusion (MCBDF) code.

The Bayesian approach is advantageous in that it enables disparate data to be combined in a mathematically tractable way and confers a high level of error tolerance. Whilst the core of MCBDF is a Bayesian inference process that generates hypotheses, evaluates them and constructs pdfs for the source-term, it also requires a dispersion model to produce concentration predictions at the sensor locations, against which the hypotheses can be evaluated. This means that the accuracy of the source-term estimation output from MCBDF depends to a large degree upon how well the concentration fluctuations recorded by the sensors can be correlated with those represented by the concentration probability density function (pdf) defined by the dispersion model.

CONFIGURATION OF MCBDF
The dispersion model used in MCBDF is the Dstl Urban Dispersion Model (UDM), which is a Gaussian puff model. UDM was developed, as its name suggests, for predicting the dispersion of hazardous material in urban environments. At present, MCBDF has only been tested against cases where the data is from sensors in flat open terrain. In this situation, UDM predicts the dispersion based on surface roughness values and has a very fast run time that enables MCBDF to operate in near real-time on a standard desk-top PC.

The usual inputs to MCBDF consist of:
1. A defined spatial area within which to carry out the inference process.
2. The concentration signals output from a sensor array.
3. Meteorological data in the form of a wind profile.

Given the above inputs, MCBDF is generally run to solve the source-term estimation problem for nine parameters[1]. These are: the source location (x and y), the release time (t), the release mass (m), the wind vectors (u and v), roughness length (r), and Monin-Obukhov length (L). Whilst it is possible to let MCBDF infer the wind speed and direction without providing any initial information, it is more efficient to provide a wind profile, and variances that enable it to hypothesise limited variations. It is important to note that whilst the meteorology might vary over the release area, the meteorological input to UDM consists of a single wind vector at a specified reference height. This reference wind is then taken to apply to the entire domain, and to be invariant with time.

CALCULATION OF CONCENTRATION VARIANCE
The efficiency and accuracy of the source-term estimation process in MCBDF depends upon correlating the concentration mean and variance value predictions output by UDM with the sensor data. This is achieved through a concentration pdf of an assumed form. The method of determining the variance values has a particular impact on the accuracy of the pdf because of the large range of fluctuations and intermittency that may be present in real data.

At present, UDM implements a relatively simple calculation for concentration variance. This defines the overall concentration variance value, c_{var}, due to a number of over-lapping puffs at a point as (Ratcliffe *et al.* 2009) :

[1] Values for more parameters can be inferred, such as meteorological conditions at a number of points if data from multiple meteorological sensors are input, but the computational time rises rapidly.

$$c_{\text{var}} = \frac{\bar{r}^2 \bar{c}^2}{\bar{G}} \tag{1}$$

In equation (1) $\bar{r}$ is the average fluctuation intensity of the puffs; $\bar{c}$ is the average concentration of the puffs; and $\bar{G}$ the average Gaussian factor of the puffs. The fluctuation intensity, *r*, of a puff is determined from:

$$r = \sqrt{\frac{\sigma_{ex}\sigma_{ey}\sigma_{ez}(1+K^2)}{\sigma_{ix}\sigma_{iy}\sigma_{iz}} - 1} \tag{2}$$

In equation (2) *K* is the internal fluctuation constant, which has a fixed value: *K*= 0.3. The subscripts 'e' and 'i' refer to ensemble and instantaneous puffs, and σ_x , σ_y and σ_z are the puff spreads in the x, y and z directions respectively. The relationships used to derive the instantaneous spread values (detailed in Ratcliffe *et al.* 2009) ensure that they are always a fraction of the ensemble values. This means that generally r >1; and the fluctuations are generally greater than the mean concentration and a wide range of variance values are produced.

It is recognised that the simple approach defined by equations (1) and (2) does not capture the full physics of the processes that determine the concentration variances. They also reflect a bias towards representing concentration fluctuations in the urban situation, rather than the open terrain. This is important, as the magnitudes of concentration fluctuations in a plume passing through obstacles are substantially less than those in a plume dispersing across open terrain (e.g. Davidson *et al*. (1995)).

TESTING OF MCBDF WITH TRIALS DATA

The first set of trial data against which the effectiveness of MCBDF was assessed was taken from Dipole Pride 26 (DP26). This revealed that the concentration series against which MCBDF was making its inference had little in common with the measured concentration time series. This is illustrated in Figures 1 and 2.

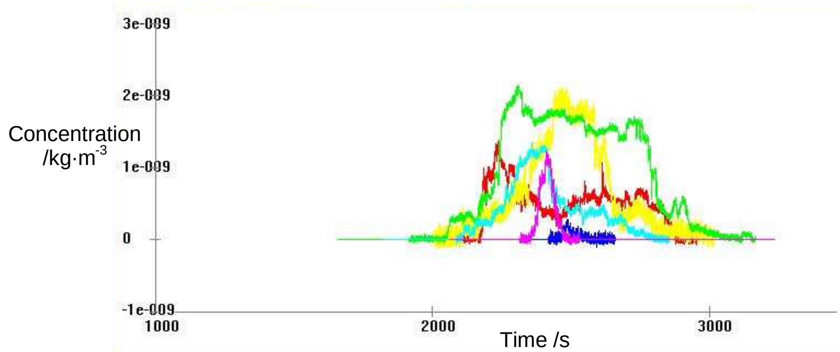


Figure 1. Dipole Pride 26 Case 12b: high frequency concentration data from TGA-4000 devices.

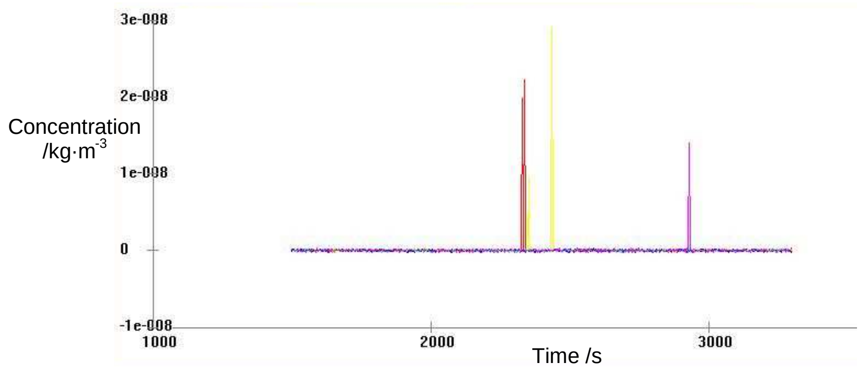


Figure 2. Dipole Pride 26 Case 12b: simulated data 4 m·s^{-1}, ensemble, clipped-Gaussian.

The concentration time series values generated within MCBDF that are shown in Figure 2 were derived by assuming that the concentration mean and variance values output from UDM referred to a clipped-Gaussian distribution. This assumption was consistent with the results of trials data acquired by Dstl and other organisations (including that in Davidson *et al.* 1995), and with the approach implemented in the Second-order Closure Integrated puff model (SCIPUFF, Sykes *et al.* 2008). However, it did not result in a data series from which MCBDF could make a sensible inference. Further investigation showed that if the mean and variance values were assumed to apply to a standard Gaussian distribution, then MCBDF was provided with a concentration time series that was qualitatively much more realistic (illustrated in Figure 3), and led to reasonable results.

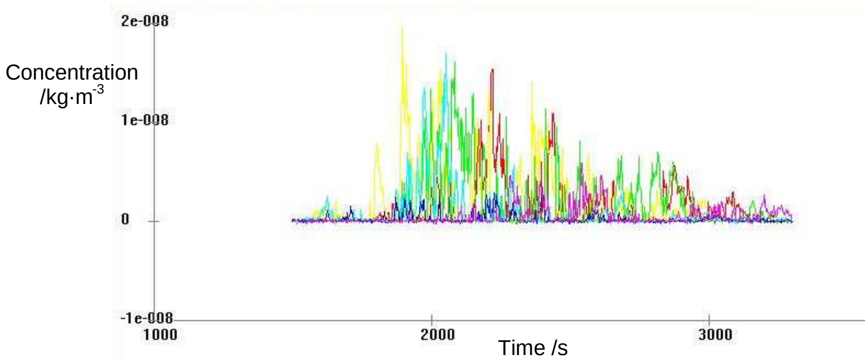


Figure 3. Dipole Pride 26 Case 12b: simulated data 4 m·s^{-1}, ensemble, Gaussian.

In 2008 MCBDF was applied, along with a number of other inverse methods, in a 'blind' test against a series of 35 test cases derived from the FUSION Field Trial 2007 (FFT07) dataset by the Institute for Defense Analyses. Based on previous experience with the DP26 data, the mean and variance values output from UDM were assumed to refer to an unclipped-Gaussian distribution. In addition, for the purpose of the FFT07 exercise, MCBDF was configured to output a most likely hypothesis for the release. This hypothesis was selected by considering the likelihoods of hypotheses for all nine parameters; and not simply the most likely location, most likely release mass, etc.

Overall, the ‘blind test’ exercise showed that the results from MCBDF were generally good in terms of release location and time. However, it was also evident that it systematically under-estimated the release-mass. This is illustrated by the actual and estimated release masses for the four cases shown in Table 1. The errors in the release masses indicated the need for a greater understanding of the appropriateness of the variance calculation in UDM, and how the concentration mean and variance values output from UDM should be used by MCBDF.

Table 1. Actual release masses, and those estimated by MCBDF for FFT07 test cases.

Case	Actual release mass	MCBDF release mass
No.	(kg)	(kg)
16	0.698	0.185
22	1.159	0.294
61	1.159	0.292
70	0.698	0.231

MEAN AND VARIANCE INVESTIGATION

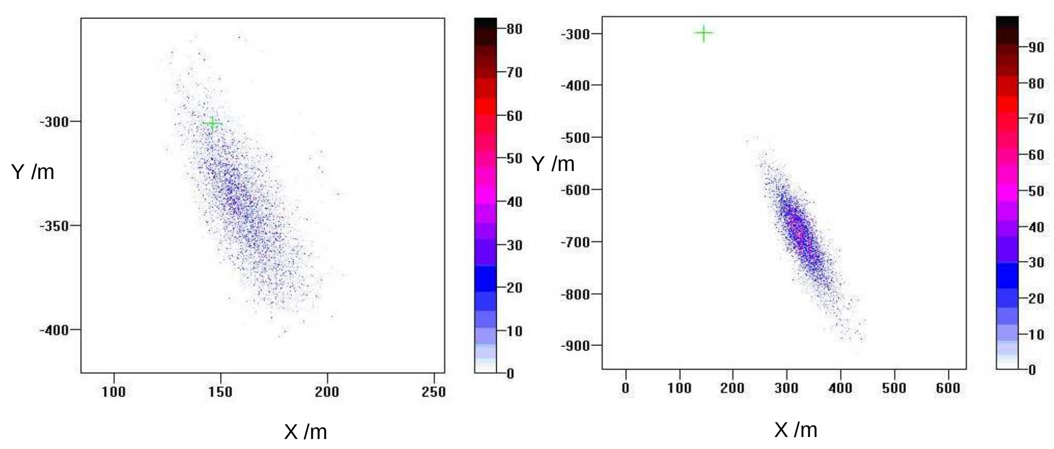
To address the need identified above, it was decided to carry out a comparative investigation. For each case MCBDF was first run assuming that the mean and variance values output from UDM referred to an unclipped-Gaussian distribution; the exercise was then repeated assuming that the mean and variance values output from UDM referred to a clipped-Gaussian distribution.

A set of five test cases were selected for the assessment. These consisted of: DP26 case 12b, and FFT07 cases 16, 22, 61, and 70. The FFT07 test cases selected were those from which the best results had been obtained in the FFT07 comparative exercise. Once the results were obtained, a comparison of the results for the FFT07 cases, when the unclipped-Gaussian assumption was made, showed that:

- The true source location was always within the location pdf.
- The most likely hypothesised release mass was typically between 20% and 40% of the true value, and the true value was generally outside the mass pdf.
- The release time pdf consistently indicated a later release time than was actually the case.

The results from DP26 Case 12b showed similar characteristics; except that the predicted release time was early.

When the results of the four FFT07 cases processed with the clipped Gaussian assumption were examined, the main observation was that the true source location was only within the location pdf for case 70. In all the other cases the location pdf was centred so far upwind of the true location that it did not contain it. This is illustrated in Figure 4, which shows the locations pdfs for Case 22. The true release location is indicated by the green cross, and the significance of individual results by the colour scale. The error in the release locations pdf that resulted from the clipped Gaussian assumption, led to earlier release time, and larger release mass predictions. The result for DP26 case 12b was very poor in all respects; with none of the true release parameters being within the solution pdfs.



(a) (b)

Figure 4. Comparison of locations pdfs for FFT07 case 22 for (a) unclipped-Gaussian and (b) clipped -Gaussian assumptions.

DISCUSSION

There were a number of significant differences in DP26 and FFT07 cases that could have affected the performance of MCBDF. These were:

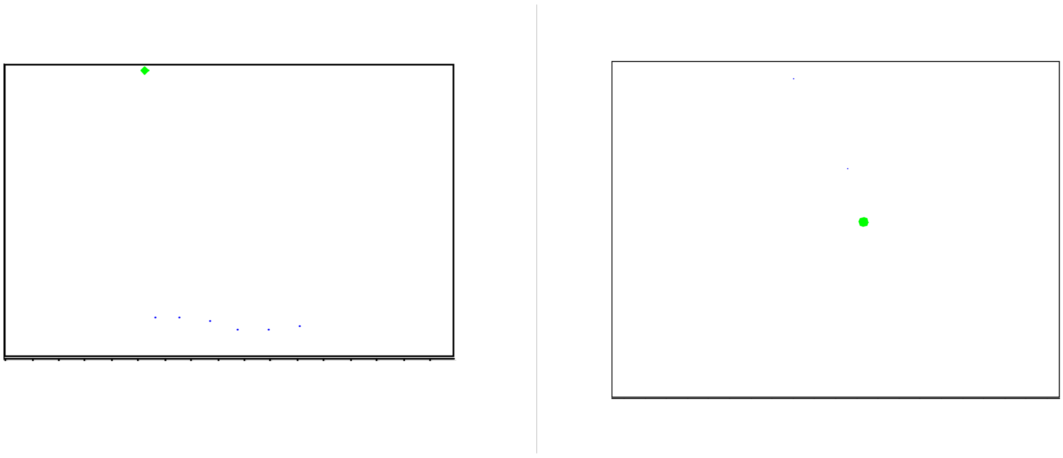
- The distances over which the dispersion concentrations were recorded were quite different. While the DP26 releases took place over more than 10 km, the FFT07 releases took place over about 700 m.
- The DP26 trials involved releases of SF_6, whilst those in FFT07 used polypropylene.
- The high frequency concentration data was recorded on TG-4000 instruments in DP26, while digiPIDs were used in FFT07.

All the test cases presented challenges to MCBDF. The DP26 case was challenging because of the distance and time over which the dispersion took place (around 10 km and half an hour), and the variation in meteorology that existed over the test area (as described by Biltcroft (1998)). However, the six TG-4000 sensors were favourably placed in a line at more than 70 degrees to the nominal downwind direction. Conversely, in FFT07 the distances and times over which the measurements were taken were limited to a few hundred metres and a minute of two; but the test cases only had data from two or three sensors with little crosswind separation. The geospatial differences are illustrated in Figure 5.

In addition to the inherent differences in the trials data, differences could have existed in how the data was processed in MCBDF. The sensor models used in MCBDF for processing the DP26 and FFT07 data were both simple Gaussian spread concentration models. Although there was some uncertainty regarding the values in the sensor models, it is not believed these had any substantial effect on the performance of MCBDF, as any errors were dominated by the concentration variance assumptions.

The better results obtained from the unclipped-Gaussian cases were not attributed to the fact that it provided a better model of the variances; but rather because it provided more sensor data and more vague hypotheses. This meant that significantly less weight was attached to each piece of data, which helped MCBDF construct sensible pdfs. The availability of a significant amount of vague data at each time-step helps MCBDF, because it does not currently take account of the past history of outputs from each sensor due to processing power limitations. This means that each likelihood is taken to be independent of the previous likelihood.

Two possible reasons for the consistent under-estimation of release-mass were firstly that the unclipped Gaussian assumption would be expected to result in an effective loss of mass; and secondly, that the predicted variances were too large. The second of these was investigated simply factoring the variances. No consistent benefit was observed, however, which suggested that solution to the variance calculation needed to be revised to improve the overall quality of the solution.



(a) (b)

Figure 5. Sensor positions (blue) and release locations (green) for DP26 Case 12b (a), and FFT07 case 22 (b).

For MCBDF to function well, its dispersion model must provide concentration fluctuations that are broadly comparable to those that are recorded by real sensors, reflecting the large uncertainties, and spatial and temporal variations, that can be expected in the input data. The results of the study have shown that the current UDM variance calculation, coupled with the unclipped-Gaussian assumption enables MCBDF to provide a partial solution in non-urban environments, but will lead to an under-estimate of the release mass. The study has shown that to significantly improve the results produced by MCBDF a more complex fluctuations model is required than the one currently used. Moreover, the variance model should reproduce the fluctuation distribution appropriate to the area: open terrain, or the urban environment, for example.

CONCLUSIONS

The study has led to the following conclusions:

- When MCBDF is used with the current version of UDM, a partial solution can be obtained if the concentration mean and variance values output from UDM are treated as referring to an unclipped-Gaussian distribution. The output pdfs will capture the true source location, and release time (within a reasonable interval), but not the release mass.
- More accurate outputs from MCBDF will require a variance model in UDM that provides more realistic spatial and temporal variances.
- Inverse modelling requires variance calculations appropriate to the environment, open terrain and urban, for example.

REFERENCES

Biltcroft, C.A., 1998: Dipole Pride 26: Phase II of Defense Special Weapons Agency Transport and Dispersion Model Validation. Dugway Proving Ground Meteorology and Obscurants Division Customer Report DPG-FR-98-001.

Davidson, M.J., K.R. Mylne, C.D. Jones, J.C. Phillips, R.J. Perkins, J.C.H. Fung, and J.C.R. Hunt, 1995: Plume Dispersion Through Large Groups of Obstacles: - A Field Investigation.

Ratcliffe, K., J. England, N. Hood, 2009: The Urban Dispersion Model (UDM) v2.18 Software Technical Documentation. Dstl/TR32228 v1.1.

Sykes, R.I., S.F. Parker, D.S. Henn, B. Chowdury, 2008: SCIPUFF Version 2.4 Technical Documentation. L-3 TMSD, Contract No. DTRA01-03-D-0013.