International Congress on Environmental Modelling and Software

5th International Congress on Environmental Modelling and Software - Ottawa, Ontario, Canada - July 2010

Jul 1st, 12:00 AM

The Spatial Variation of the Maximum Possible Pollutant Concentration from Steady Sources

N. Mole

R. J. Munro

The Spatial Variation of the Maximum Possible Pollutant Concentration from Steady Sources

N Mole [a], **R J Munro**[b]

[a]*School of Mathematics and Statistics, University of Sheffield, Hicks Building, Sheffield S3 7RH, U.K. (N.Mole@sheffield.ac.uk)*

[b]*Environmental Fluid Mechanics Group, Process and Environmental Research Division, Faculty of Engineering, University of Nottingham, Nottingham NG7 2RD, UK (Rick.Munro@nottingham.ac.uk)*

Abstract: When considering the possible hazard or nuisance associated with a release of toxic or malodorous gas into the atmosphere, large concentrations are especially important, but relatively little work has been done on measuring or modelling the distribution of large concentrations of a contaminant dispersing in a turbulent flow. We have previously applied statistical extreme value theory to field experiment measurements, analysing large concentrations at a handful of positions. We have also been involved in using a related moment-based method to give more comprehensive spatial coverage for a steady source in a wind tunnel. In the latter case, however, we did not calculate confidence intervals for the estimates. In the present paper we analyse measurements from steady source wind tunnel measurements, with particular emphasis on the spatial variation of the estimated maximum possible concentration. We use bootstrapping to obtain confidence intervals for the moment-based method, and use an improved version of this method. We show that this method agrees well with the results from maximum likelihood fitting to exceedances of a high threshold. We find that the centreline maximum concentration, normalised by the centreline mean concentration, increases downwind from a value just greater than 1 near the source, to a peak value of about 5, before decreasing again. Across the plume the maximum concentration only varies slowly. These observations are explained in terms of the physical processes of turbulent advection and molecular diffusion.

Keywords: Atmospheric dispersion; maximum concentration; turbulence; molecular diffusion; generalised Pareto distribution.

1 INTRODUCTION

For releases of toxic or malodorous gases into the atmosphere, large concentrations are of particular importance for the assessment of hazards or nuisance. Atmospheric boundary layer flows are turbulent, so pollutant concentrations are random, and need to be analysed in terms of their probability distributions.

In this paper we are particularly concerned with the probability distribution of large concentrations, for which the most appropriate framework is provided by statistical extreme value theory. Let $P_c(\theta; \theta_T)$ be the distribution function of the concentration Γ, conditional on Γ being above a threshold θ_T, i.e.

$$P_c(\theta; \theta_T) = \text{Prob}(\Gamma \leqslant \theta | \Gamma > \theta_T).$$

The corresponding probability density function (pdf) is

$$p_c(\theta; \theta_T) = \frac{d}{d\theta} P_c(\theta; \theta_T).$$

Subject to some regularity conditions, Pickands [1975] showed that for large threshold θ_T

$$p_c(\theta; \theta_T) \approx g(\theta - \theta_T; k, a) \qquad \text{for } \theta > \theta_T,$$

where $g(s; k, a)$ is the pdf of the generalised Pareto distribution (GPD), given by

$$g(s; k, a) = \frac{1}{a} \left(1 - \frac{ks}{a} \right)^{\frac{1}{k} - 1}. \tag{1}$$

Here k is the shape parameter and a (> 0) is the scale parameter.

This result has been applied to the statistical modelling of extreme values by fitting to exceedances of a high threshold, usually fitting by maximum likelihood (ML). Examples of this approach are given by Davison and Smith [1990]; Leadbetter [1991] and, for the present application, Mole et al. [1995]; Anderson et al. [1997]; Schopflocher [2001] and Munro et al. [2001]. For pollutant concentration the maximum possible value, θ_{MAX}, is finite (bounded above by the largest source concentration). Thus we expect that $k > 0$, so that $g(\theta; k, a)$ has a finite upper endpoint a/k. We then have

$$\theta_{\mathrm{MAX}} = \theta_T + \frac{a}{k}.$$

It is straightforward to show from (1) that increasing the threshold θ_T decreases the scale parameter a linearly, but does not change the form of the asymptotic distribution, nor the value of the shape parameter k.

Motivated by the desire to use a physical model for the concentration moments to estimate properties of the large concentrations, Mole et al. [2008] introduced an alternative method. The central argument there was that high order absolute moments m_n are dominated by the largest concentration values, so that

$$m_n = \mathrm{E}\left\{\Gamma^n\right\} \approx \eta \int_0^{\theta_{\mathrm{MAX}}} \theta^n g(\theta; k, a)\, d\theta$$

for some constant η. (Without loss of generality θ_T was taken to be zero – otherwise a can simply be rescaled.) Mole et al. [2008] showed that this implied that, for sufficiently large n,

$$\frac{m_{n-1}}{m_n} \approx \frac{1}{a}\left(\frac{1}{n}\right) + \frac{k}{a} = \frac{1}{a}\left(\frac{1}{n}\right) + \frac{1}{\theta_{\mathrm{MAX}}}. \tag{2}$$

This allows the parameters k, a and θ_{MAX} to be estimated from the linear relationship between the ratio of successive moments, m_{n-1}/m_n, and $1/n$. Mole et al. [2008] used this method to estimate the parameters from a theoretical model for the concentration moments, and from the experimental data of Sawford and Tivendale [1992] from a steady line source release in a wind tunnel.

In the present paper we extend the work of Mole et al. [2008] by applying this moment-based method to the wind tunnel data in a more thorough manner, including the estimation of confidence intervals for the parameter estimates. We also compare these results with those obtained by the more conventional method of ML fitting to exceedances of a high threshold.

2 THE ESTIMATION METHODS

2.1 The Moment-Based Method

Mole et al. [2008] fitted (2) by least squares, using the values $n = 4$ to $n = 8$. The straight line fit was quite good, but there was a suggestion of curvature which would have altered the estimates if larger values of n had been used. They used the data of Sawford and Tivendale [1992], from a steady line source release in wind tunnel grid turbulence.
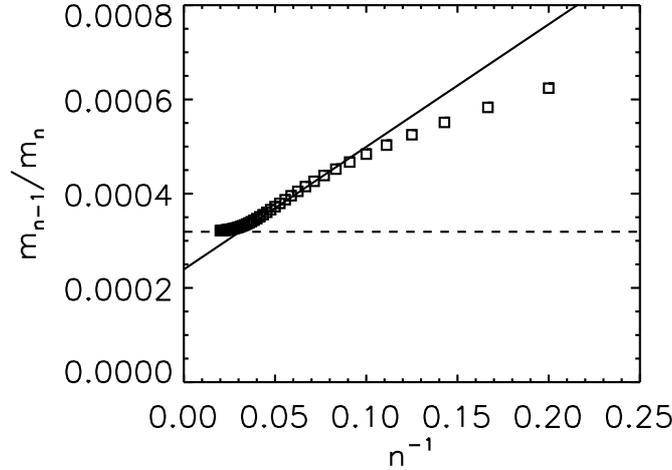
Figure 1. The ratio of successive moments, m_{n-1}/m_n, against $1/n$. The squares are the values calculated from the data, and the solid line is the fit of (2) using the maximum gradient. The dashed line is $1/\Gamma_{\text{MAX}}$, where Γ_{MAX} is the largest measured concentration. This case is on the mean plume centreline, at a distance of $100\,\text{mm}$ downwind of the source.

For a finite size dataset, for large n the moments m_n become dominated by the largest measured value Γ_{MAX}. Thus, using very large n to fit (2) will result in θ_{MAX} being underestimated (by Γ_{MAX}). Conversely, at small n (2) would not be expected to hold. So in fitting (2) to the data a balance needs to be struck between using small and large values of n. We have carried out some preliminary analysis, on simulated datasets, which suggests that a reasonable choice is to fit the straight line through the point of maximum gradient when m_{n-1}/m_n is plotted against $1/n$. For typical cases this will give the largest estimate for θ_{MAX} that is possible from (2), and for hazard assessment purposes will therefore tend to be on the conservative side. Figure 1 shows a typical plot of m_{n-1}/m_n against $1/n$, with the line fitted using the maximum gradient.

We estimate confidence intervals for the parameters k and θ_{MAX} by bootstrapping. We use the studentized bootstrap method with a log transformation of the data [for details of the method see e.g. Davison and Hinkley, 1997].

2.2 The Maximum Likelihood Method

Here we fit the GPD to excesses above a high threshold, using maximum likelihood. The thresholds are chosen using mean excess plots (i.e. mean residual life plots). Confidence intervals are estimated using profile likelihood [Davison and Smith, 1990]. To take account of possible dependence between concentration exceedances when estimating the confidence intervals, declustering is carried out. This involves choosing a cluster separation time s. If the time between successive exceedances is greater than s then they are deemed to belong to separate clusters, and only the maximum concentration value from each cluster is used in the ML fitting. We choose s using the method of Ferro and Segers [2003], which is based on estimating the extremal index. For details of the extremal index see Leadbetter et al. [1983, p. 67]; it can be interpreted as the reciprocal of the mean cluster size [Leadbetter, 1983].

3 RESULTS

We apply the estimation methods to the experimental data of Sawford and Tivendale [1992], which were also described in Sawford and Sullivan [1995] and Mole et al. [2008]. The experiments used a steady line source release in grid turbulence, in a wind tunnel. Concentration measurements were
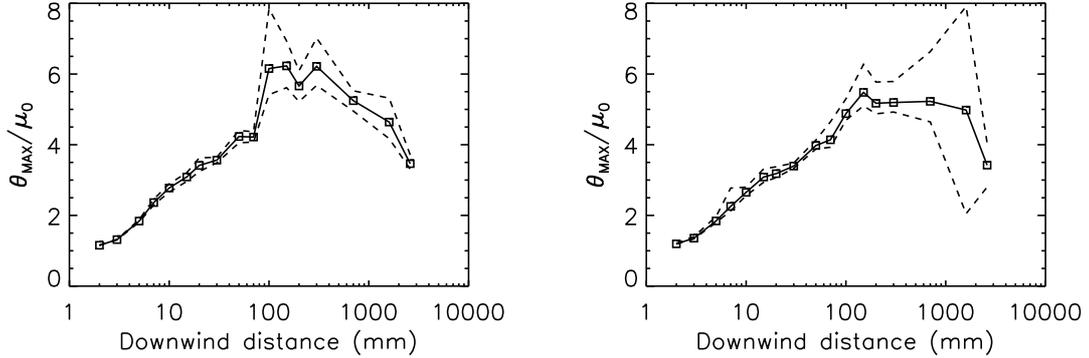
Figure 2. Left panel: the moment-based estimate of the maximum possible concentration θ_{MAX}, as a function of downwind distance from the source. θ_{MAX} is normalised by the centreline mean concentration μ_0. The squares and solid line are the estimates of $\theta_{\mathrm{MAX}}/\mu_0$, and the dashed lines give the 95% confidence intervals, estimated using bootstrapping.
Right panel: as left panel, but these are the maximum likelihood estimates. The confidence intervals are estimated using profile likelihood (except for the two furthest downwind positions, where the profile likelihood calculations would not converge properly, and Wald confidence intervals are given instead).

made at many downwind and crosswind positions. The sampling frequency was 4096 Hz, and at each position measurements were made for 20 separate, approximately independent, intervals, each of 1 second duration.

Figure 2 shows the moment-based estimates of θ_{MAX} on the mean plume centreline, as a function of downwind distance X from the source, and also the corresponding ML estimates. θ_{MAX} is normalised by the centreline mean concentration μ_0. Up to about 100 mm downwind the two methods agree very well. Beyond 100 mm the two methods differ more, but the confidence intervals are larger, and the differences are probably not very significant. $\theta_{\mathrm{MAX}}/\mu_0$ increases steadily from a value just over 1 at 2 mm downwind, to a maximum of about 5 at about 200–300 mm, and then decreases to about 3.5 at 2600 mm.

This pattern agrees qualitatively with the results of Mole et al. [2008], obtained using the moment ratios from $n = 4$ to $n = 8$, but the values here are roughly 20% larger. For the moment-based method this is to be expected, since here we are fitting to the maximum gradient, but the agreement with the ML estimates suggests that the maximum gradient method is superior to that used by Mole et al. [2008].

Mole et al. [2008] argued that the variation of $\theta_{\mathrm{MAX}}/\mu_0$ with downwind distance X shown in Figure 2 is what would be expected on physical grounds. For a steady source, close to the source we expect θ_{MAX} to tend to the largest source concentration θ_2, and μ_0 to tend to the mean source concentration θ_1. Thus, as $X \to 0$, we expect $\theta_{\mathrm{MAX}}/\mu_0 \to \theta_2/\theta_1 \geqslant 1$. For most sources θ_2/θ_1 is likely to be close to 1.

For a passive, conserved scalar, like the concentration in these experiments, the variation of $\theta_{\mathrm{MAX}}/\mu_0$ as one goes away from the source is determined by the balance between the effects of the two physical processes involved. These are advection by the turbulent velocity, and molecular diffusion (which is the only process which can change the concentration in a fluid element). The mean concentration μ is hardly affected by molecular diffusion, and so is controlled by turbulent advection. The maximum concentration θ_{MAX} can only be altered through the action of molecular diffusion.

Most atmospheric releases have large Péclet number $Pe = ul/\kappa$, where u and l are velocity and length scales for the turbulent fluctuations, and κ is the molecular diffusivity. If the length
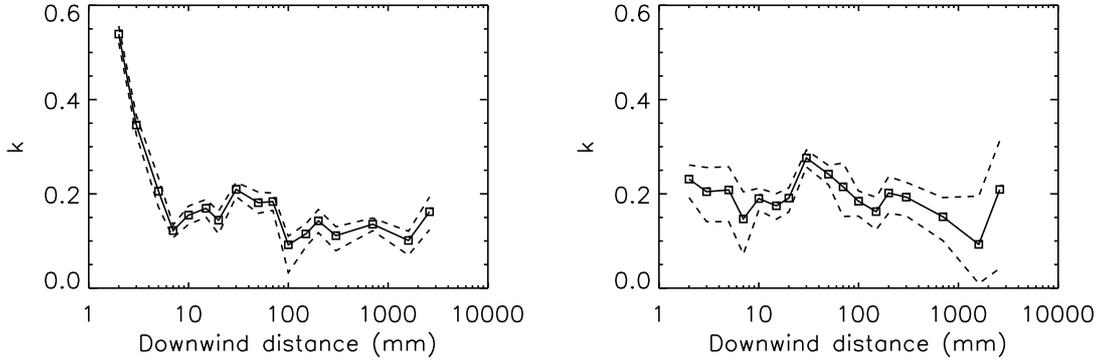
Figure 3. Left panel: the moment-based estimate of k, as a function of downwind distance from the source. The squares and solid line are the estimates of k, and the dashed lines give the 95% confidence intervals, estimated using bootstrapping.
Right panel: as left panel, but these are the maximum likelihood estimates. The confidence intervals are estimated using profile likelihood (except for 1600 mm, where the Wald confidence interval has been used).

scales for the velocity and concentration fields are comparable, then the ratio of the time scales for diffusion and advection is Pe, so advection acts much more quickly than diffusion. As the plume becomes stretched out into thin sheets and strands by turbulent advection, the concentration length scale becomes smaller and diffusion acts more quickly. The conduction cutoff length is defined as $\left(\nu\kappa^2/\epsilon\right)^{1/4}$, where ν is the kinematic viscosity and ϵ is the turbulent energy dissipation rate per unit mass, and the Schmidt number is defined as ν/κ. For pollutants with Schmidt number of order 1 or greater, which is usually the case, once the concentration length scale reaches the conduction cutoff length, then a balance is reached between advection and diffusion, where their time scales are the same and the concentration length scale does not decrease any more.

Thus, provided the source size is large compared with the conduction cutoff, near the source advection acts much faster than diffusion. This means that μ_0 is reduced much more quickly than θ_{MAX}, so $\theta_{\text{MAX}}/\mu_0$ increases away from the source. Further downwind the concentration length scale decreases, and since the largest concentrations are found in the thin sheets and strands, diffusion acts to reduce θ_{MAX} more quickly. Conversely, as the plume width increases, advection reduces μ_0 more slowly. Eventually the rate of reduction of μ_0 becomes less than that of θ_{MAX}, so $\theta_{\text{MAX}}/\mu_0$ reaches a peak and then decreases with downwind distance.

If there is an upper bound on the turbulent length scales, as is the case for wind tunnel grid turbulence, then eventually the plume width is much greater than the largest turbulent length scale, and Mole et al. [2008] argued that in this case $\theta_{\text{MAX}} \to \mu$ as $X \to \infty$, so on the centreline we expect $\theta_{\text{MAX}}/\mu_0 \to 1$. In practice, wind tunnels are not long and wide enough to approach this limit, and in field experiments one would expect non-stationarity and inhomogeneity to make this limit difficult to observe.

Figure 3 shows the moment-based and ML estimates of k. The moment-based estimates show a rapid decrease in values within 10 mm of the source. Beyond 10 mm there is no obvious pattern, with values generally between 0.1 and 0.2. The ML estimates have no obvious pattern throughout, with values between 0.1 and 0.3. The suggestion is that k does not vary much with downwind distance, with values of order 0.2 or a little less. This is different from the results of Mole et al. [2008], where k was roughly equal to the reciprocal of $\theta_{\text{MAX}}/\mu_0$, decreasing to about 0.2 at 200-300 mm downwind, before increasing again.

Close to the source we expect most of the weight of the pdf of concentration to be close to the source concentration values, and away from zero. For a uniform source with concentration θ_0, this pdf would be $\delta(\theta - \theta_0)$, and for a non-uniform source we expect the pdf to be slightly broadened
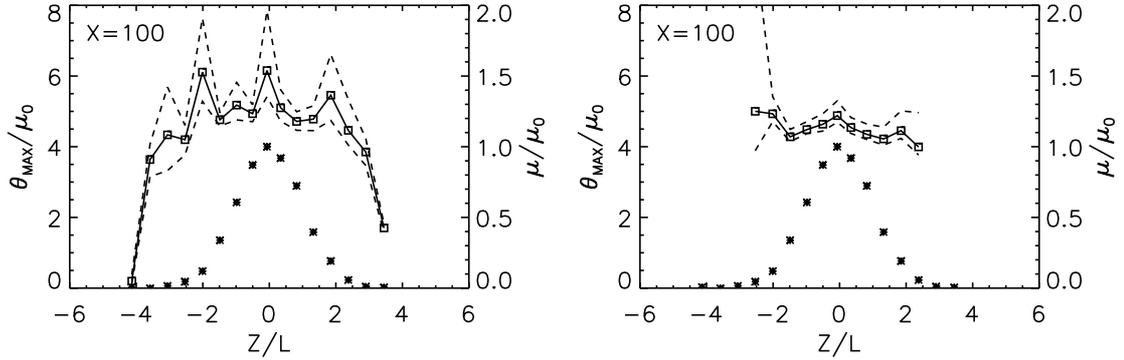
Figure 4. Left panel: the moment-based estimate of $\theta_{\text{MAX}}/\mu_0$, at downwind distance $100\,\text{mm}$, as a function of crosswind distance Z from the mean plume centreline, normalised by the mean plume width L. The squares and solid line are the estimates of $\theta_{\text{MAX}}/\mu_0$, and the dashed lines give the 95% confidence intervals, estimated using bootstrapping. The asterisks show the mean concentration μ, normalised by the centreline value μ_0.
Right panel: as left panel, but these are the maximum likelihood estimates. The confidence intervals are estimated using profile likelihood. Values of $\theta_{\text{MAX}}/\mu_0$ are not plotted for the points furthest from the centreline, since convergence was not obtained there.

from this. For (1) to give a peak away from $\theta = 0$ requires $k > 1$, so we would expect to have large values of k very close to the source. Similarly, very far downwind, Mole et al. [2008] argued that the concentration pdf tends to $\delta(\theta - \mu)$, so the same argument for k holds. So the expectation would be that k is large near the source, decreases to a minimum, and then increases again very far downwind. Figure 3 does not show any clear evidence for the latter, which may just be because we do not have measurements sufficiently far downwind. The left-hand panel of Figure 3 does show evidence for larger values of k near the source, but since this is not supported by the right-hand panel, it is not clear how reliable this is.

Figure 4 shows the moment-based and ML estimates of $\theta_{\text{MAX}}/\mu_0$, as a function of crosswind distance Z, at downwind distance $X = 100\,\text{mm}$. Z is normalised by the mean plume width L, which is defined as the standard deviation of the crosswind profile of mean concentration. Also shown on the figure are the values of μ/μ_0, whose crosswind variation is close to Gaussian. At the positions furthest from the centreline, convergence was not obtained for ML, so no points are shown for those positions in the right panel.

For $|Z/L|$ less than about 3, $\theta_{\text{MAX}}/\mu_0$ appears to be fairly constant. Further from the centreline the moment-based method gives a rapid decrease in $\theta_{\text{MAX}}/\mu_0$. However, this far from the centreline larger concentrations occur only very rarely. To obtain accurate estimates of the distribution of larger concentrations much longer time series will be required than for positions close to the centreline. This is reflected in the lack of ML convergence. So we believe that this apparent rapid decrease in $\theta_{\text{MAX}}/\mu_0$ does not reflect the physical reality.

The crosswind variation of μ/μ_0, and the value of L, is determined by turbulent advection. Very far from the centreline the positions are much further from the source than are those positions near the centreline, so diffusion has had longer to act. Thus, we expect $\theta_{\text{MAX}} \rightarrow 0$ as $|Z/L| \rightarrow \infty$. At small values of X, where advection reduces μ_0 more quickly than diffusion reduces θ_{MAX}, we expect that the decrease of θ_{MAX} away from the plume centreline will be on a much larger length scale than L, so we expect $\theta_{\text{MAX}}/\mu_0$ to be fairly constant for $|Z/L|$ less than 2 or 3.

At values of X close to the downwind peak in $\theta_{\text{MAX}}/\mu_0$, we still expect this to be true because it depends on the cumulative effect while the fluid elements travel from the source to these downwind distances. So this argument is in agreement with the observations in Figure 4.
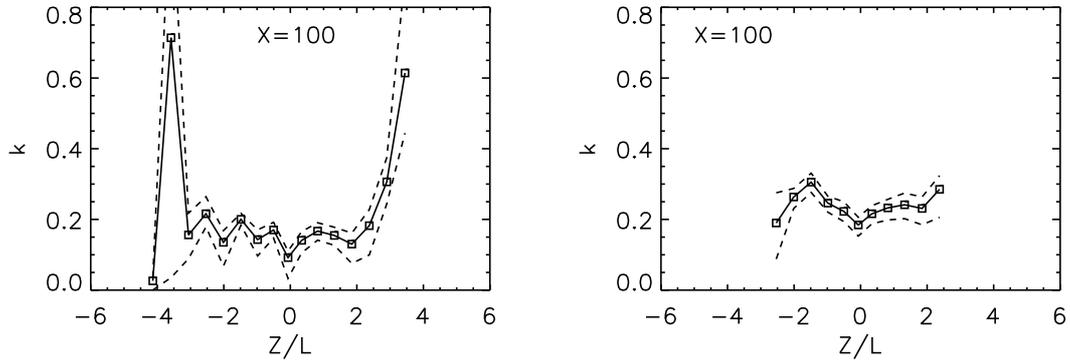
Figure 5. Left panel: the moment-based estimate of k, at downwind distance $100 \, \mathrm{mm}$, as a function of Z/L. The squares and solid line are the estimates of k, and the dashed lines give the 95% confidence intervals, estimated using bootstrapping.

Right panel: as left panel, but these are the maximum likelihood estimates. The confidence intervals are estimated using profile likelihood. Values are not plotted for the points furthest from the centreline, since convergence was not obtained there.

Very far downwind we expect that $\theta_{\mathrm{MAX}} \to \mu$, so $\theta_{\mathrm{MAX}}/\mu_0 \to \mu/\mu_0$, and $\theta_{\mathrm{MAX}}/\mu_0$ will decrease on the same length scale L as μ/μ_0. The points in Figure 7(a) of Mole et al. [2008] suggest that the wind tunnel measurements do not extend far enough downwind to identify this regime.

Figure 5 shows the crosswind variation of the moment-based and ML estimates of k. For $|Z/L|$ less than about 3 there is no clear pattern, with moment-based values between about 0.1 and 0.2, and the ML estimates a little larger. As for $\theta_{\mathrm{MAX}}/\mu_0$, we believe that the moment-based estimates at the positions furthest from the centreline can be discounted, and that the evidence is that k only varies slowly across the plume. This is what we would expect from the physical arguments given above for $\theta_{\mathrm{MAX}}/\mu_0$, since k is determined by the distribution of large concentrations, which are controlled by molecular diffusion.

4   DISCUSSION

We have used two methods for estimating θ_{MAX} and k. One is based on the expected behaviour of high order concentration moments, and the other fits (1) directly to concentrations above a high threshold, using maximum likelihood. The two methods agree reasonably well for all the cases shown, and are in very close agreement for θ_{MAX} on the mean plume centreline up to $100 \, \mathrm{mm}$ downwind of the source.

The results show that on the centreline $\theta_{\mathrm{MAX}}/\mu_0$ increases from a value slightly larger than 1 very near the source, to a value of about 5 at about 200–$300 \, \mathrm{mm}$ downwind, before decreasing again. In the crosswind direction, at $100 \, \mathrm{mm}$ downwind, both $\theta_{\mathrm{MAX}}/\mu_0$ and k vary much more slowly than μ/μ_0. These results are in agreement with the qualitative predictions based on arguments about the balance between the physical processes of turbulent advection and molecular diffusion.

These results provide encouragement to proceed with attempts to develop quantitative models for $\theta_{\mathrm{MAX}}/\mu_0$ and of other properties of the distribution of large concentrations. Such models would then enable hazard assessment to be carried out for practical applications involving releases of toxic gases in the atmosphere. Some discussion of possible modelling approaches was given in Mole et al. [2008].

Grid turbulence provides an approximation to homogeneous isotropic turbulence, for which it is relatively easy to interpret results and develop models. However, in practical applications releases occur in inhomogeneous boundary-layer turbulence, which, unlike grid turbulence, does not decay

with downwind distance. An alternative would be to use wind tunnel boundary layer releases like those analysed by Xie et al. [2007].

To avoid the limitation on downwind distances imposed by the dimensions of a wind tunnel, it would be desirable to repeat the analysis for field experiments. There are, however, some difficulties with attempting this. In addition to the usual problems with field experiments of non-stationary conditions and terrain effects, there is a specific difficulty relating to measuring large concentrations. Large concentrations are found at the smallest spatial scales present in the concentration field, so sensors with very good spatial resolution are needed to measure them. Such sensors will be ones which can only make measurements at a single point, so obtaining a wide spatial coverage is expensive. On the other hand, a measurement system like lidar which gives wide spatial coverage, cannot resolve sufficiently small scales to give reliable measurements of the largest concentrations.

REFERENCES

Anderson, C. W., N. Mole, and S. Nadarajah. A switching Poisson process model for high concentrations in short-range atmospheric dispersion. *Atmos. Environ.*, 31:813–824, 1997.

Davison, A. C. and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, 1997.

Davison, A. C. and R. C. Smith. Model for exceedances over high thresholds. *J. R. Statist. Soc. B*, 52:393–442, 1990.

Ferro, C. A. T. and J. Segers. Inference for clusters of extreme values. *J. R. Statist. Soc. B*, 65: 545–556, 2003.

Leadbetter, M. R. Extremes and local dependence in stationary sequences. *Z. Wahrsch. Ver. Geb.*, 65:291–306, 1983.

Leadbetter, M. R. On a basis for 'peaks over thresholds' modelling. *Statist. Probab. Lett.*, 12: 357–362, 1991.

Leadbetter, M. R., G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Series*. Springer, New York, 1983.

Mole, N., C. W. Anderson, S. Nadarajah, and C. Wright. A generalized Pareto distribution model for high concentrations in short-range atmospheric dispersion. *Environmetrics*, 6:595–606, 1995.

Mole, N., T. P. Schopflocher, and P. J. Sullivan. High concentrations of a passive scalar in turbulent dispersion. *J. Fluid Mech.*, 604:447–474, 2008.

Munro, R. J., P. C. Chatwin, and N. Mole. The high concentration tails of the probability density function of a dispersing scalar in the atmosphere. *Boundary-Layer Met.*, 98:315–339, 2001.

Pickands, J. Statistical inference using extreme order statistics. *Ann. Statist.*, 3:119–206, 1975.

Sawford, B. L. and P. J. Sullivan. A simple representation of a developing contaminant concentration field. *J. Fluid Mech.*, 289:141–157, 1995.

Sawford, B. L. and C. M. Tivendale. Measurements of concentration statistics downstream of a line source in grid turbulence. In *Proc. 11th Australasian Fluid Mechanics Conf., 14–18 Dec. 1992*, pages 945–948, University of Tasmania, Hobart, 1992.

Schopflocher, T. P. An examination of the right-tail of the PDF of a diffusing scalar in a turbulent flow. *Environmetrics*, 12:131–145, 2001.

Xie, Z.-T., P. Hayden, A. G. Robins, and P. R. Voke. Modelling extreme concentrations from a source in a turbulent flow over a rough wall. *Atmos. Environ.*, 41:3395–3406, 2007.