INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE
UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
INTERNATIONAL CENTRE FOR SCIENCE AND HIGH TECHNOLOGY
c/o INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS 34100 TRIESTE (ITALY) VIA GRIGNANO, 9 (ADRIATICO PALACE) P.O. BOX 586 TELEPHONE 040-224572 TELEFAX 040-224575 TELEX 460449 APH I

SMR/760-19

"College on Atmospheric Boundary Layer and Air Pollution Modelling"
16 May - 3 June 1994

"Modelling Concentration Fluctuations in Air Pollution"

P.C. CHATWIN
University of Sheffield
School of Mathematics and Statistics
Sheffield, United Kingdom

Please note: These notes are intended for internal distribution only.

13 Modelling Concentration Fluctuations in Air Pollution

N. MOLE, P.C. CHATWIN
Department of Applied and Computational Mathematics, University of Sheffield, Sheffield, United Kingdom

P.J. SULLIVAN
Department of Applied Mathematics, University of Western Ontario, London, Canada

The assessment of damage likely to be caused by pollutants in the atmosphere is a problem of great practical importance. This assessment is necessary in order to answer political and planning questions such as: whether a pollutant source should be allowed, and where it should be sited; what level of emissions should be allowed from a source; and what contingency measures would be needed to deal with an accidental release of a harmful pollutant.

Assessing likely damage involves three separate modelling components. Firstly, the characteristics of the source need to be known – these include things like the rate and duration of release, source size, initial buoyancy and momentum. In some cases, for example steady releases from power station stacks, these may be quite well defined. In other cases, for example ruptured storage tanks, modelling may be required to determine their likely values (or ideally their probability distributions).

The second component is the transport of the pollutant through the atmosphere. This will always involve advection by the wind, and molecular diffusion, but may well involve other processes too. These can include buoyancy effects for heavy or light gases, two-phase releases with thermodynamic effects, and non-conservative processes such as deposition, chemical reaction and radioactive decay. In general surface and atmospheric conditions need to be taken into account – surface roughness, complex terrain, low level atmospheric stability, depth of mixing layer etc.

The final component is the effect on people, or on some aspect of the environment, resulting from a given exposure to a pollutant. Typical hazards to be considered are those resulting from flammable, toxic and

Modelling Change in Environmental Systems. Edited by A.J. Jakeman, M.B. Beck and M.J. McAleer

radioactive pollutants. Chapter 12 discusses aspects of this human impact problem.

The main interest in this chapter is with modelling pollutant transport; it will be assumed that the source characteristics are well defined. The environmental impact of pollutant exposure has to be dealt with to the extent of determining which are the appropriate functions of pollutant concentration that need to be modelled. While the more complex aspects of transport listed above are important in practice, there is as yet no fully successful model even for very simple pollutants and atmospheric and surface conditions. Modelling more complicated situations is difficult to justify scientifically at present. For this reason we shall, unless otherwise stated, restrict attention to passive conserved pollutants, flat terrain, and neutral stability. We shall also consider only short ranges (up to at most a few kilometres) from the source – otherwise it is necessary to have as input the spatial mean wind field, or to model it with a numerical weather prediction type model. Large area sources, such as car exhaust gases from a city, will also not be considered directly. The related modelling problem of pollutant transport in rivers is covered in this volume in Chapter 4.

QUANTITATIVE DESCRIPTION

The atmospheric flows in which pollutants are transported are always turbulent, so the velocity $Y(x,t)$ at position x and time t, and the pollutant concentration $\Gamma(x,t)$, are random variables. Figure 1 shows a passive gas cloud dispersing in a wind tunnel experiment. The random fluctuations in concentration resulting from the turbulence can be seen clearly. Thus the quantities which have to be modelled will be appropriate probabilistic properties of the field $\Gamma(x,t)$. Which property is appropriate depends on the particular application. The practical importance of concentration fluctuations must be stressed – the mean concentration alone is insufficient for hazard analysis, as is clear from the cases considered below.

For flammable gases the main questions of interest are whether an ignition source at (x,t) will cause ignition of the gas, and what portion of the gas cloud will burn up once ignition has occurred at some point. The probability of ignition at (x,t) is given by $P(\theta_L < \Gamma(x,t) < \theta_U)$ where $P(\cdot)$ denotes the probability of $\cdot$ and θ_L and θ_U are the lower and upper flammable limits, respectively (e.g. for methane θ_L and θ_U are approximately 5% and 15% by volume). If $p(\theta;x,t) = \mathrm{dP}(\Gamma(x,t) < \theta)/\mathrm{d}\theta$ is the probability density function (pdf) of $\Gamma(x,t)$ then

Figure 1. Three views of a neutrally buoyant gas cloud in a wind tunnel (taken from Trial 27A of Hall *et al.*, 1982). (Crown Copyright. Reproduced by permission)

$$\text{probability of ignition at } (x,t) = \int_{\theta_L}^{\theta_U} p(\theta;x,t)\, d\theta.$$

This integral is then the quantity needing to be modelled. For flexibility, and because it is the natural quantity to model in order to gain scientific understanding, it is desirable to model $p(\theta;x,t)$ itself. The cloud burn-up problem would require the joint pdf $p(\theta_1,\ldots,\theta_n;x_1,\ldots,x_n;t_1,\ldots,t_n)$ for $\Gamma(x_1,t_1),\ldots,\Gamma(x_n,t_n)$ for large n (in theory, for all n). This is a much more difficult problem, and will not be considered in detail. (Not only is it a much more difficult modelling problem, but the difficulties with obtaining data for validation, discussed below, are also much more severe than for the single-point pdf.)

For toxic gases, toxicological experiments suggest that a measure of the level of damage resulting from exposure at a fixed point x during a time interval $[t-\frac{1}{2}T,t+\frac{1}{2}T]$ is the dose $D_n(x,t,T)$ where

$$D_n(x,t,T) = \int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T} \Gamma^n(x,s)\, ds \qquad (1)$$

(Griffiths, 1991). The exponent n depends on the particular toxic gas, and may be considerably greater than 1 (ten Berge *et al.*, 1986). Figure 2 shows an example of a time series of D_n, with $n=1$, for a steady release. The evidence for equation (*1*) is not very strong, being based mainly on exposure of small mammals such as mice and rats for differing *constant* concentrations Γ, for differing exposure times T. However, it is clear from studies such as ten Berge *et al.* (1986) that the product of exposure time and time mean concentration, i.e. $D_1(x,t,T)$, is not in general a good measure of the level of damage, and equation (*1*) is as good a measure as is currently available. So for toxic gases one would like to be able to model the pdf of the dose $D_n(x,t,T)$. This will depend on the joint pdf of the concentration for all the times in the interval $[t-\frac{1}{2}T,\ t+\frac{1}{2}T]$.

It should be noted that while the above quantities are those which are necessary to answer the kind of questions required by a proper planning process, they are usually not the ones used by regulatory bodies. This is partly because regulatory bodies have concentrated on quantities which are easier to model, like the mean concentration, but also because the necessary questions have often not been posed. In particular the importance of concentration fluctuations about the mean has been ignored

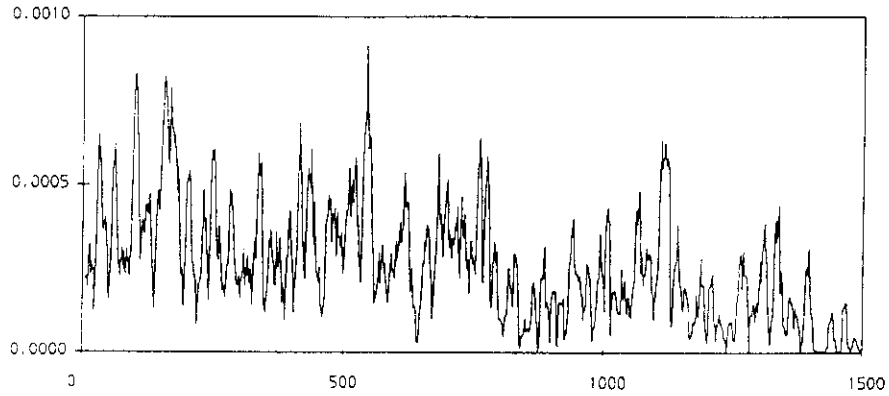

Figure 2. Time series of $T^{-1}D_n(t,T)$ for $n=1$, where the concentration Γ has been non-dimensionalised by the source concentration. A steady continuous release in the field (Mole and Jones, 1993), T = 10s. Horizontal axis is time in seconds

or undervalued, and probabilistic arguments have been avoided. Further discussion of these points can be found in Chatwin and Sullivan (1993).

The governing equations for the problem we are then considering are the Navier-Stokes equation for $Y(x,t)$, and

$$\frac{\partial \Gamma}{\partial t} + Y \cdot \nabla \Gamma = \kappa \nabla^2 \Gamma \qquad (2)$$

where κ is the molecular diffusivity. From equation (2) the following equation for $p(\theta;x,t)$ can be derived (Pope, 1985; Chatwin, 1990):

$$\frac{\partial p}{\partial t} + \nabla \cdot (pE\{Y|\Gamma = \theta\}) = \kappa \nabla^2 p - \kappa \frac{\partial^2}{\partial \theta^2} (pE\{(\nabla \Gamma)^2 | \Gamma = \theta\}) \qquad (3)$$

where $E\{\cdot \mid \cdot\}$ is the conditional expected value (or ensemble mean).

Similarly one can derive equations for the dose $D_n(x,t,T)$ and its pdf $p_{D_n}(\phi;x,t,T)$:

$$\frac{\partial D_n}{\partial t} + \mathrm{Y}\cdot\nabla D_n = \kappa\,\nabla^2 D_n - \kappa\, n(n-1)\int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T}\Gamma^{n-2}(x,s)[\nabla\Gamma(x,s)]^2\,ds$$

$$-\int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T}\{\mathrm{Y}(x,s)-\mathrm{Y}(x,t)\}\cdot\nabla[\Gamma^n(x,s)]\,ds \qquad n \geq 1 \tag{4}$$

$$\frac{\partial p_{D_n}}{\partial t} + \nabla\cdot\left(p_{D_n}E\{\mathrm{Y}\,|\,D_n=\phi\}\right) = \kappa\nabla^2 p_{D_n} - \kappa\frac{\partial^2}{\partial\phi^2}\left(p_{D_n}E\left\{(\nabla D_n)^2\,|\,D_n=\phi\right\}\right)$$

$$+\kappa n(n-1)\frac{\partial}{\partial\phi}\left[p_{D_n}E\left\{\int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T}\Gamma^{n-2}(\nabla\Gamma)^2 ds\,|\,D_n=\phi\right\}\right]$$

$$+\frac{\partial}{\partial\phi}\left[p_{D_n}E\left\{\int_{t-\frac{1}{2}T}^{t+\frac{1}{2}T}[\mathrm{Y}(x,s)-\mathrm{Y}(x,t)]\cdot\nabla(\Gamma^n)ds\,|\,D_n=\phi\right\}\right] \tag{5}$$

Equation (*4*) shows that while D_1 is conserved, i.e.

$$\int_{a.s.} D_1 dV = \text{constant}$$

where *a.s.* denotes an integral over all space, diffusion acts to destroy D_n for $n>1$. This is analogous to the moments of concentration of order greater than 1, which are also destroyed by diffusion (see a later section and Chatwin and Sullivan, 1990a).

In addition to modelling the relevant functions of concentration for fixed source and atmospheric conditions, it is also important to be able to assess the environmental impact of change in these conditions. For changes on timescales greater than dispersion timescales, this will be possible by modifying model inputs appropriately once models which can handle a range of source and atmospheric conditions have been developed and validated. For changes on short timescales, models which explicitly model changing conditions must be used. From the scientific point of view, this should only be attempted once constant conditions can be modelled with confidence.

DIFFERENT MODELLING APPROACHES

LOW-ORDER MOMENTS OF CONCENTRATION

In principle equation (2), together with the Navier-Stokes equation, could be solved directly (direct numerical simulation, or DNS). However, to yield accurate results the full range of physically important scales, from the Batchelor and Kolmogorov microscales to the large energy-generating scales, would have to be resolved. In the atmosphere this would imply scales from $O(10^{-4})$ to at least $O(10^4\text{m})$ in the horizontal and $O(10^3\text{m})$ in the vertical. Together with the associated temporal resolution required, this implies that such an exercise requires computing power orders of magnitude greater than that currently available. Even for high Reynolds number pipe flow, current computing power is far short of that required (see Reynolds, 1990; Speziale, 1991). The latter estimates that for pipe flow at Reynolds number 5×10^5, economically feasible DNS would require a computer 10^7 times faster than the Cray YMP.

Instead, a standard modelling approach has been to derive the Reynolds equations for the mean velocity $U(x,t) = E\{Y(x,t)\}$ and the mean concentration $C(x,t)=E\{\Gamma(x,t)\}$. Letting $Y = U + u$ and $\Gamma = C+c$, and taking the mean of equation (2) gives

$$\frac{\partial C}{\partial t} + U\cdot\nabla C - \kappa\nabla^2 C = -\nabla\cdot E\{uc\} \tag{6}$$

A similar equation for U can be derived from the Navier-Stokes equation. Equation (6) exhibits the well-known closure problem – the right-hand side involves an unknown term which is nonlinear in the fluctuating quantities u and c. By similar means an equation for this unknown term can be derived:

$$\frac{\partial}{\partial t}E\{u_i c\} + E\{u_i u_j\}\frac{\partial C}{\partial x_j} + U_j\frac{\partial}{\partial x_j}E\{u_i c\} + E\{u_j c\}\frac{\partial U_i}{\partial x_j} = -\frac{\partial}{\partial x_j}E\{u_i u_j c\} - \frac{1}{\rho}E\{c\frac{\partial p}{\partial x_i}\} + \kappa E\{u_i\nabla^2 c\} + \nu E\{c\nabla^2 u_i\} \tag{7}$$

where p is the pressure fluctuation, ρ is the density, and ν is the kinematic viscosity. A similar equation can be obtained for $E\{u_i u_j c\}$. These equations contain further unknown fluctuation products such as $E\{u_i u_j c\}$ and $E\{c\partial p/\partial x_i\}$.

With presently available computer power, the dividing line between small and large scales is usually of O(10-100 m).

The most common approach is to use a simple eddy viscosity/diffusion closure for these terms, although sometimes a more sophisticated version is used with an extra energy equation to give the eddy viscosity/diffusivity. Reviews of LES applied to turbulence are given by Rogallo and Moin (1984), and Reynolds (1990). Examples of applications of LES to dispersion are provided by Moeng (1984), Nieuwstadt and de Valk (1987) and Glendening and Burk (1992). LES seems more promising for atmospheric applications than second-order closures, but still suffers from some of the same problems with closure assumptions and boundary conditions. The latter have recently been addressed by Mason and Thomson (1992). Chapter 17 in this volume deals with the application of LES to ocean modelling.

Random walk models

Another popular type of model is that of random walks. This involves modelling dispersion through the random walks of discrete particles. Let $p_1(x,t|y,s)\mathrm{d}x\mathrm{d}t$ be the probability that a particle trajectory through (y,s) passes through a volume $\mathrm{d}x\mathrm{d}t$ surrounding (x,t), and let $p_2(x_1,x_2,t_1,t_2|y_1,y_2,s_1,s_2)\mathrm{d}x_1\mathrm{d}t_1\mathrm{d}x_2\mathrm{d}t_2$ be the analogous probability for two particle trajectories. Then

$$E\{\Gamma(x,t)\} = \int_{s<t} p_1(x,t|y,s)S(y,s)\mathrm{d}y\mathrm{d}s$$

$$E\{\Gamma(x_1,t_1)\Gamma(x_2,t_2)\}$$

$$= \int_{s_1<t_1,s_2<t_2} p_2(x_1,x_2,t_1,t_2|y_1,y_2,s_1,s_2)S(y_1,s_1)S(y_2,s_2)\mathrm{d}y_1\mathrm{d}s_1\mathrm{d}y_2\mathrm{d}s_2$$

where $S(y,s)$ is the source strength. Thus calculation of p_1 and p_2 allows the mean and variance of concentration to be derived. The densities p_1 and p_2 are usually calculated by Monte Carlo releases of particle pairs satisfying

If it is desired to model higher moments like $E\{c^2\}$, then similar equations can also be derived for them. For the appropriate set of equations to be used to model the desired set of moments, it is necessary to close the equations by expressing the unknown terms as functions of the quantities which are being modelled explicitly. Closure approximations of this sort form the basis for one popular class of models of turbulent dispersion, namely eddy diffusion and second-order closure models. It should, however, be emphasised that such approximations are not based on sound scientific foundations in general.

Eddy diffusion and second order closure models

Eddy diffusion models assume that $E\{u_i c\} = -\kappa_T \, \partial C/\partial x_i$, where the *eddy diffusivity* κ_T is specified (possibly in terms of local properties of the velocity and concentration fields). In more sophisticated versions κ_T and the eddy viscosity ν_T are expressed in terms of the turbulent kinetic energy $K = E\{\frac{1}{2} u_j u_j\}$ and the turbulent dissipation rate $\epsilon = 2\nu E\{u_{ij} u_{ij}\}$. Equations for these quantities (which have to be closed) are then solved. This is the so-called K-ϵ method. In this method the variance can be modelled, unlike the simple eddy diffusion models which only predict the mean. Variants of this method model length or vorticity scales instead of the rate ϵ.

A further development of this approach was the second-order closure models. In these, additional evolution equations for the Reynolds stresses $E\{u_i u_j\}$ are solved, after having themselves been empirically closed (Newman *et al.*, 1981). In all of these equations the closures are usually based on theory for isotropic limits and on principles like coordinate-frame invariance, and quantities which must be positive (like K) should remain positive in the models (realisability). These principles lead to forms for the closures involving undetermined constants which are usually set by fitting to simple laboratory experiments. Useful reviews of these models are provided by Launder (1990) and Speziale (1991). The latter is for turbulence models only, but also includes eddy viscosity models.

As discussed in these reviews, second-order closure models have been more successful than K-ϵ models for relatively simple engineering flows in laboratories (e.g. rotating or swirling channel flows). Atmospheric applications include Mellor and Yamada (1974), Lewellen and Teske (1976), Yamada (1977), Andrén (1990) and Lobocki (1992). The general problem with these closure models are the neglect of nonlocal and history effects in making the closures, and problems near solid boundaries where

$$dX = V dt$$
$$dV_i = a_i(X,V,t)dt + b_{ij}(X,V,t)d\xi_j$$

where $X=(x_1,x_2)$, $V=(v_1,v_2)$ and v_1,v_2 are the particle velocities. The $d\xi_j$ are taken to be increments of independent Wiener processes (i.e. they have a Gaussian distribution with zero mean and variance dt, and non-overlapping increments are independent). The choice of $B_{ij} = \frac{1}{2}b_{ik}b_{jk}$ determines the distribution of $b_{ij}d\xi_j$, so the model is fixed by the choice of a_i and B_{ij}. Criteria for this choice include that the short-time behaviour agrees with theory, and that an initially well-mixed particle distribution should remain so. Recent examples of such models are Stapountzis *et al.* (1986), Sawford and Hunt (1986), Thomson, (1986,1990) and Kaplan and Dinar (1989).

Random walk models have the advantage over closure models that they can include the effect of length scales. However, if desirable properties of the above kind are satisfied, then for fixed y_1 the ensemble of trajectories $x_1(t)$ depends on y_2 (Thomson, 1990). This is obviously undesirable, although simulation results suggest it might not be too important in practice. The main drawback to these models is the oversimplified turbulence properties. In addition, the computing time required to generate adequate statistics is large.

PDF OF CONCENTRATION

While the models described above can predict low-order moments of concentration, they cannot be used to predict the pdf. A possible approach is to attempt to solve equation (*3*) for the pdf. This is only possible if closure assumptions are used, analogous to those described for the solution of equations (*6*) and (*7*). An overview of such methods is provided by Pope (1985). They suffer from all of the problems of the closures discussed above, and methods used so far also assume that the pdf is approximately Gaussian.

An alternative method is to use models for the low-order moments to determine the parameters in a pre-chosen form of pdf. The chosen form would be one which fits observations to a desired degree of goodness. If different forms fit well under different circumstances then they can be utilised as appropriate. This general approach has been demonstrated by Jakeman *et al.* (1988), who term it the hybrid method. If it is deemed necessary to use a pdf with more than two parameters, then an alternative

to modelling higher-order moments using models of the above type is to use a simpler physical or empirical model relating the different moments. For example a model like that of Chatwin and Sullivan (1990b) can be used. As is described briefly in the next section, this is based on a combination of physical arguments from the case of zero molecular diffusion, and of experimental observation.

It is typical of atmospheric dispersion that the concentration is intermittent, i.e. there is a significant probability of the observed concentration being indistinguishable from zero (there are difficulties with making a precise definition of this property – see for example Sreenivasan, 1985; Chatwin and Sullivan, 1989; Mole and Jones, 1993). As a consequence the pdf $p(\theta)$ is often idealised as

$$p(\theta) = (1 - \gamma)\delta(\theta) + \gamma f(\theta) \qquad \theta \geq 0 \qquad (8)$$

where $f(\theta)$ is the pdf conditional on the concentration being non-zero and γ is the intermittency. Commonly suggested pdfs $f(\theta)$ for fitting to data include the lognormal, exponential and truncated normal. A physical justification (see Csanady, 1973) has been proposed for the lognormal, but the underlying assumptions are not well-founded, and it does not appear to fit experimental data very well. The exponential fits reasonably well in the fringes of plumes, but not very well on the plume centreline, and of these three, the truncated normal gives the best all-round fit (Mylne and Mason, 1991), which might be expected from its flexibility of shape. Use of the truncated normal is also sometimes justified on the grounds that it is the two-parameter maximum entropy solution (e.g. Lewellen and Sykes, 1986).

The goodness of fit has generally not been carried out very rigorously. (There is, however, a body of techniques which can be applied to *estimate* parameters of the above distributions, and the Weibull and gamma distributions; and to *identify* the most appropriate distribution to use in practice. See Bai *et al.*, 1991, for new maximum likelihood methods; Bai *et al.*, 1990, for a comparison of methods of moments versus maximum likelihood; and Bai *et al.*, 1992, for the identification methodology. Taylor *et al.*, 1991, describe a related computer program.) The goodness of fit has also tended to be judged by the main bulk of the distribution, so the tail of the distribution may not be well predicted by using these pdfs. Furthermore, these fits have been made to data which include effects of noise and instrument smoothing (as discussed below), as well as explicitly applied time averaging. Thus the degree of applicability of these fits to

the pdf of the true concentration is not clear.

SOME RECENT IDEAS

A major source of difficulty in modelling turbulent dispersion is the interaction of molecular diffusion and turbulent advection. While molecular diffusion has little effect on the mean concentration, it plays an important role through dilution, which causes the dissipation of concentration variance (and, indeed, higher-order moments of concentration). Thus a framework in which the effects of diffusion and advection are decoupled may allow relatively simple models to be developed. Here we shall briefly describe two possible ways of doing this.

The first is to consider fluid which has emanated from the source separately from that which has not. This was fully discussed by Chatwin and Sullivan (1989). If $\kappa = 0$ then for a source of uniform concentration there is a precisely defined intermittency $\pi_0 = C/\theta_0$, where θ_0 is the source concentration, and the pdf of concentration is

$$p(\theta) = (1-\pi_0)\delta(\theta) + \pi_0\delta(\theta-\theta_0) \tag{9}$$

When $\kappa > 0$ the pdf can be written exactly as

$$p(\theta) = (1-\pi_0)g(\theta) + \pi_0 f(\theta)$$

where g is the pdf conditional on being in non-source fluid, and f is the pdf conditional on being in source fluid. Since C is little different from when $\kappa = 0$, π_0 can be estimated as C/θ_0. Furthermore, since the highest concentrations occur in source fluid, the distribution of high concentrations will be determined by f.

The moments based upon (9) can all be expressed as functions of C and θ_0. Chatwin and Sullivan (1990b) used physical arguments to generalise these expressions to the case of non-zero molecular diffusion. This allows the moments to be expressed in terms of C and a local mean concentration scale C_0 (e.g. the maximum mean concentration in a plume cross-section, or the maximum mean concentration at a particular instant in an instantaneously released cloud). Two parameters α and β are also

required. Experimental observations have been shown to be well fitted by these expressions (see Chatwin and Sullivan, 1990b). For the self-similar cases considered by Chatwin and Sullivan (1990b) these parameters are constants, but in general they may depend on x and t. Examples of the latter are provided by Moseley (1991) and Sawford (1992, personal communication).

The second method is to integrate over the whole cloud or over the whole plume cross-section. The easiest case to consider is that of an instantaneously released cloud. The following theory has been developed in Chatwin and Sullivan (1990a), Moseley (1991), Sullivan and Ye (1992a,b) and Sullivan and Heagy (1992). When equation (2) is multiplied by $n\Gamma^{n-1}$ and integrated over all space ($a.s.$) the following equation results on taking the mean:

$$\frac{\partial}{\partial t} \int_{a.s.} E\{\Gamma^n\} \mathrm{d}V = -\kappa n(n-1) \int_{a.s.} E\{\Gamma^{n-2}(\nabla\Gamma)^2\} \mathrm{d}V \quad n \geq 1 \qquad (10)$$

Moseley (1991) used a closure scheme for the right-hand side of equation (10). Use of equation (10) for $n = 2,3$ together with an assumed Gaussian profile for C allowed predictions to be made for the downwind variation of α and β in a plume. These agreed qualitatively with experimental results. Working with the cloud-averaged quantities has the advantages that the choice of the closure scheme is probably less critical, and that when enough spatially distributed measurements are available, convergence of observed statistics will require fewer replications, or a shorter run for a steady case, than for point quantities. This approach is therefore more realistic in terms of demands on data.

Chatwin and Sullivan (1990a) and Sullivan and Ye (1992a) developed the cloud-average approach to apply to the pdf. A new quantity $\bar{p}(\theta,t)$ can be defined by

$$\bar{p}(\theta,t) = \frac{\theta}{Q} \int_{a.s.} p(\theta;x,t) \mathrm{d}V$$

where Q is the total mass of material released. This has the following interpretation: $\bar{p}(\theta,t)\mathrm{d}\theta$ is the expected proportion of the total mass of material which is found in the concentration interval $[\theta,\theta+\mathrm{d}\theta]$. This is a quantity which should be capable of robust estimation given a suitable measurement system. The litmus fluid technique of Baines and Corriveau

(1992) is one which is particularly suited to this. The moments of $\bar{p}$ are related simply to the moments of Γ, integrated over all space. For applications where damage occurs above some threshold, $\bar{p}$ provides the expected proportion of total mass above the critical level. The quantity $\bar{p}$ is not affected by turbulent advection except indirectly – only molecular diffusion can alter $\bar{p}$ from its initial delta function (for a release of uniform source concentration).

MODEL VALIDATION AND UNCERTAINTY

Assessment of the relative performance of the models described in the second section of this chapter is not easy. In addition to qualitative arguments about the shortcomings of particular models, one would like to be able to make quantitative comparisons for particular cases of interest. However, when models appear in the published literature, quantitative comparison with other models is usually limited. Model comparison based on evaluations against experimental observations is made difficult by the frequent use of different data sets, or of different features of the predictions, for evaluations of each model. Major benefits could result from a more consistent and coordinated approach to model evaluation, as advocated by Britter (1991).

The scientific appropriateness of individual aspects of models can be examined through theoretical argument and by comparison with experimental data (for example relationships assumed in closure schemes can be tested directly against data), and this is obviously an important source of improvements to models. However, the ultimate test of a practical model is how good its end predictions are. For this reason the validation of model predictions against experimental observations is vital, both for the assessment of model performance, and for the identification of weaknesses and the suggestion of potential improvements. This process must include the consideration of sources of uncertainty, as distinct from model errors resulting from incorrect representation of the physical processes involved in dispersion.

Over the past decade an increasing effort has been devoted to model validation and to the assessment of uncertainty. This effort began with the recommendations of the 1980 American Meteorological Society workshop at Woods Hole (Fox 1981), and has been continued by a range of studies, including Fox (1984), Smith (1984), Cox and Tikvart (1990), Weil *et al.* (1992) and Hanna *et al.* (1992). Recently a workshop was held at Risø

(Olesen and Mikkelsen, 1992) to discuss objectives for short-range atmospheric dispersion models, where these aspects were among those considered. However, the range of models to which these analyses has been applied is limited. Mostly these have been operational models based on Gaussian plumes, and on box models for heavy gas dispersion. To our knowledge this type of intercomparison of models against each other and against data has not been carried out for the types of statistical models described above.

Here we shall briefly summarise some of the techniques currently in use. We shall also consider a more general application suitable for the new generation of models which, as described above, will make predictions for more general parameters than just the mean concentration. Our comments here will apply to general models of atmospheric dispersion, and not just to those considered in our second section.

In this general context four questions need to be posed:

Question 1: Which quantities are being predicted (and so need to be validated against observations)? In most models for which validation has so far been carried out only the mean (or an often ill-defined average, e.g. a time average) has been predicted, but in general one would want to validate models for higher-order moments and for the pdf of concentration at the very least.

Question 2: Which ensemble are these quantities being predicted for? At one extreme the ensemble may be for a very specific set of physical conditions, e.g. a particular mean wind speed and direction and atmospheric and surface conditions. At the other extreme the ensemble would include all conditions experienced at a particular site over the course of many years. The latter case is not likely to be attempted except by purely statistical models, but the other end of the scale is unlikely to be precisely attained either, since data for validation will usually include at least some variation in mean wind speed, direction etc.

Question 3: What size observational sample is available for the ensemble being considered? There may be only one observation, for example for an instantaneous atmospheric release, or there may be a large sample from which parameter estimates and their confidence limits may be relatively easily derived.

Question 4: How significant are the effects of instrument smoothing and

noise on the observations?

The approach to validation will differ according to the answers to these questions. The currently well-established methods apply to the case when only the ensemble mean concentration is predicted, and when only one sample is available for each ensemble. In order to analyse model errors, the sources of uncertainty in both model predictions and in observations need to be identified and quantified.

The major sources of uncertainty in observations are measurement effects associated with instrumental shortcomings, and natural variability resulting from the random nature of turbulent dispersion. In principle it is possible to reduce the measurement uncertainty through instrument improvements. This will also have the effect of making the perceived natural variability (which will increase) closer to the true natural variability. The natural variability produces an irreducible uncertainty in a *single* observation of concentration, but *no* uncertainty in probabilistic quantities such as the mean and variance of concentration. Uncertainty in the estimation of the latter quantities results from sampling error, and can in principle be reduced by increasing the sample size. There is also uncertainty arising from the indeterminate nature of the ensemble which the observations represent. This is because of the difficulty of distinguishing between trends (representing change in environmental conditions) and long-term random variations for time series of steady releases. Similarly, the indeterminate nature of such variations causes difficulty in forming a meaningful ensemble for repeated unsteady releases. This problem is one which ought to be given more attention than has hitherto been the case, and places severe limits on how significantly uncertainty in the mean and variance can be reduced.

There are two obvious sources of uncertainty in model outputs. The first is errors in coding the computer programs, which are likely to be more significant the larger the model becomes. The second is uncertainties in inputs, which are then propagated through the model in a manner dependent on the details of the particular model. Inputs for validating a model will include parameters defining the ensemble of interest (including details of the emission and of the ambient atmosphere), which will have been derived from the set of observations with which the model output is to be compared. These inputs will inevitably contain the observation uncertainties described above. In the general predictive application of a model the output uncertainties resulting from input uncertainties will be significant, and it is now abundantly clear and generally recognised that any outputs must be accompanied by estimates of their uncertainty;

otherwise the model should not be used. Such estimates can be given for a specific model if the appropriate input distributions can be estimated (Lewellen and Sykes, 1989).

Suppose the population parameter being predicted is λ, and let the predicted and observed values of this parameter be λ_p and λ_o respectively. We shall label the different ensembles by a superscript i, where $i = 1,\ldots,m$; thus λ_o^i is the observed value in the ith ensemble. The approach which is now becoming standard is outlined by Weil *et al.* (1992). It treats the case where λ is the mean concentration and only one observation is available for each i, so this observation must serve as λ (with large natural uncertainty). Statistics (e.g. the geometric mean and variance) are calculated for λ_p^i/λ_o^i using all the i values, and confidence limits are estimated for these statistics using bootstrap techniques (Hanna, 1989). A good model is one which gives a confidence interval for the mean value of λ_p/λ_o which includes 1, with relatively little scatter. In this case, however, the scatter includes the natural variability of λ_o and the related uncertainty in inputs, and so may be large, regardless of how good the model is. One ought also to examine the pdf of λ_p/λ_o – ideally it would have a peak at or close to 1. A model which gave an acceptable confidence interval for the mean, but had two peaks in the pdf to either side of 1 would not be satisfactory since it would probably not give good results for most ensembles. It would only appear acceptable when results were averaged over many ensembles. This general approach assumes that λ_p^i and λ_o^i for different i are representative of the ensemble for fixed i, except perhaps for some scale factor dependent on i (as discussed by Weil *et al.*, 1992). The variation of model performance with i is examined through the variation of residuals λ_p/λ_o with particular physical parameters, e.g. downwind distance, after grouping similar ensembles together. This can suggest the physical conditions for which the model is poor, and for which improvements to the model should be sought. If there were no uncertainty in λ_p^i and λ_o^i this approach would give a measure of the overall performance of the model for a range of ensembles. When λ is the mean concentration and more than one observation is available for each ensemble this method can still be used with all observed concentrations being used as estimates of λ. However, when λ is some other parameter, e.g. the standard deviation of concentration, the observed concentrations in a particular ensemble must be combined to give an estimate for λ before such an analysis can be performed.

When there is more than one observation for each ensemble then estimates of the natural uncertainty (and of the input uncertainty) can be

made. The non-parametric bootstrap methods can then be used to derive confidence intervals for λ_o^i for each i. In such a case the input variability can be used to estimate confidence intervals for λ_p^i (as in Lewellen and Sykes, 1989). If the confidence intervals for λ_p and λ_o do not overlap then the model performance should be deemed unsatisfactory. Statistics for λ_p/λ_o could also be generated using the set of λ_p and the set of bootstrap estimates for λ_o. If statistics indicating the success of a model over all ensembles i were desired then the values of λ_p^i/λ_o^i can be combined over all i, and statistics calculated in the standard ways. These statistics can be used to reject poorer models by estimating the significance of differences between models.

A drawback to the above methods is that differences between λ_p and λ_o after taking uncertainties into account (e.g. by constructing confidence intervals) may not be entirely a result of model errors. In particular measurement errors are systematic. For example, lack of instrument resolution in space and/or time causes smoothing of the true concentration so that the observed standard deviation and extreme values of concentration are reduced. In addition measurement noise always increases the observed standard deviation of concentration. It is possible to model concentrations smoothed in space-time in a manner corresponding to that of the measuring instrument, so that the smoothing is no longer considered as a measurement error, but this adds extra difficulties to the modelling. Furthermore, it is the unsmoothed concentration which is physically important to the applications described above, so it is better to treat instrument smoothing and noise as measurement errors. For effective model evaluation it is then desirable to reduce these effects as much as possible. Instrument improvements will allow some progress in this respect, but will never entirely remove them. Signal processing techniques will also allow some of these errors to be reduced (Mole, 1990).

FUTURE WORK AND PROSPECTS

In addition to the problems already discussed there are two significant areas where further modelling work is required for the applications described here. For flammable gases there is the burn-up problem mentioned in the introduction. Because of the spatial and temporal resolution required over a wide area for *each realisation* this problem has not really been tackled in models or experiments. Perhaps the best hope

of immediate progress is empirical work based on laboratory laser sheet experiments. In the case of toxic gases more work is required on the toxicological effects (both short and long term). In particular a better idea is needed of what function of the concentration field is a suitable measure of harm. Assuming the best available estimate to be the dose as defined in the introduction, the best present modelling approach is probably again to seek empirical extensions to existing work.

In both of these applications a significant role is played by high concentrations, and regulations are often framed in terms of high concentrations. Modelling of the whole pdf is unlikely to give good results for high concentrations, so it may well be desirable to use a separate model for the tail of the pdf. Little use has been made of statistical extreme value theory in this field – one of the few examples is given in Davison and Smith (1990). This theory gives extreme value distributions which hold asymptotically in the limit of large sample size, for particular classes of stochastic process (for example, at the simplest level, when sample members are independent). This provides an advantageous basis for the development of empirical models.

An obstacle to further model improvement is provided by the lack of good quality experimental data. There are two drawbacks to the experimental data. Firstly, there is a shortage of field experiments, in particular replicated experiments, with which to validate and develop models. This is especially important for unsteady cases such as those resulting from accidental releases, or when the meteorology is changing rapidly. Even when the release is steady, it is far from clear that atmospheric conditions are sufficiently stationary for convergent and meaningful statistics to be obtained from a single release. There has been a tendency on the part of experimentalists to vary release conditions and source-receptor arrangements, when it might have been profitable to use the experimental time for replicated experiments under near-identical conditions. Although expensive, this is an area which needs attention for modelling to progress.

Secondly, poor resolution, noise, baseline drift, and uncertainties in the accuracy and stability of calibrations all corrupt measurements. (These problems can be compounded by poor analogue-digital conversion and inappropriate thresholding.) Improvements in instrument design and in analytical techniques will alleviate these problems but never entirely remove them.

The most fundamental mathematical problem which hinders modelling is the closure problem, discussed in the introduction. The only real hope of avoiding it is by repeated numerical solution of the Navier-Stokes

equation and equation (2) with all relevant scales resolved. As discussed above, the necessary computer power is orders of magnitude beyond that currently available. While increases in computer power will enable advances, this objective is unlikely to be achieved in the foreseeable future. Even if it is achieved, the problem of defining an appropriate ensemble for practical applications will remain.

ACKNOWLEDGEMENTS

We would like to thank the following organisations for their support: the UK Chemical and Biological Defence Establishment and Commission of the European Community (PCC), the National Science and Engineering Research Council of Canada (PJS), and the UK Health and Safety Executive (NM). We are also grateful to David Hall of Warren Spring Laboratory for allowing us to use the photographs in Figure 1, and to Liz Clarke for providing us with the graph in Figure 2.

REFERENCES

Andrén,A. 1990 Evaluation of a turbulence closure scheme suitable for air pollution applications. *Journal of Applied Meteorology*, 29: 224-239.

Bai, J., Jakeman, A.J. and McAleer, M.J. 1990 Percentile estimation of the three-parameter gamma and lognormal distributions: methods of moments versus maximum likelihood. *Mathematics and Computers in Simulation*, 32: 167-172.

Bai, J., Jakeman, A.J. and McAleer, M.J. 1991 A new approach to maximum likelihood estimation of the three-parameter gamma and Weibull distributions. *Australian Journal of Statistics*, 33: 397-410.

Bai, J., Jakeman, A.J. and McAleer, M.J. 1992 Estimation and discrimination of alternative air pollution models. *Ecological Modelling*, 64: 89-124.

Baines, W.D. and Corriveau, A.F. 1992 Mixing in turbulent jets as revealed by pH indicator. *Proceedings 2nd Caribbean Conference on Fluid Dynamics*, St. Augustine, Trinidad, pp. 142-148.

Britter, R.E. 1991 The evaluation of technical models used for major-accident hazard installations. Report to CEC DG XII.

Chatwin, P.C. 1990 Statistical methods for assessing hazards due to dispersing gases. *Environmetrics*, 1: 143-162.

Chatwin, P.C. and Sullivan, P.J. 1989 The intermittency factor of scalars in turbulence. *Physical Fluids A*, 1: 761-763.

Chatwin, P.C. and Sullivan, P.J. 1990a Cloud-average concentration statistics. *Mathematics and Computers in Simulation*, 32: 49-57.

Chatwin, P.C. and Sullivan, P.J. 1990b A simple and unifying physical interpretation of scalar fluctuation measurements from many turbulent shear flows. *Journal of Fluid Mechanics*, 212: 533-556.

Chatwin, P.C. and Sullivan, P.J. 1993 Quantitative models for environmental prediction:

a review. In *Mixing and Transport in the Environment*, K.J.Beven, P.C. Chatwin, J.Millbank (eds), John Wiley & Sons Ltd: Chichester.

Cox, W.M. and Tikvart, J.A. 1990 A statistical procedure for determining the best performing air quality simulation model. *Atmospheric Environment*, 24A: 2387-2395.

Csanady, G.T. 1973 *Turbulent Diffusion in the Environment*. D.Reidel.

Davison, A.C. and Smith, R.L. 1990 Models for exceedances over high thresholds (with discussion). *Journal of Royal Statistical Society B*, 52: 393-442.

Fox, D.G. 1981 Judging air quality model performance. *Bulletin of American Meteorological Society*, 62: 599-609.

Fox, D.G. 1984 Uncertainty in air quality modeling. *Bulletin of American Meteorological Society*, 65: 27-36.

Gifford, F.A. 1959 Statistical properties of a fluctuating plume dispersion model. *Advances in Geophysics*, 6: 117-137.

Glendening, J.W. and Burk, S.D. 1992 Turbulent transport from an arctic lead: a large-eddy simulation. *Boundary-Layer Meteorology*, 59: 315-339.

Griffiths, R.F. 1991 The use of probit expressions in the assessment of acute population impact of toxic releases. *Journal of Loss Prevention in Process Industries*, 4: 49-57.

Hall, D.J., Hollis, E. J. and Ishaq, H. 1982 A wind tunnel model of the Porton dense gas spill field trials. Warren Spring Laboratory Report LR394(AP).

Hanna, S.R. 1989 Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods. *Atmospheric Environment*, 23: 1385-1398.

Hanna, S.R., Chang, J.C. and Strimaitis, D.G. 1992 Hazardous gas model evaluation with field observations. Submitted to *Atmospheric Environment*.

Jakeman, A.J., Simpson, R.W. and Taylor, J.A. 1988 Modeling distributions of air pollutant concentrations-III. The hybrid deterministic-statistical distribution approach. *Atmospheric Environment*, 22: 163-174.

Kaplan, H. and Dinar, N. 1989 The distribution function of the fluctuations of a passive scalar in homogeneous isotropic turbulence. *Journal of Applied Meteorology*, 28: 1176-1184.

Launder, B.E. 1990 Phenomenological modelling: present and future? In *Whither Turbulence? Turbulence at the Crossroads*, J.L.Lumley (ed.), Lecture Notes in Physics 357, Springer-Verlag: Berlin, pp. 439-485.

Lewellen, W.S. and Sykes, R.I. 1986 Analysis of concentration fluctuations from lidar observations of atmospheric plumes. *Journal of Climate and Applied Meteorology*, 25: 1145-1154.

Lewellen, W.S. and Sykes, R.I. 1989 Meteorological data needs for modeling air quality uncertainties. *Journal of Atmospheric and Oceanic Technology*, 7: 759-768.

Lewellen, W.S. and Teske, M. 1976 Second order closure modelling of diffusion in the atmospheric boundary layer. *Boundary-Layer Meteorology*, 10: 69-90.

Lobocki, L. 1992 Mellor-Yamada simplified second-order closure models: analysis and application of the generalized von Karman local similarity hypothesis. *Boundary-Layer Meteorology*, 59: 83-109.

Mason, P.J. and Thomson, D.J. 1992 Stochastic backscatter in large-eddy simulations of boundary layers. *Journal of Fluid Mechanics*, 242: 51-78.

Mellor, G.L. and Yamada, T. 1974 A hierarchy of turbulence closure models for planetary boundary layers. *Journal of Atmospheric Science*, 31: 1791-1806.

Moeng, C.-H. 1984 A large-eddy simulation model for the study of planetary

boundary-layer turbulence. *Journal of Atmospheric Science*, 41: 2052-2062.

Mole, N. 1990 Some intersections between turbulent dispersion and statistics. *Environmetrics*, 1: 179-194.

Mole, N. and Jones, C.D. 1993 Concentration fluctuation data from dispersion experiments carried out in stable and unstable conditions. To appear in *Boundary-Layer Meteorology*.

Moseley, D.J. 1991 *A Closure Hypothesis for Contaminant Fluctuations in Turbulent Flow*. M.Sc. dissertation, University of Western Ontario.

Mylne, K.R. and Mason, P.J. 1991 Concentration fluctuation measurements in a dispersing plume at a range of up to 1000 m. *Quarterly Journal of Royal Meteorological Soci*[illegible] 117: 177-206.

Newman, G.R., Launder, B.E. and Lumley, J.L. 1981 Modelling the behaviour of homogeneous scalar turbulence. *Journal of Fluid Mechanics*, 111: 217-232.

Nieuwstadt, F.T.M. and de Valk, J.P.J.M.M. 1987 A large eddy simulation of buoyant and non-buoyant plume dispersion in the atmospheric boundary layer. *Atmospheric Environment*, 21: 2573-2587.

Olesen, H.R. and Mikkelsen, T. (eds) 1992 Proceedings of the workshop on Objectives for Next Generation of Practical Short-Range Atmospheric Dispersion Models, 6-8 May 1992, Risø.

Pasquill, F. and Smith, F.B. 1983 *Atmospheric Diffusion*, 3rd edition. Ellis Horwood: Chichester.

Pope, S.B. 1985 Pdf methods for turbulent reactive flows. *Progressive Energy Combustion Science*, 11: 119-192.

Reynolds, W.C. 1990 The potential and limitations of direct and large eddy simulations. In *Whither Turbulence? Turbulence at the Crossroads*, J.L.Lumley (ed.), Lecture Notes in Physics 357, Springer-Verlag: Berlin, pp. 313-342.

Rogallo, R.S. and Moin, P. 1984 Numerical simulation of turbulent flows. *Annual Reviews in Fluid Mechanics*, 16: 99-137.

Sawford, B.L. and Hunt, J.C.R. 1986 Effects of turbulence structure, molecular diffusion and source size on scalar fluctuations in homogeneous turbulence. *Journal of Fluid Mechanics*, 165: 373-400.

Smith, M.E. 1984 Review of the attributes and performance of 10 rural diffusion models. *Bulletin of American Meteorological Society*, 65: 554-558.

Speziale, C.G. 1991 Analytical methods for the development of Reynolds-stress closures in turbulence. *Annual Reviews in Fluid Mechanics*, 23: 107-157.

Sreenivasan, K.R. 1985 On the fine-scale intermittency of turbulence. *Journal of Fluid Mechanics*, 151: 81-103.

Stapountzis, H., Sawford, B.L., Hunt, J.C.R. and Britter, R.E. 1986 Structure of the temperature field downwind of a line source in grid turbulence. *Journal of Fluid Mechanics*, 165: 401-424.

Sullivan, P.J. and Heagy, W.K. 1992 A new approach to some environmental dosage problems. 14th Canadian Congress of Applied Mechanics, 30 May-4 June 1993, Queen's University, Kingston, Ontario.

Sullivan, P.J. and Ye, H. 1992a Further comment on *cloud-averaged* concentration statistics. Submitted to *Mathematics and Computers in Simulation*.

Sullivan, P.J. and Ye, H. 1992b Cloud average probability density and its application. 14th Canadian Congress of Applied Mechanics, 30 May-4 June 1993, Queen's University, Kingston, Ontario.

Taylor, J.A., Bai, J., Jakeman, A.J. and McAleer, M. 1991 PROFIT: a computer program for selecting and fitting probability distributions to data. *Proceedings 13th IMACS World Congress on Computation and Applied Mathematics*, R. Vichnevetsky and J.J.H. Miller (eds), Trinity College, Dublin, pp. 1957-1962.

ten Berge, W.F., Zwart, A. and Appelman, L.M. 1986 Concentration-time mortality response relationship of irritant and systemically acting vapours and gases. *Journal of Hazardous Materials*, 13: 301-309.

Thomson, D.J. 1986 On the relative dispersion of two particles in homogeneous stationary turbulence and the implications for the size of concentration fluctuations at large times. *Quarterly Journal of Royal Meteorological Society*, 112: 893-894.

Thomson, D.J. 1990 A stochastic model for the motion of particle pairs in isotropic high-Reynolds-number turbulence, and its application to the problem of concentration variance. *Journal of Fluid Mechanics*, 210: 113-153.

Weil, J.C., Sykes, R.I. and Venkatram, A. 1992 Evaluating air-quality models: review and outlook. *Journal of Applied Meteorology*, 31: 1121-1145.

Yamada, T. 1977 A numerical experiment on pollutant dispersion in a horizontally-homogeneous atmospheric boundary layer. *Atmospheric Environment*, 11: 1015-1024