Journal of Hazardous Materials, 16 (1987) 75–101
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

ESTIMATES OF THE MEAN CONCENTRATION AND VARIANCE FOR THE THORNEY ISLAND PHASE I DENSE GAS DISPERSION EXPERIMENTS

K.K. CARN

Department of Applied Mathematics and Theoretical Physics, The University of Liverpool, P.O. Box 147, Liverpool L69 3BX (Great Britain)

(Received January 30, 1987; accepted April 3, 1987)

Summary

This paper describes an analysis of the Thorney Island data using nonparametric statistical techniques. Particular emphasis is placed on methods; some of which are used to obtain estimates of $\bar{C}(\boldsymbol{x}, t)$, the mean concentration, and $\overline{c^2}(\boldsymbol{x}, t)$, the concentration variance, for this particular series of experiments. Qualitative and semi-quantitative relationships between these variables are sought using simple stochastic models. A principal objective is to characterise the relationship and hence draw inferences for dense gas dispersion.

Useful estimates of the time dependence of $\bar{C}$ and $\overline{c^2}$ are an important achievement of this investigation. Significantly the variability, as characterized by $R(C) = \overline{c^2}/\bar{C}^2$, is shown to be large. This is an interesting conclusion that may have important consequences for the assessment of hazards in dense gas dispersion.

1. Introduction

The dispersion of dense contaminants in the atmosphere is an important practical problem because of the hazards that are associated with many typical admixtures. Specifically, petroleum products are dangerous because of their flammability while other materials, such as chlorine or ammonia, are toxic when inhaled in large doses. These hazards have been addressed by several authors [1, 2] with the conclusion that $C(\boldsymbol{x}, t)$, the contaminant concentration by volume, is essential for a realistic assessment of the hazard. In reality, as emphasised by Chatwin [3], however $C(\boldsymbol{x}, t)$ is a random variable because the dispersion takes place in a medium that is invariably turbulent – the atmospheric boundary layer [4]. The statistical properties of $C(\boldsymbol{x}, t)$ are thus required for a description of this process in practice. They are the subject of the numerical investigations that are presented here in this paper.

It is appropriate to mention first the characterisation of $C(\boldsymbol{x}, t)$ by its complete set of moment quantities $\overline{C^r}(\boldsymbol{x}, t)$ [5]. These may be defined by an operation of ensemble averaging on C like the one described in Chatwin [3].

The most important for practical purposes are the first two moments $\bar{C}(\boldsymbol{x}, t)$, the ensemble mean concentration, and $\overline{C^2}(\boldsymbol{x}, t)$, the ensemble mean square concentration. The concentration variance,

$$\overline{c^2}(\boldsymbol{x}, t) = \overline{\{C(\boldsymbol{x}, t) - \bar{C}(\boldsymbol{x}, t)\}^2} = \{\overline{C^2}(\boldsymbol{x}, t) - \bar{C}^2(\boldsymbol{x}, t)\} \tag{1.1}$$

is thus inevitably a measure of the deviation of $C(\boldsymbol{x}, t)$ from its mean value in any one realisation of the process. This lies behind the proposal in Chatwin [3] to characterise the variability of the concentration field by the fluctuation intensity $R(C) = \overline{c^2}/\bar{C}^2$. Hence, if $R(C)$ is large, it is necessary to make use of both $\bar{C}(\boldsymbol{x}, t)$ and $\overline{c^2}(\boldsymbol{x}, t)$ for the assessment of hazards.

Experimental evidence, from a number of sources, summarised in [6] indicates that this is indeed the case for some dispersion phenomena at small scales. However, very little information about the magnitude of $R(C)$ in large scale dispersion processes is available at the present time. This is particularly true in the case of instantaneous releases where many current mathematical models (such as those summarised in [7]) do not even acknowledge the statistical nature of the phenomenon. This situation has changed recently with the availability of data from the Thorney Island heavy gas dispersion experiments which is suitable for a study of $\bar{C}(\boldsymbol{x}, t)$ and $\overline{c^2}(\boldsymbol{x}, t)$. The work described within is thus intended to remedy deficiencies in our knowledge of these properties for a particular ensemble of the Thorney Island experiments.

Some results of research along these lines and brief details of the method have been published before in Carn, Sherrell and Chatwin [8]. However, the presentation here is more comprehensive and directed towards choosing a suitable framework for the analysis of the Thorney Island experiments. Particular emphasis is placed on methods, some of which are new, and some proposals are made for future research. Thus Section 2 contains a detailed description of the methods used for the analysis of the data set, Section 3 discusses the context in which such an analysis should be interpreted and Section 4 describes the numerics of the procedure. Sections 5 and 6 contain the bulk of the discussion of the analysis and its results. In summary this investigation concludes that $R(C)$ is of order unity, consistent with most observations of small scale processes.

These results have been obtained without the use of any questionable assumptions about the dispersion of dense gases in the atmosphere and retain the full space–time dependence of the concentration field not allowed for in the only other relevant study by Davies [9]. This work may thus have important applications and implications for hazard analysis in the future.

2. Methods for the statistical estimation of multivariate regression functions

The problem that we face with the analysis of data from the Thorney Island experiments is the estimation of functions, like $\bar{C}(\boldsymbol{x}, t)$, which depend contin-

uously on space and time. Classical statistical theory provides several methods which are useful for this purpose given a sample of realisations of C, i.e. of values $C(\mathbf{x})$ at all points $\mathbf{x}$ in the domain of definition. An example illustrating a modern approach to this problem has been described by Davies [9]. The difficulty with the Thorney Island data set is that we do not have even one complete realisation of $C(\boldsymbol{x}, t)$. This is because, by necessity, concentration measurements were made only at discrete locations in space and time.

In this section we are concerned with methods for the solution of problems of the above type in statistical estimation. Hence though some of this material may be skipped much of it is an essential preliminary to the analysis described in Sections 3 and 4.

Consider first the estimation of $\bar{C}(\boldsymbol{x}, t)$, the first order regression function. The classical approach to this problem would be to fit a regression model to the data as described for example in Kendall and Stuart [10]. Thus assume that N data values $y_i = C(\boldsymbol{x}_i)$ are available and that $m(\mathbf{x}|\boldsymbol{\alpha})$, a parametric model for $\bar{C}(\mathbf{x})$, is specified. Hence,

$$y_i = \{C(\mathbf{x}_i) + c(\mathbf{x}_i)\} = \{m(\mathbf{x}_i|\boldsymbol{\alpha}) + \epsilon(\mathbf{x}_i)\}; \; i = 1, \ldots N \tag{2.1}$$

where the ϵ_i, $i = 1, \ldots n$ are the data residuals; normally assumed to be independent, identically distributed random variables (i.i.d.). The parameters $\boldsymbol{\alpha}$ need thus only be chosen, as can be achieved by the minimisation of an appropriate loss functional $S(\boldsymbol{\alpha})$. Conventionally use is made of the following,

$$S(\boldsymbol{\alpha}) = 1/N \sum_{i=1}^{N} \{y_i - m(\mathbf{x}_i|\boldsymbol{\alpha})\}^2 \tag{2.2}$$

in least squares regression.

The trouble with this approach is that it presumes the functional dependence of $m(\;)$ on $\mathbf{x}$ and $\boldsymbol{\alpha}$ is prescribed. This is the source of considerable difficulties since, in dense gas dispersion, no such model is known which retains a full dependence on the spatial coordinates $\boldsymbol{x}$. We could of course accept this deficiency and make use of a box model for $\bar{C}(\mathbf{x})$ [11]. Fortunately recent developments in statistical estimation theory suggest a viable alternative procedure. This is to develop a nonparametric model for the regression function [12], so called because it involves much weaker assumptions about the functional dependence of $m(\;)$ on $\mathbf{x}$ and $\boldsymbol{\alpha}$.

The methods we shall make use of in this work are in fact derived from nonparametric estimates of the joint probability density of concentration and sampling location $f(\theta, \mathbf{x})$. As described in Prakasa Rao [12], such an estimate is,

$$f_N(\theta, \mathbf{x}) = \{\frac{1}{N} D^{4+1}\} \sum_{i=1}^{N} J\{(\theta - \theta_i)/D, \; (\mathbf{x} - \mathbf{x}_i)/D\} \tag{2.3}$$

where $J(\theta, \mathbf{x})$ is a $4+1$ variate probability density and D is a bandwidth parameter that has yet to be specified. Note that we have,

$$\mathbf{x} = (\boldsymbol{x}, t) = (x_1/\beta_1, x_2/\beta_2, x_3/\beta_3, t/\beta_4) \tag{2.4}$$

where the $\beta_j, j = 1, \ldots 4$ are scaling parameters. Hence, because,

$$m(\mathbf{x}) = \int_0^1 \theta f(\theta \mid \mathbf{x}) \mathrm{d}\theta \tag{2.5}$$

$$f(\theta \mid \mathbf{x}) = (f(\theta, \mathbf{x}) / \int f(\theta, \mathbf{x}) \mathrm{d}\theta\} \tag{2.6}$$

it is appropriate to estimate $\bar{C}(\boldsymbol{x})$ by,

$$m_N(\mathbf{x}) = \{\int \theta f_{\mathrm{N}}(\theta, \mathbf{x}) \mathrm{d}\theta / \int f_{\mathrm{N}}(\theta, \mathbf{x}) \mathrm{d}\theta\} \tag{2.7}$$

Following arguments described in Prakasa Rao [12] this may be shown to reduce to the form,

$$m_N(\mathbf{x} \mid \boldsymbol{\alpha}) = \sum_{i=1}^{N} C_i K\{(\mathbf{x} - \mathbf{x}_i)/D\} / \sum_{i=1}^{N} K\{(\mathbf{x} - \mathbf{x}_i)/D\} \tag{2.8}$$

provided,

$$J(\theta, \mathbf{x}) = \mathrm{L}(\theta) K(\mathbf{x}) : \int \theta \mathrm{L}(\theta) \mathrm{d}\theta = 0 \tag{2.9}$$

assumptions whose validity is discussed in Carn [13]. Hence it remains only to specify the "Kernel" function $K(\mathbf{x})$ for eqn. (2.8). This may be taken to have the form,

$$K(\mathbf{x}) = \begin{cases} \text{Constant} \times (1 - |\mathbf{x}|^2) \text{ for all } \quad \forall |\mathbf{x}| \leqslant 1 \\ \qquad\qquad 0 \qquad\qquad \text{for all } \quad \forall |\mathbf{x}| > 1 \end{cases} \tag{2.10}$$

according to theoretical arguments summarised in Gasser and Muller [14] and Sacks and Ylvisaker [15].

Clearly eqn. (2.8) provides a class of regression estimators for $\bar{C}(\mathbf{x}, t)$ which depend on the definition of the bandwidth parameter D. Three possibilities in particular are of importance. Thus when;

2(i) $D(\mathbf{x}{:}\mathbf{x}_i) = h_N$, a constant, eqn. (2.8) is a Kernel type estimator;

2(ii) $D(\mathbf{x}{:}\mathbf{x}_i) = H_i(\mathbf{x})$, the distance from $\mathbf{x}$ to its K-th nearest neighbour, eqn. (2.8) is a nearest neighbour estimator;

2(iii) $D(\mathbf{x}{:}\mathbf{x}_i) = H\kappa_i$, the distance from $\mathbf{x}_i$ to its K-th nearest neighbour, eqn. (2.8) is a variable Kernel estimator.

Some details of the statistical properties of these estimators are described in Carn [13]. It is worth mentioning here that, provided, the ϵ_i are i.i.d. and $N \to \infty$ so that $ND_N^n \to 0$, $m_N(\mathbf{x})$ is asymptotically unbiased, normally distrib-

uted and of minimum possible variance [14]. This provides the statistical justification for the use of such methods in regression analysis.

Any one of 2(i)–(iii) may thus be used for the estimation of $\bar{C}(\mathbf{x}, t)$ from the Thorney Island data given a set of optimum parameters $\boldsymbol{\alpha} = (\boldsymbol{\beta}, D)$ for the procedure. To choose these we could make use of a number of possibilities including eqn. (2.2), i.e. ordinary least squares. Theoretical evidence due to several authorities however, summarised in Carn [13], suggests the use of an alternative. Hence the minimisation of the Cross Validation function,

$$S(\boldsymbol{\alpha}) = CV = 1/N \|\mathbf{B} \cdot (\mathbf{I} - \mathbf{A}) \cdot \mathbf{y}\|^2 : \mathbf{B} = \{1/(1 - a_{jj})\delta_{uj}\}_{N \times N} \qquad (2.11)$$
$$\mathbf{y} = \{\mathbf{A} \cdot \mathbf{y} + \boldsymbol{\epsilon}\} : \mathbf{A} = \{a_{ij}\}_{N \times N}$$

is recommended and is supported by the numerical evidence presented in Rice [16]. Setting $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$ at the minimum we may thus estimate $\bar{C}(\mathbf{x})$ by,

$$\hat{C}(\boldsymbol{x}) = m_N(\mathbf{x} \mid \hat{\boldsymbol{\alpha}}) \qquad (2.12)$$

i.e. $m_N(\boldsymbol{x}, \boldsymbol{y} \mid \hat{\boldsymbol{\alpha}})$.

To estimate $\overline{c^2}$ note the following inequality,

$$\mathscr{L}(\boldsymbol{y}^2) \geqslant \{\mathscr{L}(\boldsymbol{y})\}^2 \qquad (2.13)$$

[13], valid whenever $\mathscr{L}$ is linear in the data values. Then using the same optimum parameters as above and the definition contained in eqn. (1.1), such an estimate is,

$$\hat{c}^2(\mathbf{x}) = \{m_N(\mathbf{x}, \boldsymbol{y}^2 \mid \hat{\boldsymbol{\alpha}}) - \{m_N(\mathbf{x}, \boldsymbol{y} \mid \hat{\boldsymbol{\alpha}})\}^2\} \qquad (2.14)$$

This statistical methodology, i.e. eqns. (2.8), (2.11) and (2.14), has been used to obtain the estimates of $\bar{C}(\mathbf{x})$, $\overline{c^2}(\mathbf{x})$ that are described in Section 5 of this paper.

Finally we note that an interesting alternative to eqn. (2.8) is suggested by the observation that $C(\boldsymbol{x}, t)$ satisfies the following integral constraint:

$$\int C(\boldsymbol{x}, t)\, \mathrm{d}\boldsymbol{x} = \int \bar{C}(\boldsymbol{x}, t)\, \mathrm{d}\boldsymbol{x} = V_g \qquad (2.15)$$

i.e. the conservation of mass condition. Thus it may be appropriate to estimate $C(\mathbf{x})$ by a time dependent probability density function. Such an estimator [13] is:

$$M_N(\boldsymbol{x}, t) = V_g \{f_N(\theta(\boldsymbol{x}), t) / \int f_N(\theta(\boldsymbol{x}), t)\, \mathrm{d}\boldsymbol{x}\}$$
$$= V_g \{\sum_{i=1}^{N} C_i K\{(\boldsymbol{x} - \boldsymbol{x}_i)/D, (t - t_i)/D\} / \sum_{i=1}^{N} C_i I\{(t - t_i)/D\}\} \qquad (2.16)$$

$$I(t) = \int K(\boldsymbol{x}, t)\, \mathrm{d}\boldsymbol{x} \qquad (2.17)$$

though its statistical properties are unknown at the present time. Note also that eqn. (2.16) is nonlinear in the data values and so precludes the use of eqn. (2.14) for the estimation of $\overline{c^2}$. As a consequence we refer to eqn. (2.16) only tentatively in this study.

All the estimates described in this section retain the fully three dimensional dependence of the concentration field. Their significance is that they embody no questionable physical assumptions, unlike those in current parametric models of dense gas dispersion.

3. A model ensemble of the Thorney Island experiments

To apply the methodology of Section 2 to the Thorney Island data we must first specify an ensemble for the analysis of these experiments [3]. As a preliminary it is useful to review the characteristic features of the data set as apparent, for example, from McQuaid and Roebuck [17]. Thus typically 10,000 concentration values $\{C(\boldsymbol{x}_i, t_j)\}$ are available for each of the 13 Thorney Island Trials 7–19. These were recorded on the observation site of roughly 500 m square and cover several hundreds of seconds at 0.6 s intervals. Not more than 73 spatially separated concentration sensors detected gas, at fixed locations, at four heights of 0.4, 2.4, 4.4 and 6.4 m above the ground. The horizontal separation of the recording sites was thus typically of order 100 m (see Fig. 3.1) which may be compared with the initial dimensions of the contaminant distribution; cylindrical of radius 7 m and of height 13 m.

While it may seem that this is ample data for the regression procedure of Section 2 to yield reliable estimates of $\bar{C}(\mathbf{x})$, $\overline{c^2}(\mathbf{x})$ there are in fact some essential difficulties with such an approach. Thus observations reveal that a typical horizontal dimension of the dispersing cloud is of order 100 m. In addition the values of $C(\boldsymbol{x}_i, t_j)$ recorded in any one realisation of $C(\boldsymbol{x}, t)$ are likely to be strongly correlated from point to point in space, at least as a consequence of the conservation of mass condition eqn. (2.15). These features are statistically undesirable and very poor estimates of $\bar{C}(\mathbf{x})$, $\overline{c^2}(\mathbf{x})$ are likely to result. More observations of $C(\boldsymbol{x}, t)$ are to be desired but none are available as each of the Phase I trials was distinct.

These difficulties can fortunately be resolved by an appropriate choice of the system ensemble for the dispersion process [3]; that is by a precise description of the conditions under which two distinct trials can be considered equivalent. This definition can include the variation of the physical parameters of the experiment within certain limits and is justified physically provided it remains relevant to the practical problem under consideration. It is thus necessary to choose an ensemble which is appropriate for the purpose of hazard assessment. We shall discuss several possibilities for this choice in this section.

It is important first to classify the Thorney Island experiments according to some measure of their physical compatibility. The Richardson number

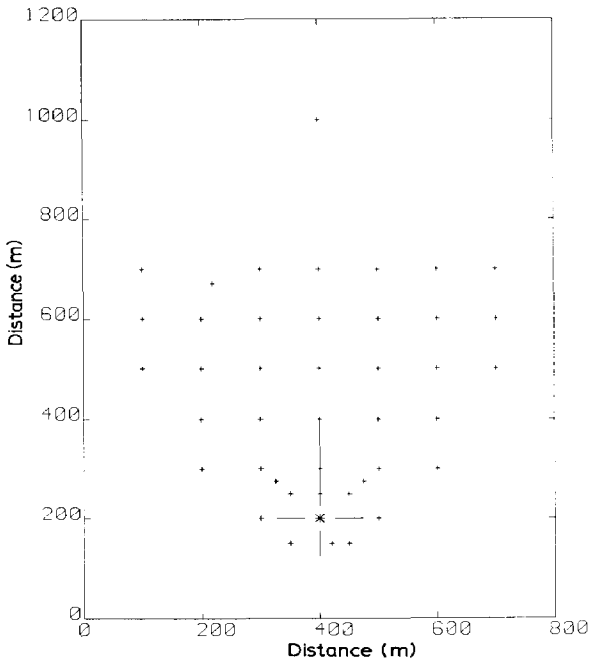

Fig. 3.1. The distribution of sampling across Trials 7–19: + concentration sensors, * origin of release.

$$Ri = \{g(\rho_1 - \rho_0)/\rho_0\}\{h/u^2\} \tag{3.1}$$

(see McQuaid and Roebuck [17]) is used almost universally for this purpose. Here $(\rho_1 - \rho_0)/\rho_0$ is the initial excess density, h is the initial cloud height and u is characteristic of the turbulent fluid velocity. Values of these parameters for the Thorney Island experiments are included in Table 3.1 (taken from McQuaid and Roebuck [17]). In this study we shall make comparisons on the basis of the following alternative to eqn. (3.1):

$$Q_\kappa(t) = 1/t \int_0^t {}_q Ri(200)/{}_\kappa Ri(t)\ \mathrm{d}t \tag{3.2}$$

since this emphasizes that, in reality, Ri is a function of time as well as an ensemble mean property. Figure 3.2 displays the variation of the $Q_\kappa(t)$ across the Thorney Island experiments. It is apparent that these records can be divided into two categories within which the $Q_\kappa(t)$ are within a factor of 4; i.e. for large values Trials 13–16, 18 and 19* and for small values Trials 7–12, and 17. On

*In this study trial 16 was excluded as it is clear from data given in McQuaid and Roebuck [17] that it corresponds to the release of an untypically small volume of gas.

TABLE 3.1

Environmental conditions of the Thorney Island experiments (from McQuaid and Roebuck [17])

Trial No.	Wind Speed m/s	Stability class	Gas density	Sensors seeing gas
7	3.2	E	0.75	55
8	2.4	D	0.63	73
9	1.7	F	0.60	62
10	2.4	C	0.80	11
11	5.1	D	0.96	23
12	2.6	E	1.37	65
13	7.5	D	1.00	47
14	6.8	C/D	0.76	50
15	5.4	C/D	0.41	38
16	4.8	D	0.68	45
17	5.0	D/E	1.20	62
18	7.4	D	0.87	60
19	6.4	D/E	1.12	67

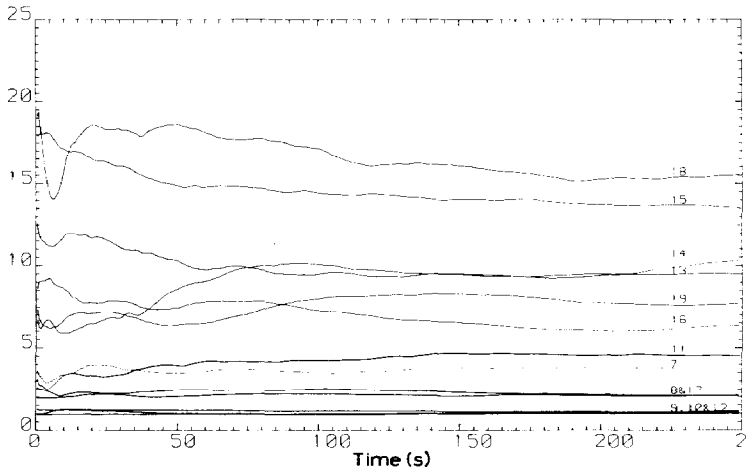

Fig. 3.2. Comparative Richardson numbers across Trials 7–19.

this basis it was decided to choose Trials 13–15, 18 and 19 as the components of a preliminary ensemble.

It thus remains, only to specify the role of the wind direction ϕ in Trials 13–15, and 19 as this is the one remaining free parameter. Of course ϕ must be defined as the time average of the instantaneous wind heading $\phi(t)$ and it is the statistical variation of this propety, with respect to a particular reference direction $\chi = \psi$, that is fundamental in characterising the ensemble. We thus have at least three distinct possibilities for the system ensemble as a consequence of the following simple symmetry assumptions:

3(i) The statistical variation of $\phi(t)$, measured absolutely in the plane, is laterally symmeytric about $\chi = \bar{\phi}$, the ensemble mean wind direction;

3(ii) The statistical variation of $\phi(t)$, measured relative to ϕ in the plane, is laterally symmetric about $\chi = \phi$, the realised wind direction;

3(iii) The statistical variation of $\phi(t)$, measured with respect to any particular direction ω in the plane, is radially symmetric about $\chi = \omega$.

In the ensemble corresponding to 3(i) equivalent realisations of $C(\boldsymbol{x}, t)$ are compared by superimposing samples directly in the horizontal plane. This is because $\phi(t)$ varies with respect to a well defined fixed direction $\chi = \bar{\phi}$. For the Thorney Island data set this means that records from the same instruments in different experiments occupy the same sites in the system ensemble. This definition has been considered quite widely and is illustrated for example in Fig. 15.2 of McQuaid and Roebuck [17].

In the ensemble corresponding to 3(ii) however equivalent realisations of $C(\boldsymbol{x}, t)$ are compared by superimposing samples so that the wind directions ϕ_i are coincident in the horizontal plane. This is because $\phi(t)$ varies with respect to $\chi = \phi$, the realised wind direction. A consequence for the Thorney Island data set is that in the system ensemble observations from the same instruments do not in general occupy the same sites. The profile of the mean cloud is rather better defined than in 3(i) and is illustrated here for Trials 13–15, 18 and 19 in Fig. 3.3.

In ensemble 3(iii) we have yet another alternative to 3(i) or (ii). Here equivalent realisations of $C(\boldsymbol{x}, t)$ are compared by superimposing samples from all radially equivalent points in the horizontal plane. This is because, as $\phi(t)$ is radially symmetric, the statistical properties of $C(\boldsymbol{x}, t)$ are dependent only on the same t, the height h and the horizontal radial coordinate r. The distribution of sample sites is intermediate to that of (i) or (ii) as may be deduced by an examination of Fig. 3.3.

The use of either of the ensembles 3(i)–(iii) is thus an equally valid basis for an analysis of the Thorney Island experiments. However each is accrued with its own peculiar advantages and disadvantages. Thus although 3(i) allows corresponding instruments to be compared it admits no progress as regards the spatial definition in the data set. This situation is reversed for ensemble 3(ii) which is more characteristic of a single realised cloud (or perhaps of experiments in wind tunnel). Finally in ensemble 3(iii) we have a perspective on these experiments, appropriate to dispersion in an isotropic flow, which achieves the greatest condensation of the spatial data set. It remains to decide which of these alternatives is the more relevant for an assessment of hazards in dense gas dispersion.

The suggestion advanced in this paper is that the most appropriate framework for a study of this problem is in fact 3(ii). This is because in any given release of a heavy gaseous material serious hazards are most likely to occur

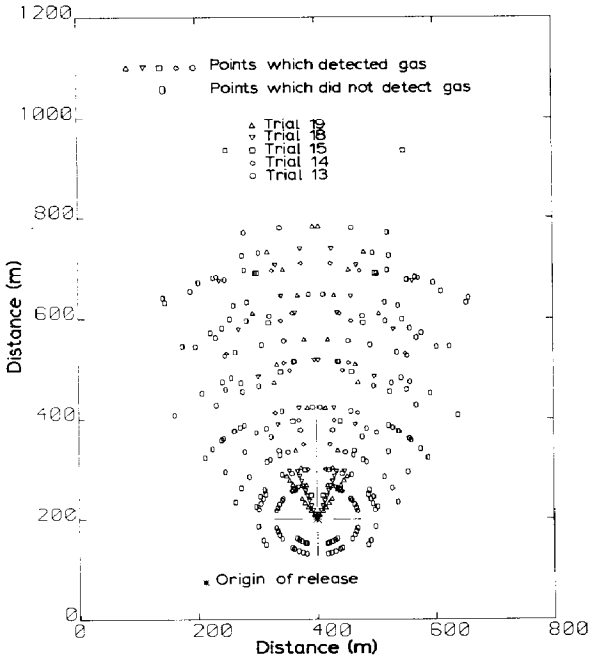

Fig. 3.3. A realisation of ensemble 3(ii) for Trials 13–15, 18 and 19.

astride a path coincident with the mean flow direction. It is thus important to have estimates of $\bar{C}(\mathbf{x})$, $\overline{c^2}(\mathbf{x})$ for dispersion in an ensemble of releases consistent with such a realisation. For this reason we shall consider the estimation of these statistical properties from Trials 13 to 15, 18 and 19 in an ensemble of type 3(ii) using the statistical methodology described in Section 2.

4. Algorithmic and numerical details of the estimation procedure

Before discussing the algorithmic details involved in obtaining estimates of $\bar{C}$, $\overline{c^2}$ from the Thorney Island data we require to construct a realisation of the ensemble 3(ii). Estimates of the wind direction ϕ were obtained by choosing stable values of the time mean wind heading,

$$\hat{\phi} = 1/t \int_0^t \phi(t)\, dt \tag{4.1}$$

at the origin. Such values were found to occur 200 s after the release of the contaminant and were used to superimpose the data from Trials 13 to 15, 18 and 19. A reflection across the line $X = 400$ m was used to obtain the lateral

symmetry implicit in the definition of 3(ii). The final result was thus as illustrated in Fig. 3.3.

After sensible truncation the data set available for the estimation of $\bar{C}$ and $\overline{c^2}$ was,

$$\Delta = \{C(\boldsymbol{x}_i, t_j): i=1, \ldots 318; j=1, \ldots 416\}: \boldsymbol{x} = (x, y, 0.4\ \mathrm{m}) \tag{4.2}$$

because $C(\boldsymbol{x}, t)$ was observed to be a rapidly decreasing function of z. Thus approximately 250 s of the record were retained. Of the 318 spatial sampling locations 148 (including lateral symmetry) detected gas while the remainder did not.

Estimates of $\bar{C}(\boldsymbol{x}, t)$ etc. may thus be obtained from eqn. (4.2) using the methods described earlier. There remain however some theoretical difficulties associated with such an analysis that make it desirable to prefer a slightly modified procedure. It is apparent to begin with that there is a much greater profusion of temporal data than occurs in either of the two spatial directions. As seriously, it is implicit in the construction of 3(ii) that the spatial data sample displays a measure of point to point statistical independence while with respect to time the record is very strongly correlated. The data set Δ is thus not directionally isotropic.

To take account of the first of these difficulties it is only necessary to introduce scaling parameters $\beta_j, j=1, \ldots 4$ as earlier, i.e. eqn. (2.4) and carry out the analysis with this in mind. The second however is less easy to deal with. The approach employed here, which cannot be justified at present except on intuitive grounds, was to partition Δ into τ distinct subsets with respect to time i.e.,

$$\Delta = \{\Delta_1 \cup \Delta_2 \cup \ldots \Delta_p \cup \ldots \cup \Delta\tau\} \tag{4.3}$$

$$\Delta_p = \{C(\boldsymbol{x}_i, t_{(j-1)\tau+p}): i=1, \ldots 318; j=1, \ldots 416/\tau\} \tag{4.4}$$

Then choosing τ so that the Δ_p are relatively statistically independent as analysis can be applied to each of the Δ_p in turn and a final estimate obtained as the average of these values. This suggests new definitions,

$$m^*(\mathbf{x}{:}\Delta) = 1/\tau \sum_{p=1}^{\tau} m(\mathbf{x}{:}\Delta_p) \tag{4.5}$$

$$M^*(\mathbf{x}{:}\Delta) = 1/\tau \sum_{p=1}^{\tau} M(\mathbf{x}{:}\Delta_p) \tag{4.6}$$

$$\hat{c}^2 = \{m^*(\boldsymbol{y}^2) - m^*(\boldsymbol{y})^2\} \tag{4.7}$$

which were in fact employed in preference to the simpler procedure of Section 2.

Consider now the estimation of $\bar{C}(\boldsymbol{x}, t)$ from (4.2). A specific form for the

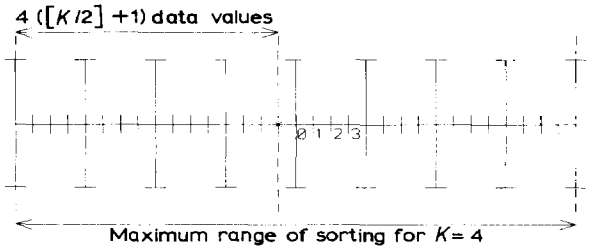

Fig. 4.1. Sorting nearest neighbours from an array along a line.

basic estimator in eqn. (4.5) must first be chosen from the class of alternatives 2(i)–(iii). The irregular distribution of the spatial data points makes it sensible to prefer a definition which is data adaptive i.e. either of 2(ii) or (iii). The first of these alternatives – the method of nearest neighbours – was adopted because it is computationally the least complicated. It was thus necessary to construct the set of K-th nearest neighbours for the set of points comprising eqn. (4.2). The optimum values of K and α were chosen by minimisation of the Cross Validation function eqn. (2.11).

The construction of the set of K-th nearest neighbours to a point for a data sample like eqn. (4.2) in fact involves surprising little computational penalty. A valuable pre-processing of the data set for this purpose has been described by Friedman, Basket and Schuster [18]. This may be combined with any one of the rapid sorting procedures available [19] to yield an efficient algorithm for this problem.

To minimise computer storage requirements however a different approach was implemented which takes account of the homogeneity of the data set with respect to time. Careful thought shows that the set of K-th nearest neighbours to any fixed time subset of eqn. (4.2) is the same for all time $0.6([\kappa/2]+1) \leqslant t \leqslant 0.6(416-[\kappa/2])$. Hence the size of the sorting problem is never greater than $318([\kappa/2]+1)$ values and there are only $318[(\kappa/2)+1]$ such calculations to perform. The full calculation is of course slightly more complicated since the set of nearest neighbours is required for all 416 time points for each of the subsets Δ_p, $p=1, \ldots \tau$. This problem is illustrated in Fig. 4.1 for the case $\tau=4$, $K=7$. Evidently it is necessary to sort $318(2\kappa+1)$ data values (at most) on $\tau([\kappa/2]+1)$ occasions for half of the 318 spatial observation sites (because of lateral symmetry). Fortunately this is only necessary for one of the partitions of (4.2) because of symmetry. As only small values of τ and K are likely to be encountered in practice this approach is not substantially less efficient than that of Friedman, Basket and Schuster [18]. A simple pass sorting method [19] was considered satisfactory as a consequence.

A particular advantage of the method of nearest neighbours, subject to the use of a Kernel function like eqn. (2.10), is that it allows an efficient construction of the regression function eqn. (4.5). To see this it is only necessary to

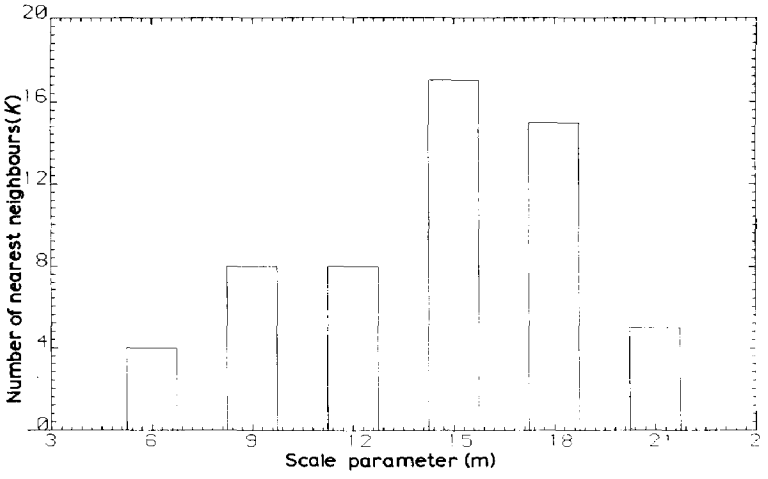

Fig. 5.1. Value of K, at the minimum of Cross Validation, for various values of the scale parameter.

note that by definition $K(\mathbf{x}/H_\kappa)=0$: $\forall\,|\mathbf{x}|>H_\kappa$. Hence the operation of sorting nearest neighbours automatically prescribes the data sample for use in eqn. (4.5). No more than $(2\kappa+1)$ additions need to be performed as a result to obtain each estimate of $\bar{C}(\boldsymbol{x}, t)$. This would not be the case if a kernel with unbounded support were to be used in place of eqn. (4.10) or if another method (say the variable Kernel estimator 2(iii)) was employed instead.

It remains only to discuss the choice of optimum parameters for the procedure and the estimation of $\overline{c^2}$ from eqn. (4.2). In principle this can be achieved by the minimisation of the Cross Validation function but in practice a satisfactory optimisation is not possible because of the size of the data set involved. Hence it was necessary to compromise and place emphasis on obtaining the optimum value of the number of nearest neighbours K. The dimensionality of the problem was reduced by choosing $\beta_1=\beta_2$ and β_3 (clearly not a critical parameter in this analysis) was chosen by empirical examination of the data set to be 2.0 m. A value of $\tau=4$ was chosen to minimise the computational problems associated with the construction problem. Hence it was necessary to determine only one additional parameter $\delta=(\beta_1{:}\beta_4)$ m/s since the operation of sorting nearest neighbours involves only relative proportions of lengths. This value was chosen by seeking to minimise the Cross Validation function in such a way as to maximise K at the minimum. Estimates of $\overline{c^2}$ were obtained via eqn. (4.7) using the same optimum parameters.

5. The results of the regression analysis: estimates of the mean and variance

Fixing β_4 at 2.4 s, β_1 was allowed to vary in the range [6, 21] m increments

of 3 m. The number of nearest neighbours K was allowed to vary up to a maximum of 20. Realisations of the Cross Validation function were constructed for each of these cases and the results are as summarised in Fig. 5.1. Evidently $K=17$, $\delta=6.25$ m/s is a satisfactory choice of parameters.

These values of K and δ were used to obtain estimates of $\bar{C}(\boldsymbol{x}, t)$, $\overline{c^2}(\boldsymbol{x}, t)$ at all the points comprising (4.2) using eqns. (4.5–4.7). In view of the limitations of the optimisation procedure however it must be emphasised that these should be regarded as purely numerical experiments at the present time. Nonetheless the results are encouraging since they display a smooth continuous dependence on both the position and time unlike, for example, 'box' models. This is all the more remarkable since no parametric assumptions about dense gas dispersion are involved in this analysis.

5.1 Estimates of the mean concentration

It is interesting to compare the estimates of $\bar{C}(\boldsymbol{x}, t)$ with the data records associated with each of the 74 distinct spatial observation sites. Ten of these records are presented in Fig. 5.2 for the most part for points near to the origin of the release. A wide variety of structures are apparent though there are several features in common. Hence the estimates obtained are invariably smooth and the peak concentration value is closely comparable with that of the data record. The instantaneous discrepancy between the estimated and realised concentration however is frequently large and sometimes disturbingly so (i.e. notably Figs. 5.2(iv) and (ix)). This is indicative of substantial variability associated with the dispersion of dense gases as will be discussed further later on this paper. What is particularly striking though is the comparison between the estimated mean concentration and that characteristic of box models. Rather than the simple monotonic decay normally associated with the latter the time evolution of $\bar{C}(\boldsymbol{x}, t)$ is much more like that of a dissipative solitary wave. While it is not possible to give a theoretical justification for this observation at the present time it should be borne in mind for future research.

Several contour maps of the spatial variation of $\bar{C}(\boldsymbol{x}, t)$ were prepared to

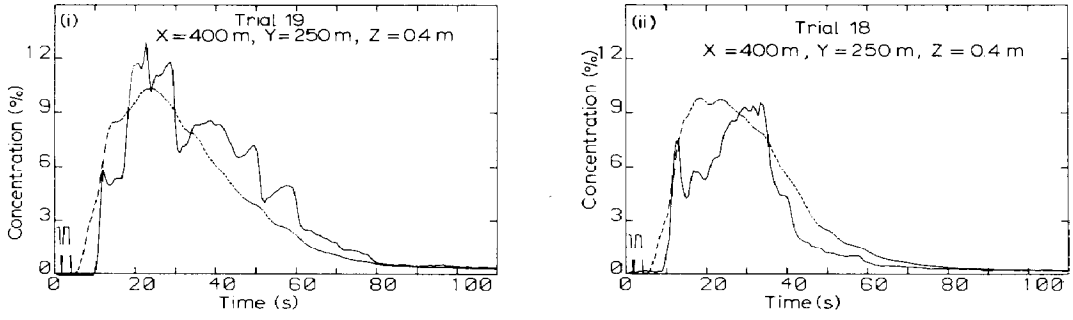

Fig. 5.2. Estimates of the mean concentration: Time record for Trials 13–15, 18 and 19; —— data, — — — estimate.

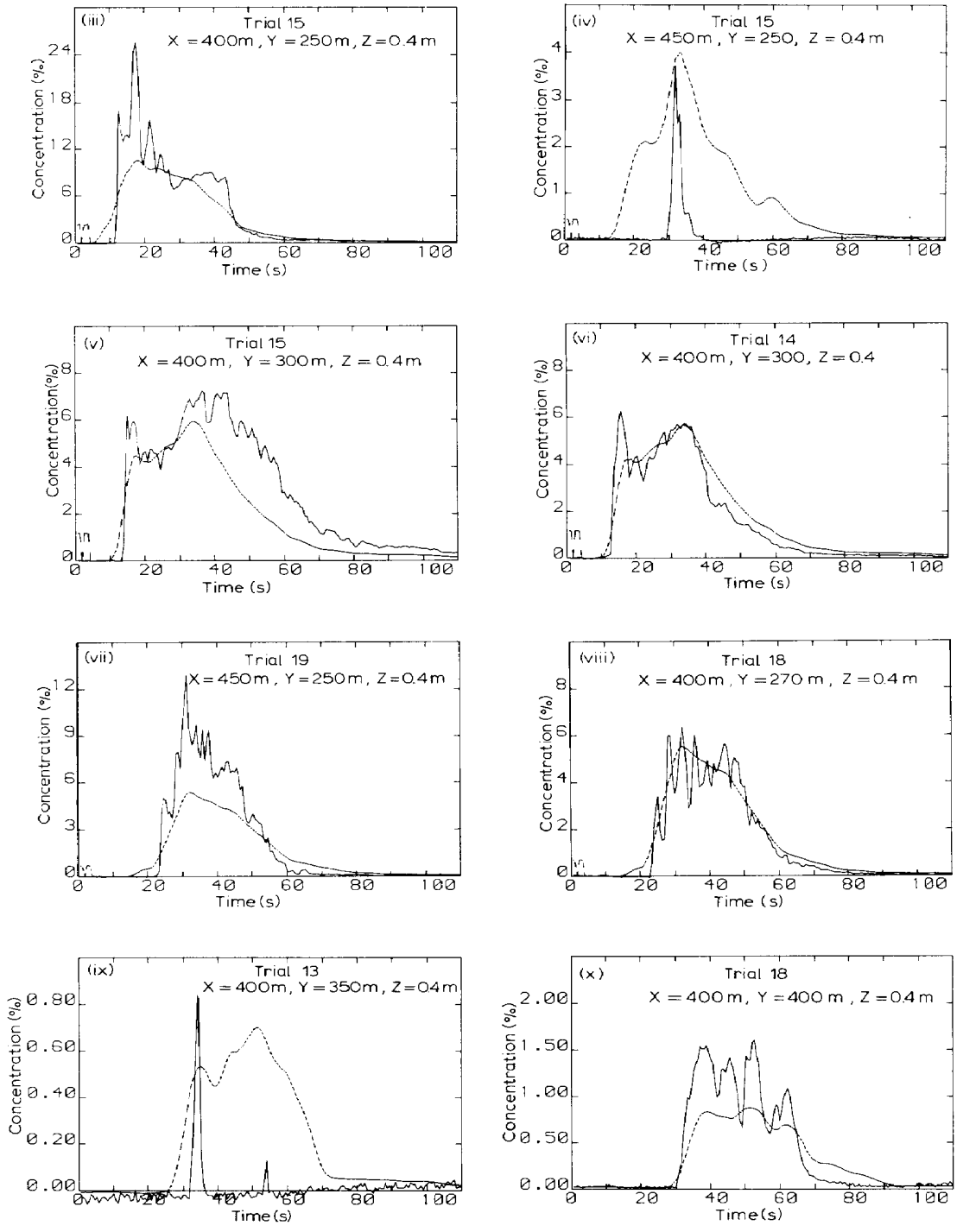

Fig. 5.2 (continued).

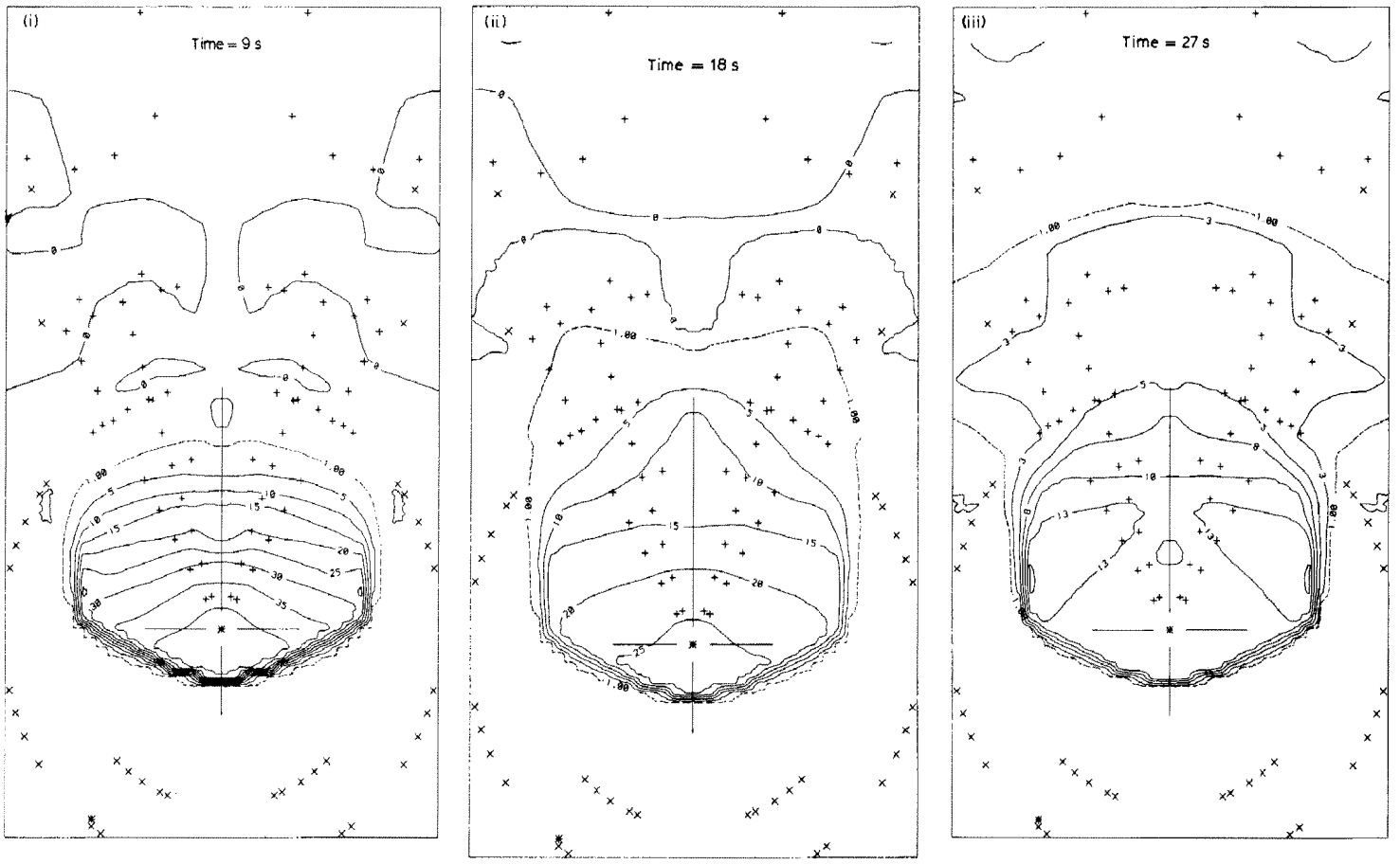

Fig. 5.3. Contour diagrams of the mean concentration: Standard estimate; * origin of release.

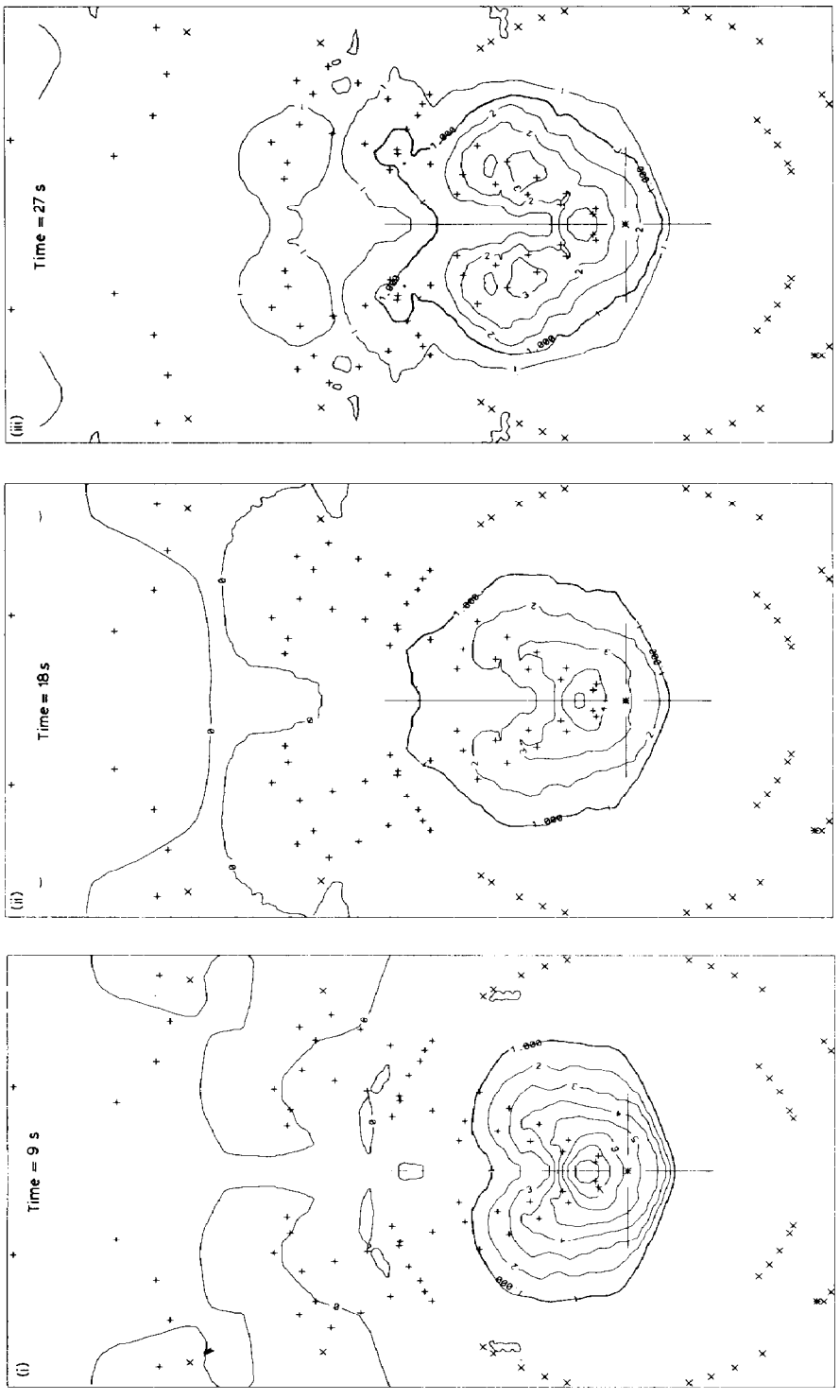

Fig. 5.4. Contour diagrams of the mean concentration: modified estimate; * origin of release.

assess the applicability of this analysis to the delineation of hazardous boundaries in dense gas dispersion. These are illustrated in Fig. 5.3 and the downwind trend in the mean concentration is certainly well indicated. Overall though these estimates are rather poor as they suffer the absence of sufficient values of $C(\boldsymbol{x}, t)$ near the origin and the invalid assumption $\beta_1 = \beta_2$. To see whether the form eqn. (4.6) produces any improvements in this respect similar estimates of $\bar{C}(\boldsymbol{x}, t)$ were produced using definition 2(i) and are displayed in Fig. 5.4. The improvement is suggestive and justifies further investigation of the estimator for this purpose.

5.2 Estimates of the concentration variance

Estimates of the concentration variance were prepared at each of the 74 spatial sitres in the data set eqn. (4.2). The square root variance is plotted against the instantaneous fluctuation in Fig. 5.5 for the same 10 points considered above. Again several features of these records are apparent. Thus the estimated fluctuation is significant in comparison with the estimated mean concentration and is also more persistent in duration. The estimates are smooth though considerable structure is apparent.

The most striking feature of these records is a characteristic double peak in the evolution of $\sqrt{c^2}$ with time. This is shown clearly in Fig. 5.5(i) and (ii) for points near to the origin of the release where the first peak is the more distinctive. Further out this diminishes in comparison with the second and a broad depression develops in the centre of the record. Eventually this structure is lost far out from the origin.

It is interesting to speculate on the origins of this behaviour. Thus one possible reason for the occurrence of the first peak is that it reflects the presence of real variations within the cloud. A prominent annular structure to the instantaneous cloud has been observed from aerial photographs in the early stages of the release and this would certainly contribute to this feature. A possible explanation of the second peak is another change in the internal state of the cloud though this is more difficult to identify. As will be shown later something of this kind should be expected on purely statistical grounds.

Some contours of the spatial variation of the square root variance are illustrated in Fig. 5.6. It is interesting to note that these estimates indicate quite substantial variation of $\sqrt{c^2}(\boldsymbol{x}, t)$ in the horizontal plane. Some of this is undoubtedly due to the instantaneous structure of the cloud which, as noted above, is dominated by an annular feature in the early stages of the dispersion. The far field persistence of these estimates is certainly greater on occasions that that of $\bar{C}(\boldsymbol{x}, t)$.

To illustrate the points raised in the above discussions it is useful to consider a simple stochastic model of $C(t)$ and to examine the records of $\bar{C}(t)$, $\sqrt{c^2}(t)$ produced by a process of ensemble averaging. When $C(t)$ is dependent on two

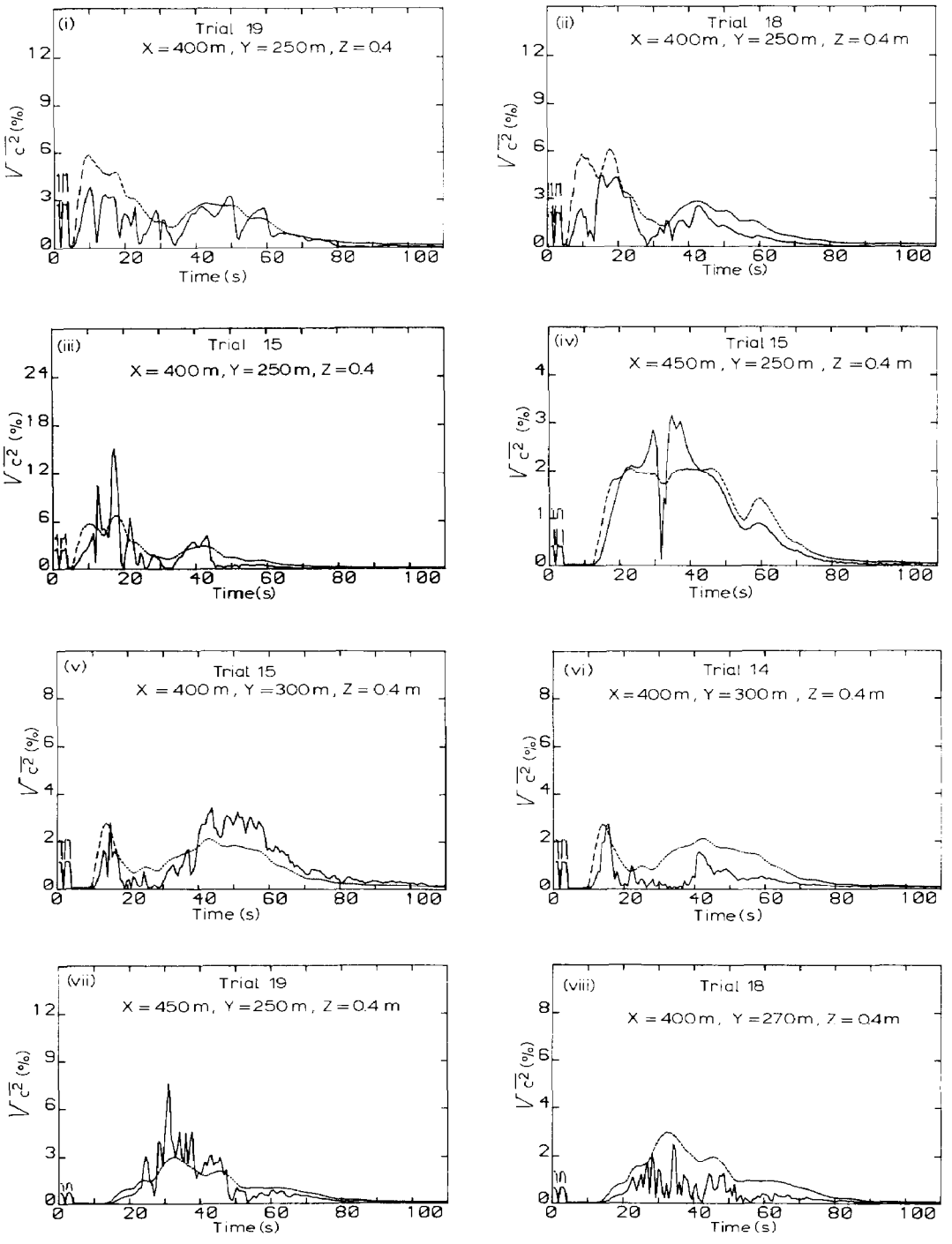

Fig. 5.5. Estimates of the concentration variance: Time record for Trials 13–15, 18 and 19; —— data, — — — estimate.

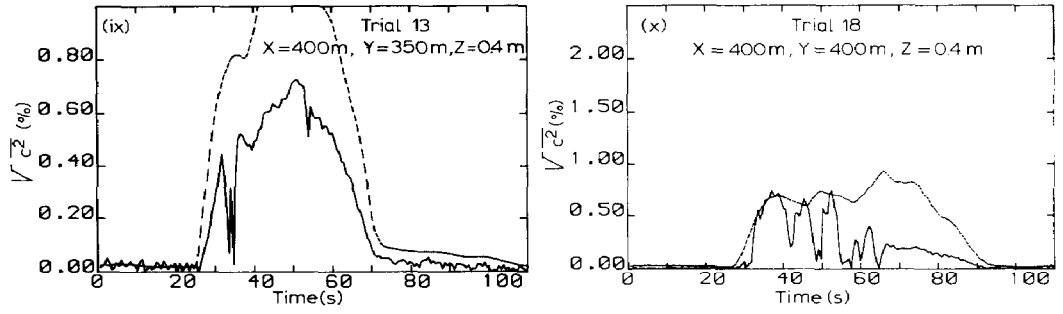

Fig. 5.5 (continued).

random parameters t_α, t_β, jointly distributed according to $p(t_\alpha, t_\beta)$, we have the following expression for the r-th order moments,

$$\overline{C^r}(t) = \int\int C^r(t|t_\alpha, t_\beta) p(t_\alpha, t_\beta)\, dt_\alpha dt_\beta \tag{5.1}$$

It suffices to consider only the simplest possible model: to identify t_α, t_β with the arrival and departure times and to assume;

$$C(t) = \begin{cases} \theta_0 \text{ for all } t_\alpha \leqslant t \leqslant t_\beta \\ 0 \quad \text{for all } t < t_\alpha, t > t_\beta \end{cases} \tag{5.2}$$

Given that t_α, t_β are distributed independently according to $\mu(t)$, $\nu(t)$ respectively we have,

$$p(t_\alpha, t_\beta) = \mu(t_\alpha)\nu(t_\beta - t_\alpha): \quad \begin{cases} t_\alpha \leqslant t_\beta < \infty \\ 0 \leqslant t_\alpha < \infty \end{cases} \tag{5.3}$$

and hence,

$$\bar{C}(t) = \theta_0 \int_0^t \int_t^\alpha \mu(t_\alpha)\nu(t_\beta - t_\alpha)\, dt_\beta\, dt_\alpha \tag{5.4}$$

$$\overline{C^2}(t) = \theta_0 \bar{C}(t) \tag{5.5}$$

To obtain specific results from the above we have to choose particular forms for the probability densities $\mu(t)$, $\nu(t)$. The following definitions lead to particularly simple results,

$$\mu(t) = \begin{cases} (1/a)\ e^{-(t-d)/a} & : t \geqslant d \\ 0 & : t < d \end{cases} : \nu(t) = \begin{cases} 1/C\ e^{-(t-b)/c} & : t \geqslant b \\ 0 & : t < b \end{cases} \tag{5.6}$$

Hence,

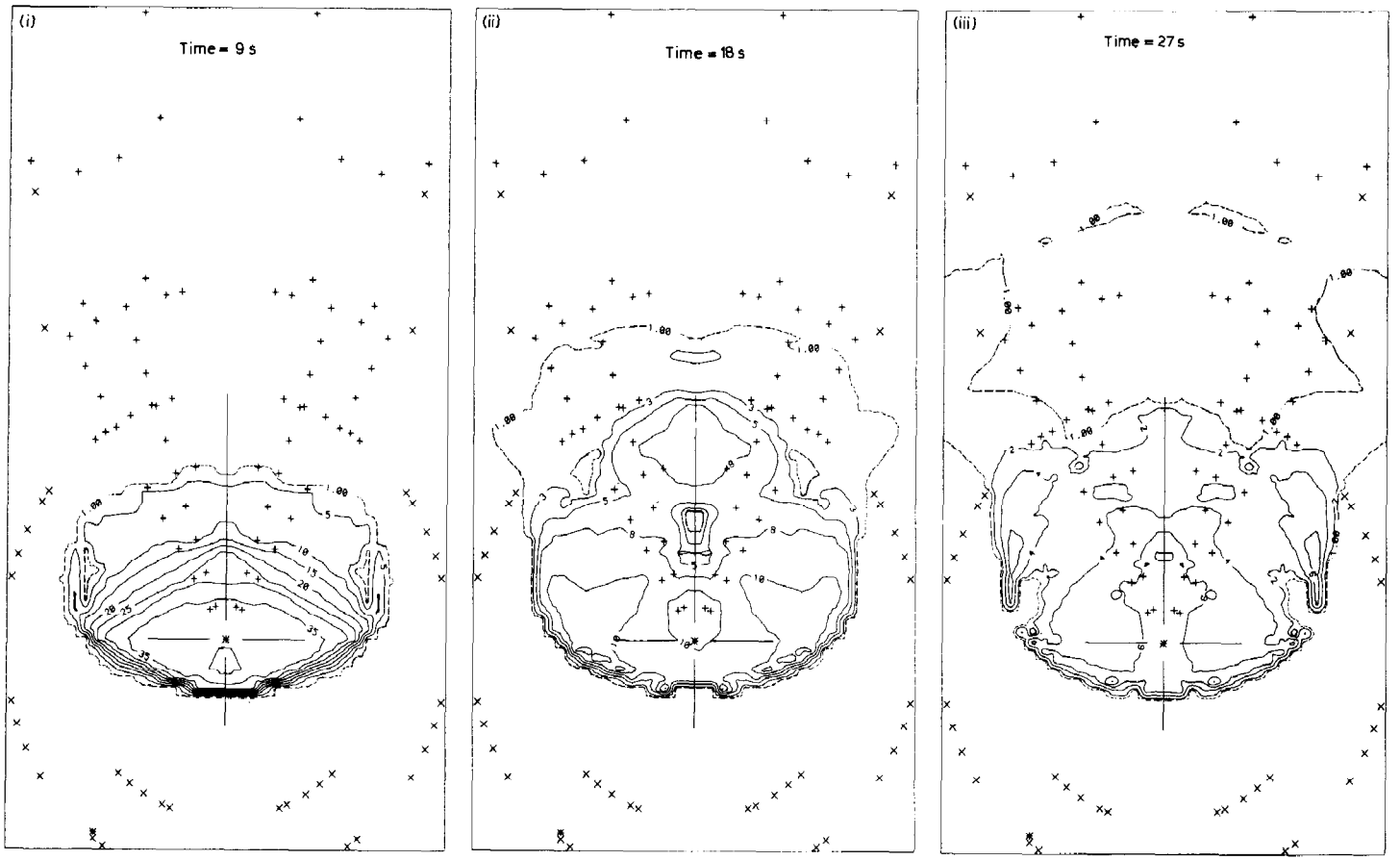

Fig. 5.6. Contour diagrams of the square root variance: * origin of release.

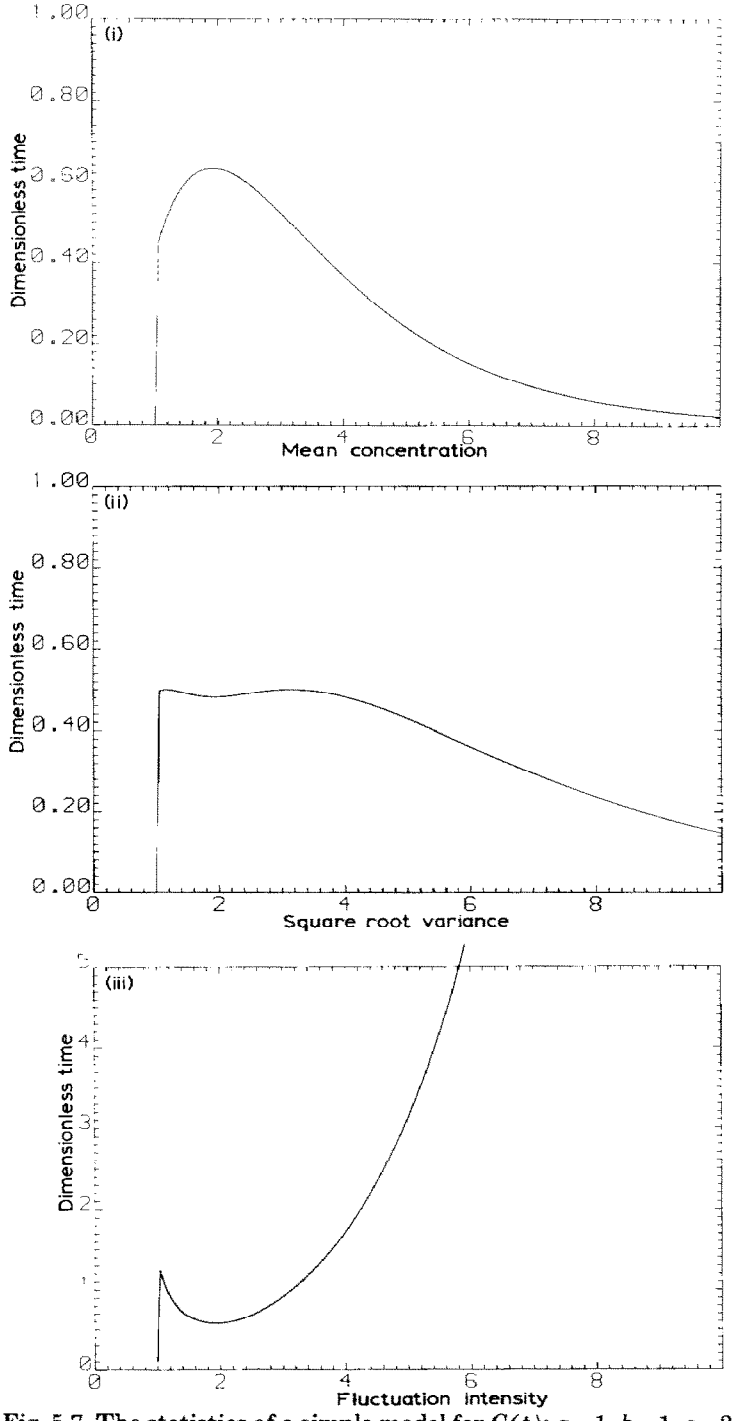

Fig. 5.7. The statistics of a simple model for $C(t)$: $a=1$, $b=1$, $c=2$, $d=1$.

$$\bar{C}(t) = \quad 0 \qquad :t<d \qquad ; \text{else} \tag{5.7}$$

$$\bar{C}(t) = \left\{\frac{e^{d/a}\, e^{b/c}}{ac}\right\}\left\{\int_d^t\int_b^\alpha e^{-t_\alpha/a}e^{-s/c}\,ds\,dt_\alpha - \int_d^{t-b}\int_b^{t-t_\alpha} e^{-t_\alpha/a}e^{-s/c}\,ds\,dt_\alpha\right\}$$

$$= \{e^{t/a}\, e^{d/a}\, (e^{b/a}-1) - \left\{\frac{c}{c-a}\right\} e^{d/a}\, e^{b/c}\, e^{-t/c}\left(e^{-d(c-a)/ac} - e^{(t-b)(c-a)/ac}\right)\Big]$$

Similarly we may calculate $\overline{C^2}$ according to,

$$\overline{c^2}(t) = \{\overline{C^2}(t) - \bar{C}^2(t)\} \tag{5.8}$$

Some illustrations of eqn. (5.7) and of $\sqrt{\overline{c^2}}(t)$, $R(C)$ are included in Fig. 5.7. Evidently these results display many of the features deduced experimentally from the Thorney Island data. Thus Fig. 5.7(i) is characteristic of the decay of $\bar{C}(x, t)$ with time while Fig. 5.7(ii) displays the same double peaked structure observed for $\sqrt{\overline{c^2}}(x, t)$ earlier. Further investigations along these lines are thus clearly to be desired. A priority would be to clarify the model dependence of these properties, particularly in respect of $C(t|t_\alpha, t_\beta)$.

6. Estimates of the variability and conclusions

The description of the variability in dense gas dispersion is ultimately an important aim of investigations like the one described above. Estimates of $R(C)$ may be obtained by simply making use of the estimates of $\bar{C}$, $\overline{c^2}$ in the definition $R(C) = \overline{c^2}/\bar{C}^2$. In this section we shall discuss the results obtained by such a procedure and indicate some possibilities for future research.

The overall magnitude of the variance ratio for this ensemble of the Thorney Island experiments is illustrated in Fig. 6.1 as a function of $\bar{C}$, the ensemble mean concentration. Evidently there is substantial variation in the value of $R(C)$ even at constant $\bar{C}$. For the most part these estimates are small, though spurious, very large, values of $R(C)$ also occur. It is to be noted that a useful bound is provided by the empirical law,

$$R(C) = \epsilon(1/\bar{C} - 1) \tag{6.1}$$

suggested by Chatwin [3] with $\epsilon = 0.04$.

Some idea of the point to point variation of $R(C)$ with time can be obtained by comparing the two sets of estimates of $\bar{C}(x, t)$, $\sqrt{\overline{c^2}}(x, t)$ in Figs. 5.2 and 5.5. The principal features of this estimate are easy to deduce because $C(t)$, $\sqrt{\overline{c^2}}(t)$ are complementary in form. Hence it appears the time dependence of $R(C)$ is characterised by two large peaks one near the beginning and one near the end of the record, where R is upwards of order unity, interspersed by a

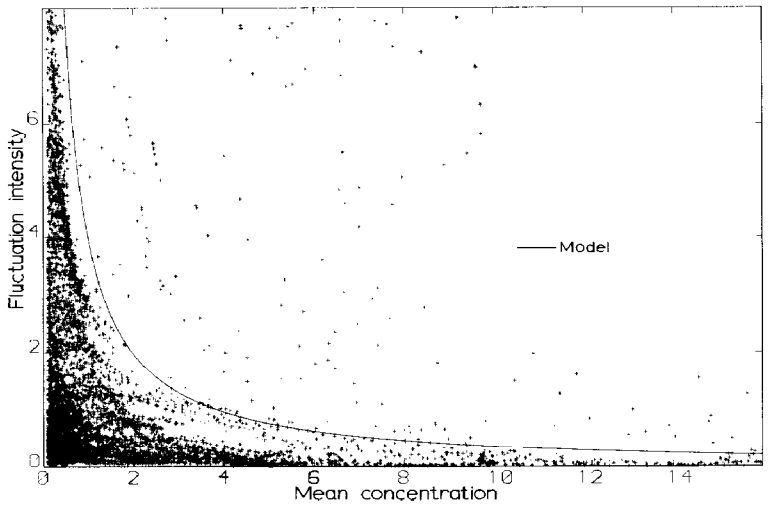

Fig. 6.1. The fluctuation intensity as a function of the mean concentration.

broad trough were $R \simeq 0.1$. It is interesting to note that this structure is apparent even when it is not foreshadowed in the behaviour of $\sqrt{c^2}(t)$, as in Figs. 5.2(vi) and 5.5(vii). Further away from the origin of the release $R(C)$ becomes almost symmetrical with respect to time and the extent of variation across the record decreases, eventually disappearing. Generally speaking $R(C)$ tends to increase with time.

Similar features are observed on comparison of the two sets of contour diagram presented earlier. For the most part inside of a narrow, rather poorly

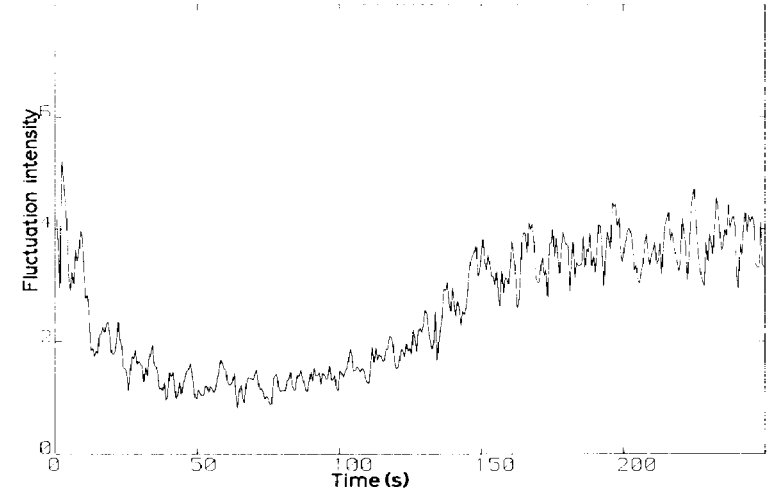

Fig. 6.2. The area averaged fluctuation intensity as a function of time.

defined, zone on the periphery of the cloud $R(C)$ is stable or slowly varying with a value about 0.2. Near the perimeter of the cloud R is larger, being close to unity. The time evolution of this property however is difficult to judge on the basis of such a limited set of estimates.

These observations are interesting because very similar conclusions have been reached by Davies [9]. These are that;

(a) $R(C)$ may be considered constant for a substantial fraction of the total record of $C(t)$,

(b) $R(C)$ may be considered constant over much of the area of the cloud.

In fact a similar conclusion may be deduced by examining the variation of $R(C)$ for the simple step model , eqn. (5.2). Including the other properties mentioned above these observations are indicative of a 'core-bulk' structure to the mean cloud.

This possibility was originally suggested in theoretical work of Chatwin and Sullivan [20] on the dispersion of very small clouds in a field of turbulence. That it should occur here in phenomena on such a large scale is perhaps rather surprising. It is probably a consequence of a broad compatibility with the basic physics of Chatwin and Sullivan's model – namely the advection of a finite cloud in turbulence with a finite scale of variation. Having said this it is important to emphasise that such behaviour has disturbing consequences. Hence the profiles of $\bar{C}(\boldsymbol{x}, t)$, $\overline{c^2}(\boldsymbol{x}, t)$. $R(C)$ can never be self similar, except asymptotically as the 'core' (i.e. the trough in the profile of $R(C)$) decreases to zero with time.

In view of this conclusion it is worthwhile to examine the possibility of defining an area averaged record for $R(C)$. Spurious large values of $R(C)$ abound as by-product of the estimation procedure and present an obstruction to this task, additional to the 'core-bulk' structure mentioned above. Nonetheless it was possible to do this by averaging over all the spatial data points at fixed time for which $0 \leqslant R(C) \leqslant 10^2$. The result is presented in Fig. 6.2.

Clearly three distinct regimes of behaviour are apparent. At small time there is an initial, very large, drop in the value of $R(C)$ from a peak about the value of 5, terminating after roughly 30 s with $R(C) \simeq 1$. The second distinctive property is quite a well defined core to the profile of R, apparent for times between 30 and 110 s from the release, where $R(C) \simeq 1$. Finally there is a transition from this value of the fluctuation intensity to a new plateau about $R(C) = 4$ that occurs roughly 160 s after the release. It is interesting to speculate on the origins of this structure to $R(C)$.

The first of these features coincides with that stage in the dispersion of a gas cloud when buoyancy forces are at their most significant. Hence the rapid fall in the value of R may well be one consequence of this influence since buoyancy is well known to dissipate $\overline{c^2}$ [3]. On the other hand this feature may well be due to deficiencies in the data set since very few observations of $C(\boldsymbol{x}, t)$ are available for $t < 30$ s. The core structure to $R(C)$ probably corresponds to a self

similar phase in the dispersion of the gas cloud. This is interesting because, if this is the case, a box model for $\bar{C}$ and $\overline{c^2}$ may well be appropriate for $30 < t < 110$ s. For times greater than about 110 s the dispersion is no-longer associated with significant buoyancy forces since $\bar{C}$ is then of the order of a few per cent. This may correspond to the final, passive stage of the dispersion.

An important conclusion to this discussion is that the variability for the ensemble of Thorney Island experiments considered in this paper is bound to be large because it is characterised by values of $R(C)$ of order unity. In view of the significance this has for the assessment of hazards in dense gas dispersion it is important to emphasise the role of the basic physical processes responsible for this phenomenon.

Concentration fluctuations in absolute diffusion are generated by two sources:

(i) by the occurrence of concentration variations within the body of the cloud;

(ii) by variations in the rate of advection of the cloud by the external flow field.

In the ensemble of experiments studied here some amount of physical variation between the component trials is also implicit.

Every indication points to the second of the above alternatives, i.e. meandering, as being the main source of variability in dense gas dispersion. This is particularly apparent in the early stages of the dispersion when the advection of the narrow, annular concentration profile generates fluctuations over a wide area. Later in the dispersion this process is less significant because the cloud is larger and less well defined. Nevertheless fluctuations due to long term variations (trends) in the rate of advection may still generate substantial variability by the same process of accretion over time. It is thus necessary to study this phenomenon with respect to a precise definition of the statistical properties of the frame of reference, i.e. the ensemble.

Several alternatives to the ensemble discussed in this investigation have already been indicated in Section 3 of this paper. One of these in particular should be singled out for future study. This is ensemble 3(ii) which, because of radially symmetry, is most suitable for a comparison with the only practical, physically based, model of dense gas dispersion – the box model. There are also computational advantages in an analysis from this perspective. Thus there is considerable justification for such an investigation, preferably using a nonlinear smoother of the form proposed in eqn. (2.16). Though only of indirect relevance to the practical problem of hazard assessment this would prove particularly useful for theoretical purposes and also in bounding $R(C)$ from above.

Acknowledgements

The work described within has been carried out under an S.E.R.C., C.A.S.E. studentship awarded in collaboration with the U.K. Health and Safety Exec-

utive. I would like to thank S.E.R.C. and the officials of the collaborating body, particularly, Dr. J. McQuaid, Mr. A. Mercer and Dr. C. Nussey. I am also grateful to Dr. J.K.W. Davies for an advance copy of his paper [9] and for much useful advice on nonparametric statistics.

Finally I would like to thank my supervisors, Professors P.C. Chatwin and G.D. Crapper, for their support and encouragement during the preparation of this paper.

References

1 A.D. Birch, D.R. Brown and M.G. Dodson, Ignition probabilities in turbulent mixing flows, Report No. M.R.S.E. 374, Midlands Research Station, British Gas, Solihull, 1980.
2 R.F. Griffith and L.C. Megson, The effect of uncertainties in human toxic response on hazard range estimation for ammonia and chlorine, Atmos. Environ., 18(6) (1984) 1195–1206.
3 P.C. Chatwin, The use of statistics in predicting the effects of dispersing gas clouds, J. Hazardous Materials, 6 (1982) 213–230.
4 J.L. Lumley and H.A. Panofsky, The Structure of Atmospheric Turbulence, Interscience Press, 1964.
5 P.C. Chatwin, The statistical description of the dispersion of heavy gas clouds, Contract 1189/01./01, Health and Safety Executive, Sheffield S3 7HQ, 1980.
6 K.K. Carn and P.C. Chatwin, Variability and heavy gas dispersion, J. Hazardous Materials, 11 (1985) 281-300.
7 C. Wheatley and D.M. Webber, Aspects of the dispersion of denser than air vapours relevant to gas cloud explosions, Report No. EUR 9592 EN, Commission of the European Communities (DG XII), Brussels, 1984.
8 K.K. Carn, S.J. Sherrel and P.C. Chatwin, Analysis of Thorney Island data: Variability and box models, In: J.R. Puttock (Ed.), Proc. I.M.A. Conference on Stably Stratified Flow and Dense Gas Dispersion, Chester, April 1986, Oxford University Press, 1987.
9 J.K.W. Davies, A comparison between the variability exhibited in small scale experiments and in the Thorney Island Phase I trials, J. Hazardous Materials, 16 (1987) 339–356.
10 M. Kendall and A. Stuart, The advanced Theory of Statistics, Vol. 2, C. Griffin & Co., 1977.
11 A.P. van Ulden, On the spreading of a heavy gas released near the ground, In: Proc. 1st International Symposium on Loss Prevention and Safety Promotion in the Process Industries, Delft, The Netherlands, 1974.
12 B.L.S. Prakasa Rao, Nonparametric Functional Estimation, Academic Press, 1983.
13 K.K. Carn, Variability and other aspects of dense gas dispersion, unpublished Ph.D. Thesis, Liverpool University, 1986.
14 T. Gasser and H.G. Muller, Kernel estimation of regression functions; in smoothing techniques for curve estimation, Springer Lecture Notes in Mathematics, No. 757, 1979.
15 J. Sacks and D. Ylvisaker, Assymptotically optimum kernel for density estimation at a point, Ann. Stat., 9 (1981) 334–346.
16 J. Rice, Bandwidth choice for non-parametric regression, Ann. Stat., 12(4) (1984) 1215–1230.
17 J. McQuaid and B. Roebuck, Large scale field trials on dense vapour dispersion, Report No. EUR 10023, Commission of the European Communities, Brussels, 1985.
18 J.H. Friedman, F. Baskett and L.J. Schuster, An algorithm for finding nearest neighbours in logarithmic time, IEEE Trans. Comput. (October 1975) 1000.
19 D.E. Knuth, The Art of Computer Engineering, Vol. 3, Sorting and Searching, Addison Wesley, 1973.
20 P.C. Chatwin and P.J. Sullivan, The relative diffusion of a cloud of passive contaminant in incompressible turbulent flow, J. Fluid Mech., 91 (1984) 337–352.