A POD model for the concentration fluctuations of gases instantaneously released in the atmosphere

Anastasios ANTYPAS (corresponding author)

Tel. and fax: (210)8991372, e-mail: t_antypas@yahoo.com

Address: 4, Pringipos Petrou St. 16673 Voula, Attiki, Greece

John BARTZIS

Address: University of West Macedonia, Dept. of Mechanical Engineering,

Salvera & Bakola St., 50100 Kozani, Greece

Abstract: In this work, the concentration signal of repeated instantaneous releases of hazardous gases into the atmosphere is decomposed into its principal components by means of the Proper Orthogonal Decomposition (POD) with the dual purpose of reconstructing it from its first (most energetic) components and predicting it in cases sharing common features with the cases from which those features were extracted by applying the POD. The analysis showed that a model can be constructed by interpolation in the range of varied parameters. The important issue of the uncertainty associated with the proposed POD model is discussed and a measure of uncertainty is provided.

Keywords: concentration, dispersion, prediction.

1. Introduction

In cases where a polluting substance is released into the environment, the need for assessing the consequences of its release arises. The question of the extent of the impact caused may be posed by governmental agencies or private operations administrations, relating the extent of the impact to regulatory directives or legal actions or both. In particular, referring to the case of hazardous (toxic or flammable) gases being released into the atmosphere, in order to quantify the environmental impact of such releases, mathematical models are constructed and employed that describe the physical process of dispersion of the released substances into the atmosphere, as well as the chemical processes involved.

Such models have been introduced decades ago and since then are being continuously improved as research carries on. The aim is to improve the accuracy of the model performance by taking into account the factors that affect it. Such factors include the fundamental physics and chemistry associated with the phenomenon of the dispersion of the released substance, as well as the mathematical and statistical techniques applied to describe the phenomenon.

1.1. Physical and mathematical - statistical background

In this study no chemical reactions will be considered, so the concentration of the pollutant in question will be treated as a passive scalar. In such a case, the concentration $\Gamma(\boldsymbol{x},t)$ at position $\boldsymbol{x}$ and time t

is determined by two processes: advection by a turbulent flow with velocity field Y(x,t), and molecular diffusion with constant diffusivity k. Γ obeys the equation

$$\frac{\partial \Gamma}{\partial t} + \succ \nabla \Gamma = k \nabla^2 \Gamma \qquad (1)$$

where Y satisfies the Navier-Stokes equations with the incompressibility condition $\nabla .\mathrm{Y} = 0$ (Monin & Yaglom vol.1, 1971).

Since all the flows encountered in the real world are turbulent, Y and hence Γ, are random (or *stochastic*) fields. That means that their values at a given location x and at a given time instant t cannot be predicted and vary from realisation to realisation, in antithesis to predictable quantities like time or space averages or *rms* concentration values often considered in practical applications. Hence, a probabilistic approach should be followed.

The probability density function (PDF) of the concentration field $\Gamma(\boldsymbol{x},t)$, denoted by $p(\theta;\boldsymbol{x},t)$, is defined by

$$p(\theta;\boldsymbol{x},t)=\frac{d}{d\theta}\{\mathrm{P}[\Gamma(\boldsymbol{x},t)\leq\theta]\} \qquad (2)$$

for all $\theta \geq 0$. The variable θ ranges over all the possible values $\Gamma(\boldsymbol{x},t)$ can take.

$\mathrm{P}[\Gamma(\boldsymbol{x},t) \leq \theta]$, the probability that $\Gamma(\boldsymbol{x},t) \leq \theta$ is the proportion of realisations forming the *ensemble* of realisations for which $\Gamma(\boldsymbol{x},t) \leq \theta$. Being a probability measure, $p(\theta;\boldsymbol{x},t)$ satisfies

$$\int_0^\infty p(\theta;\boldsymbol{x},t)\mathrm{d}\theta = 1 \qquad (3)$$

Then, the expected value (or ensemble mean) and the variance of $\Gamma(\boldsymbol{x},t)$ are defined in the usual way, by

$$\mu(\boldsymbol{x},t)=\mathrm{E}[\Gamma(\boldsymbol{x},t)]=\int_0^\infty \theta p(\theta;\boldsymbol{x},t)\mathrm{d}\theta \qquad (4)$$

and

$$\sigma^2(\boldsymbol{x},t) = \mathrm{var}[\Gamma(\boldsymbol{x},t)] = \int_0^\infty [\theta - \mu(\boldsymbol{x},t)]^2 p(\theta;\boldsymbol{x},t)\mathrm{d}\theta =$$

$$\int_0^\infty \theta^2 p(\theta;\boldsymbol{x},t)\mathrm{d}\theta - [\mu(\boldsymbol{x},t)]^2 \qquad (5)$$

respectively. The square root of the variance denoted by $\sigma(\boldsymbol{x},t)$ is the standard deviation of the concentration, otherwise known as the *rms concentration fluctuation.* Higher central moments are defined in analogy to eq.(4) with a suitable choice of the exponent. The time evolution equations of the PDF, mean and variance of concentration are given by Chatwin (1990).

The true values of these parameters are unknown, therefore in practice, in order to estimate them one must draw a random sample from the *ensemble* of realisations and compute statistics to be used as estimators of the aforementioned parameters.

1.1.1. The mean concentration

From a sample of n realisations $\Gamma_i(\boldsymbol{x},t)$ one may compute the sample mean,

$$\overline{\Gamma}_n(\mathrm{x},\mathrm{t}) = \frac{1}{n}\sum_{i=1}^{n}\Gamma\mathrm{i}(\mathrm{x},\mathrm{t}) \qquad (6)$$

defined as the arithmetic mean of the data (the subscript n is often omitted). This is itself a random variable (it varies as the sample is varied with respect to its size as well as with respect to the experimenter's choice of values of the varied parameters). When used as an estimator of $\mu(\boldsymbol{x},t)$, it possesses the property of *unbiasedness,* as well as that of *least variance,* hence being referred to as the *best unbiased estimator* of the mean (Arnold, 1990; p.264). Notably, in the limit as $n \to \infty$, eq. (5) provides

$$\mu(\boldsymbol{x},t)=\lim_{n\to\infty}\overline{\Gamma}_n(\boldsymbol{x},\mathrm{t}) \qquad (7)$$

Often, when data are not available, the mean concentration is modelled with a Gaussian, which provides a reasonably good approximation some time after release (Hanna and Drivas, 1987).

However, although the sample mean suggests itself as an important measure in estimating concentrations of harmful substances dispersing in the atmosphere under realistic conditions, it does not *per se* establish a panacea in that regard. The reason is that a predictive concentration model should also incorporate concentration fluctuations, measured by the concentration *rms* as defined above, since in many cases the standard deviation of concentration is as large as its mean or even larger (Chatwin, 1982).

1.1.2. The variance of concentration

The sample variance of the concentration field, defined by

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}[\Gamma i(x,t) - \overline{\Gamma}_n(x,t)]^2 \qquad (8)$$

is an *unbiased* estimator of the ensemble concentration variance, not possessing though the property of least variance (Arnold, 1990; p. 266). Further, the statistical noise it contains (being a random variable) is likely to be greater than that of the sample mean.

A very important contribution in the direction of modelling the concentration variance in the context of turbulent diffusion was the introduction of the Chatwin – Sullivan collapse theory (Chatwin and Sullivan, 1990), which states that a simple quadratic relationship between the mean and the variance exists, namely

$$\sigma^2 = \beta^2 \mu(\alpha\mu_0 - \mu) \qquad (9)$$

where β and α are scalar non-dimensional parameters and μ_0 is a scale for μ (e.g. its maximum).

The above equation fits well data taken from steady continuous sources for which the statistical properties of the concentration field are stationary. In these cases α and β must both be constants satisfying $1<\alpha$ and $0<\beta<1$. For instantaneous releases though, for which the statistical properties of the concentration field are non-stationary, α, β and μ_0 must be functions of downwind receptor position and time. More on mathematical models for α and β can be found in Sullivan (1990), Moseley (1991), Mole (1995), Mole *et al*. (1997) as well as in Mole *et al.* (2007).

1.2. Stochastic concentration models – motivation for use of Principal Components Analysis

The development of mathematical models predicting, ideally the PDF, or less ambitiously, the mean and variance of the concentration field has been and carries on being a major area of research as already pointed out by Chatwin and Sullivan (1994).

The main reason is that such models are much less costly than most deterministic ones, exhibiting at the same time a very good performance, often superior to that of deterministic ones that need vast computing power (unavailable in the reasonably near future) for the large number of direct numerical simulations required. They also tend to replace Gaussian models which due to the many unrealistic assumptions they include are becoming obsolete. Mole, Chatwin and Sullivan (1993) provided a detailed account on different modelling approaches, as well as methods for assessing model performance.

Hanna and Drivas (1987) also declared emphatically the need for development of concentration fluctuation models in their detailed synopsis, adding the imperative requirement of inclusion of the uncertainty assessment that should accompany the model, without which use of the model should be avoided.

Here, the less ambitious direction will be followed, i.e. the proposed model will restrict itself to describing the variance of the concentration field, with the mean being estimated by eq. (6). The methodology proposed is based on the decomposition of the covariance matrix of the concentration field into its principal components. It has been used to identify and describe the coherent structures of turbulent flows by analysing the velocity field Y into a series of orthogonal functions

(principal components) possessing the property of *optimal convergence* with success (Berkooz *et al.*, 1994; Kevlahan *et al.,* 1994; Holmes *et al.,* 1996). The term "optimal" means that only a few of these functions suffice in order to reproduce the original signal, thus playing a key role in the identification of coherent structures. Following an analogous way of thinking, the method is applied here with the hope of extracting from data sets common features characterising the process of turbulent dispersion of a passive scalar, in particular the concentration field. In such a case it would be possible to reconstruct the concentration field or make predictions based only on a few orthogonal functions whose shape would bear the common features mentioned above.

2. The problem and the proposed methodology

The problem addressed here is stated as follows:

Given as input

(a) the dimensions of a gaseous pollutant source with finite dimensions e.g. the cylinder radius R_0 and the cylinder height L (which is taken to be the length scale of the problem), in the case of a cylindrical source;

(b) the initial gas density $\rho_{gas} \Rightarrow g' = g\Delta\rho/\rho = g(\rho_{gas} - \rho_{air})/\rho_{air}$,

(c) the mean wind speed U at source height,

(d) the mean concentration field;

describe and predict the temporal evolution of the concentration of the instantaneously released gas in the atmosphere, at some position downstream.

The proposed methodology is the Proper Orthogonal Decomposition (POD), or Principal Components Analysis, presented in Holmes *et al*. (1996), the basis of the analysis of the velocity field into its principal components that provides an optimal, low-dimensional model.

2.1. POD of concentration field

Consider an ensemble $\{\Gamma_i(t)\}$ of concentration fields, each being defined in the domain $0 \le t \le T$ and thought of as a point in an infinite dimensional Hilbert space L^2 ($[0,T]$) with inner product

$$(f,g) = \int_0^T f(t)g(t)dt \quad (10)$$

The evolution in time of such a concentration field is, as already mentioned, governed by eq.(1).When the Reynolds number $Re = UL/\nu >> 1$ (U is the mean wind velocity, L is the characteristic length scale and ν the molecular viscosity), the flow is turbulent (Monin & Yaglom vol. 1, 1971), so **Y** and, consequently, Γ_i are random variables.

Hence, a statistical description of the physical phenomenon of turbulent diffusion is necessary, as discussed in the introduction.

To that end, the non-dimensional *concentration fluctuation* is defined as

$$F_i(t) = \frac{\Gamma_i(t)}{\Gamma_0} - \mathrm{E}\left[\frac{\Gamma_i(t)}{\Gamma_0}\right] = C_i(t) - \mathrm{E}[C_i(t)] \quad (11)$$

where Γ_0 is a typical concentration scale so that the value of $C(t)$ is a non-dimensional concentration representing the data set, and $\mathrm{E}[.]$ denotes an "ensemble" average.

Thus $F(t)$ is non-dimensional with $\mathrm{E}[F(t)]=0$. The purpose here is "to find a basis $\{\varphi_j(t)\}_{j=1}^{\infty}$ for L^2 that is optimal for the data set in the sense that finite-dimensional representations of the form

$$F^{(N)}(t) = \sum_{j=1}^{N} a_j \varphi_j(t) \quad (12)$$

describe typical members of the ensemble better than representations of the same dimension in any other basis" (Holmes *et al*., 1996).

In this context, the optimality problem is stated as:

$$find \max E[|(F(t),\varphi(t))|^2]$$

$$subject\ to\ \|\varphi\|^2 = 1$$

where $\|.\|$ is the standard norm in L^2,

$$\|f\| = (f, f)^{\frac{1}{2}} = [\int_0^T f^2(t)dt]^{\frac{1}{2}} \qquad (13)$$

from eq. (10).

Introducing a Lagrange multiplier λ, the functional

$$J[\varphi] = E[|(F,\varphi)|^2] - \lambda(\|\varphi\|^2 - 1) \qquad (14)$$

is constructed and the extremum condition

$$\frac{d}{d\delta} J[\varphi + \delta\psi] \,|_{\delta=0} = 0 \qquad (15)$$

is imposed, where $\varphi + \delta\psi \in L^2([0,T]), \delta \in R$, is a variation of φ.

Solving the extremisation problem it is found that the optimal basis is given by the eigenfunctions $\{\varphi_i\}$ of the covariance of concentration (Holmes *et al.,* 1996). In the finite dimensional case, where the observations $\{C_i\}$ are m-vectors (where m is the number of sampled time points), the eigenvectors $\{\varphi_i\}$ are the principal components of the mxm concentration covariance matrix. Furthermore, the eigenfunctions φ_j are mutually orthogonal in L^2 being basis vectors:

$$\int_0^T \varphi_i(t)\varphi_j(t)dt = 0, i \neq j \qquad (16)$$

Assuming no degeneracy of the eigenvalues, that is:

(i) $\lambda = 0$ is not an eigenvalue;

(ii) no multiple eigenfunctions for a given eigenvalue;

and from $\lambda_j \geq 0$, we may order the eigenvalues:

$$\lambda_j > \lambda_{j+1}. \qquad (17)$$

Then, $F(t)$ may be reproduced by a modal decomposition based on the eigenfunctions $\{\varphi_j(t)\}_{j=1}^{\infty}$:

$$F(t) = \sum_{j=1}^{\infty} a_j \varphi_j(t) \qquad (18)$$

where

$$a_j = \int_0^T F(t)\varphi_j(t)dt \qquad (19)$$

Equation (18) is called the POD, $\{\varphi_j(t)\}$ are the empirical orthogonal functions (EOF), $\{\lambda_j\}$ are the empirical eigenvalues and the (random) constants a_j are the expansion coefficients with the following statistical properties:

$$E[a_j] = \int_0^T E[F(t)]\varphi_j(t)dt = 0 \qquad (20)$$

and

$$\mathrm{cov}[a_i, a_j] = E[a_i a_j] = \lambda_i \delta_{ij} \qquad (21)$$

Thus, the expansion coefficients a_j are uncorrelated and

$$\mathrm{var}[a_j] = \lambda_j$$

Also,

$$\mathrm{var}[F(t)] = \frac{1}{\Gamma_0^2}\mathrm{var}[\Gamma(t)] = \sum_{i=1}^{\infty} \lambda_i \phi_i^2(t) \qquad (22)$$

If eq. (16) can in practice be strengthened to $\lambda_j >> \lambda_{j+1}$, in which case only a few modes will suffice to reproduce the original signal accurately, so that eq. (18) can be replaced by eq. (12), the model will be of great value for describing the data and possibly for making predictions.

Finally, when the domain of $\{\Gamma_i(t)\}$ is not bounded, as in the case of open flows considered here, it will be assumed that $\Gamma_i(t)$ decays rapidly to zero outside some bounded domain $[0,T]$, when no

measurements exist for $t>T$, where T is the length of the record of the experiment, so that the above analysis is still valid.

3. The experiments and the data

The data used here to carry out the POD were collected in the course of the experiments of Hall *et al.* (1991) which modelled the Thorney Island (UK) trials at 1/100 scale. The experiment corresponds to the sudden release of a heavy gas in the neutrally buoyant atmospheric boundary layer over a flat terrain. The apparatus used consisted of a cylinder of *height L=13cm and diameter 14cm that collapsed* upon release of its content, a dense contaminant gas, in a wind tunnel. The Richardson number R_i, characterising the gas, is defined by $R_i = g\Delta\rho L/\rho U^2$, where $\Delta\rho/\rho$ is the relative to air gas density and U the mean air velocity at the top of the cylinder. The values of Ri were 0, 0.5, 1, 2, 5, 10. Values of Ri close to zero indicate a neutrally buoyant gas, while higher values characterise heavier (denser-than-air) gases. 50 releases were made with the two highest values of R_i and 100 with the rest. Four sensors were used to record the gas concentration signal, located at 70 and 200cm along the centre-line, at heights of 0.4 and 2.4cm above the ground. In this paper, the four measuring positions are denoted by X1(70,0,0.4), X2(70,0,2.4), X3(200,0,0.4), X4(200,0,2.4). The near-field sensors at positions X1 and X2 monitored gravitational effects dominating close to the source, while the far-field sensors located at positions X3 and X4 captured atmospheric turbulence effects at a further downwind location where the gas concentration was 2% of the initial concentration upon release. Typical release plots with R_i=0 are provided in Hall et al. (1991).

The POD is applied here to the time series obtained from the experiments of Hall *et al.* described above, with the purpose of investigating the possibility of obtaining a universal model for the description of the data, as well as a predictive tool with varying parameters R_i, downwind distance from the source and height from the ground.

4. Data analysis

4.1. Comments on eigenfunctions

The sequence of resulting eigenvalues λ_i (shown in Table 1) of the concentration covariance matrix appears to satisfy $\lambda_i >> \lambda_{i+1}$ in agreement with the discussion in section 2.1. The first eigenfunction, corresponding to the largest eigenvalue, resembles a typical replication

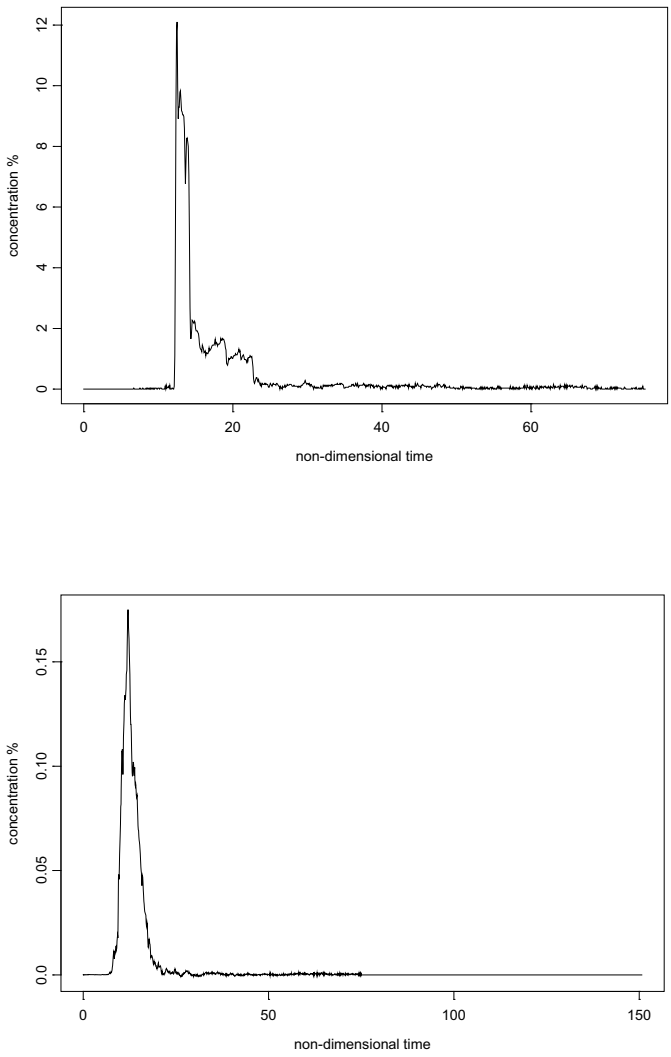

Fig. 1. Typical release (top) and first concentration eigenfunction (bottom) with Ri=0; position X.

(see Figure 1) since it carries the largest variance proportion because

$$\mathrm{var}[F(t)] = \sum_{i=1}^{\infty} \lambda_i \phi_i^2(t) \approx \lambda_1 \phi_1^2(t) \quad (23)$$

from eqs. (12) & (22), and since the plot of concentration standard deviation against time resembles that of mean concentration. Also, the fact that the plot of concentration standard deviation against time resembles that of mean concentration more with decreasing Richardson number, and particularly so for R_i=0 (Robinson, 1996), suggests that it may be possible to obtain a universal POD model for neutrally buoyant gases.

Table 1: Normalised concentration eigenvalues (relative variance proportion) corresponding to the first four eigenfunctions (modes).

R_i	Station	Normalised Eigenvalues			
0	X1	0.438	0.142	0.090	0.066
	X2	0.220	0.150	0.130	0.080
	X3	0.397	0.205	0.074	0.057
	X4	0.213	0.148	0.080	0.070
0.5	X1	0.458	0.159	0.089	0.059
	X2	0.198	0.127	0.085	0.059
	X3	0.481	0.195	0.079	0.043
	X4	0.289	0.129	0.065	0.051
1	X1	0.389	0.152	0.072	0.044
	X2	0.146	0.103	0.079	0.063
	X3	0.516	0.172	0.080	0.044
	X4	0.257	0.121	0.078	0.054
2	X1	0.475	0.171	0.093	0.045
	X2	0.207	0.156	0.083	0.054
	X3	0.609	0.126	0.057	0.027
	X4	0.202	0.076	0.056	0.040
5	X1	0.322	0.178	0.120	0.070
	X2	0.463	0.131	0.082	0.048
	X3	0.504	0.249	0.076	0.033
	X4	0.091	0.074	0.059	0.056
10	X1	0.325	0.149	0.099	0.080
	X2	0.117	0.090	0.067	0.062
	X3	0.550	0.290	0.039	0.025
	X4	0.140	0.066	0.056	0.048

The first eigenfunction tends to zero as time tends to infinity, while the long right tails observed in some cases and dealt with in section 4.2 not only do not represent actual concentration measurements (they should be attributed to instrument noise) but may potentially distort the true shape of the eigenfunction, contributing unwanted fluctuations to consequent calculations.

The first eigenfunction's smoothness decreases with Richardson number at the higher measuring stations consistent with the larger degree of intermittency in those cases.

The initial peak characterising measurements of concentration at the three highest Richardson numbers, 2, 5 and 10, at the measuring stations close to the source is exhibited by the first eigenfunction, displaying the existence of a gravity current head at the edge of these gas clouds.

The first eigenfunction's shape widens as the Richardson number increases closer to the ground, indicating that heavier gases spend a larger proportion of time near the ground, again consistent with the gravity current motion induced by these gases.

The shape of the eigenfunction is not affected when reducing the length of record down to 2/3 of the original length. The eigenvalues are almost the same, with small differences in the 3rd decimal. When the record length was reduced to 1/2 the original length, in one case (R_i =10, X3), the first eigenfunction changed slightly and in another case (R_i=5, X4) even less. In these cases, eigenvalues differ to the second decimal. In all cases, eigenvalues are not affected by truncation of record length, when rounded to one decimal. All of the above provides evidence that the shape of eigenfunctions is invariant under record length truncation when the main part of the cloud has passed, so that further measurements do not contribute. Record length appears to have an effect only when the cloud is truncated.

4.2. Normalisation and scaling

In order to be able to transfer results to another scale, so that a universal model can be obtained, the problem of normalisation and scaling of the first eigenfunctions of concentration covariance is addressed:

By applying an appropriate transformation to the eigenfunctions defined in section 2.1, it is hoped that their shapes will collapse to a common universal shape, thus removing the dependence on varied parameters. Then, by inverting the transformation, it would be possible to reproduce an eigenfunction based on the universal shape obtained, thus enabling one to transfer to another scale. That would lead to a simple model for describing the data and for making predictions. The eigenvectors φ_i ,i=1,…,N of the concentration covariance matrix defined in section 2.1 are normalised as

$$\int_0^T \phi_i^2 dt = 1 \qquad (24)$$

To obtain a common universal shape produced by the first (most energetic) eigenfunctions $\varphi_1(t)$, start by defining

$$\mu_t = \int_0^T t\varphi_1^2(t)dt \Big/ \int_0^T \varphi_1^2(t)dt = \int_0^T t\varphi_1^2(t)dt \quad (25)$$

$$\sigma_t^2 = \int_0^T (t-\mu_t)^2\varphi_1^2(t)dt \Big/ \int_0^T \varphi_1^2(t)dt = \int_0^T (t-\mu_t)^2\varphi_1^2(t)dt \quad (26)$$

$$\psi_1(s) = \sigma_t\phi_1(t) / A \qquad (27)$$

$$s = \frac{t-\mu_t}{\sigma_t} \qquad (28)$$

$$A = \int_0^T \phi_1 dt, \qquad (29)$$

$$\int_{t=0}^{t=T} \psi_1(s)ds = \int_0^T \varphi_1(t)dt / A = 1 \qquad (30)$$

where the last equality in both eqs. (25) and (26) is obtained by virtue of eq. (24).

Then, apply the scale transformations under which all scaled eigenfunctions have area equal to 1: where $dt=\sigma_t ds$ from eq. (28) .

The reason for the unusual definition of μ_t and σ_t^2 in eqs. (25) and (26) is that choosing $\varphi_1^2(t)$ as the weight in the definition of μ_t and σ_t^2, effectively removes the small fluctuations about zero in the long right tail of $\varphi_1(t)$ in experiments which measured much further into the tails, thus resulting in obtaining comparable functions as desired.

The statistics μ_t and σ_t can be interpreted as estimators of the mean signal time and its standard deviation (rms), respectively. In particular, σ_t may be regarded as a measure of the cloud width. Below is provided a physical description of the variation of μ_t and σ_t with Richardson number and spatial position.

a. The variation of μ_t

μ_t gives a central time for the cloud at the measurement position. Therefore this time is the sum of the time the cloud takes to arrive at that point and roughly half the time it takes to pass over that point. The latter contribution will be roughly proportional to σ_t , which measures the spread of the cloud.

Regarding the former contribution, it can be pointed out that comparing values of μ_t with 90%iles of cloud arriving time reported by Hall *et al.* (1991), in principle there is a direct analogy in variation. As Ri increases though, there is a disproportionate decrease of μ_t from the low near to the high near station. Also, for the three heavier gases (R_i=2,5,10) a similar decrease of μ_t is observed from low far to the high far station, which for R_i=5 and 10 is reported for the 90%ile of cloud arriving time too, suggesting that the variation pattern reported is not greatly altered.

Regarding the latter contribution now, as far as variation of μ_t with respect to height is concerned, for the three higher values of R_i (R_i=2,5,10), the cloud spends more time at the lower stations than at the higher ones due to the gravity current-like motion being more pronounced in these cases (heavier gases) and causing velocity shear stress

layers between the cloud and the surrounding air (Hunt *et al.,* 1983).

As R_i decreases, this effect is attenuated and particularly for R_i=0, μ_t is almost the same at the lower and higher stations, because a neutrally buoyant gas will diffuse faster than a heavier one that will settle on the ground due to negative buoyancy.

Concerning now the variation of μ_t with respect to downwind distance, for the three lower values of R_i (R_i=0,0.5,1), the cloud spends about twice as much time at the far stations than at the near stations as a result of the larger travel time to the downwind stations. For the three higher values of R_i (R_i=2,5,10), the time the cloud spends at the lower stations is almost tripled going from near to far stations for the same reason as above, the effect being more pronounced for heavier gases though. Also, for all values of R_i>0, one may explain the larger times spent at the downwind stations as a result of the decoupling of the cloud from the ambient flow and its motion downwind at a rate slower than the ambient wind speed.

At the higher stations now, as the cloud moves from the position close to the source (near field) to that away from the source (far field) the amount of time spent is about four times as much because upward diffusive flux dominates over downward flux produced by the mean sinking motion of the cloud (Hunt *et al.,* 1983).

b. The variation of σ_t

In the case of a neutrally buoyant gas (R_i of 0), from near to far stations σ_t increases (more than double) because spreading from turbulent dispersion as well as gravity current spreading has occurred as it moves downstream. As R_i increases, σ_t increases at the stations near the source, which, acting as near-field monitors, record fluctuations occurring during the gravity slumping phase. Thus, heavier gas clouds exhibit a wider profile in time.

At the far stations, σ_t is also increasing with increasing R_i.

A general (all stations), but not striking, exception in the increasing trend of σ_t with R_i is observed when Ri changes from 0.5 to 1. Referring to the Appendix of Hall *et al.* (1991) though, it is seen that plots of typical repetitions also possess this feature. A possible physical explanation is that these values of R_i represent the threshold of the transition from neutrally buoyant to heavier gas behaviour. A notable exception is the decrease of σ_t at position X2 when R_i is changed from 1 to 2. It should also be pointed out that in that case (R_i=2, X2), the first principal component was swapped with the second principal component (its shape resembled the shape of the second principal components produced by the POD).

4.3. Towards a universal model

Application of the normalisation procedure described above gave the normalised first eigenfunctions, which were then matched. The results are summarised in the following plot in the Figure 2, which depicts clearly the most successfully matched first eigenfunctions.

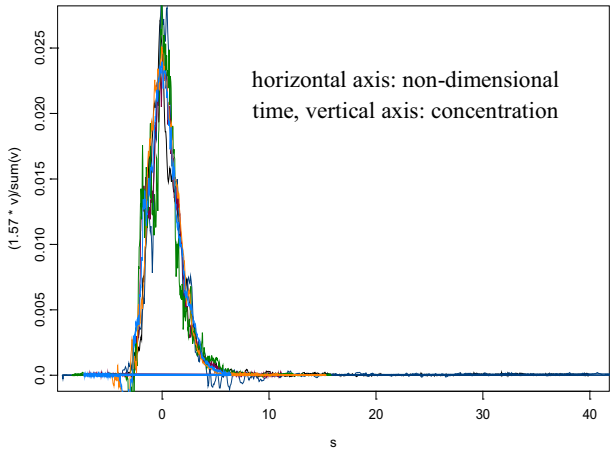

Fig. 2. Matched normalised and scaled first concentration eigenfunctions in cases with R_1=0; all stations and R_1=0.5; lower stations.

Inspection of the collapse of the shape of the first normalised eigenfunctions in Figure 2 suggests that it might be possible to describe the first eigenfunction with R_i=0 at all measuring positions and R_i=0.5 at the two lower positions by a common universal shape, so that one would be able to transfer to another downwind distance or height or Richardson number by inverting the transformations (27) and (28):

$$\phi_1'(\sigma_t s + \mu_t) = \frac{A}{\sigma_t}\psi_1(s) \quad (31)$$

where the left hand side is a predicted first eigenfunction from a given scaled eigenfunction ψ_1.

4.4. Reconstructing the signal

The first eigenfunction (principal component or mode) carries the more coarse features of the signal, since it resembles a typical repetition, the finer details being carried by the next (higher) eigenfunctions.

More specifically, the first eigenfunction, carrying most of the variability represents the largest scale of the fluctuations caused by the biggest eddies. The next eigenfunctions with less and less energy content represent smaller scale fluctuations due to smaller size eddies. According to the "energy cascade" theory (Batchelor, 1953; Tennekes and Lumley, 1972), bigger eddies supply energy to smaller ones in a hierarchical order, until in the final stage of the dispersion process all the energy will have dissipated to heat.

As provided by the theory in section 2.1, the signal can be reconstructed using only a few principal components (eq. (12)) when the sequence of the eigenvalues converges sufficiently rapidly, a condition that is generally satisfied here.

Next, a measure of the accuracy of reconstruction of the concentration signal is provided.

Since the data are conceptualised as vectors in an L^2 space (section 2), it would seem natural to choose an error formula that incorporates the L^2 norm:

The relative error (RE) in the L^2 space is defined as

$$RE_i = \frac{(\sum (C_t - Cr_t)^2)^{1|2}}{(\sum C_t^2)^{1|2}}, i=1,\ldots,n, \quad (32)$$

where C_i is the observed concentration signal and Cr_i the reconstructed concentration signal from the principal components of the covariance matrix at sampling time *t*.

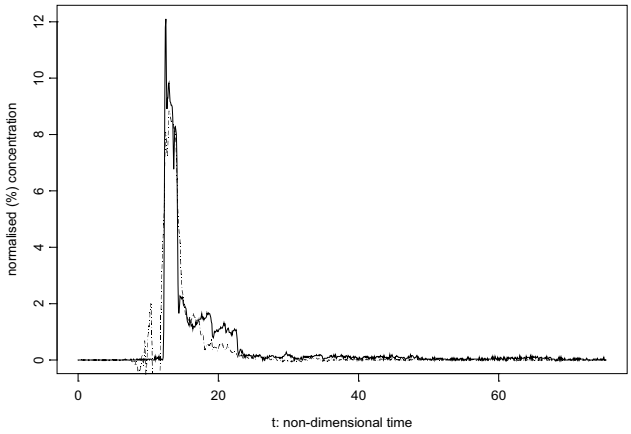

Fig. 3a. Observed (solid line) and reconstructed from first 4 eigenfunctions (dashed line) concentration signal with R_i=0 at position X1. Randomly chosen release.

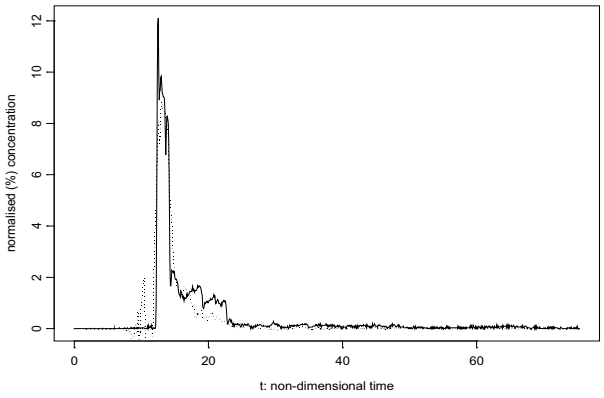

Fig. 3b. Observed (solid line) and reconstructed from first 4 eigenfunctions (dotted line) concentration signal with R_i=0 at position X1. Randomly chosen release.

Table 2: Mean relative error between observed and reconstructed signal using 2 and 4 modes.

R_i	Station	No of modes	MRE
0	X1	2	.41
		4	.34
	X2	2	.50
		4	.42
	X3	2	.32
		4	.26
	X4	2	.48
		4	.43
0.5	X1	2	.33
		4	.27
	X3	2	.60
		4	.50

The mean relative error (MRE) is the sample mean value of RE over all *n* replications:

$$MRE = \frac{1}{n}\sum_{i=1}^{n} RE_i \qquad (33)$$

The reason for choosing RE as the appropriate measure of accuracy of reconstruction of a concentration signal is that from the point of view of the end user of the model, knowing that the MRE is 10 percent means a great deal more than knowing that the mean absolute error is 123 for example (Makridakis *et al.*, 1983, p.47).

Now, from the concentration eigenvalue sequence (see Table 1) is provided an indication that initially the MRE converges relatively fast. This was verified by computing MRE, shown in Table 2, with 2 and 4 modes. When more modes (5 to 12) were included in the reconstruction of the concentration signal, differences in the MRE occurred only in the 3rd decimal place or above, indicating that the rate of convergence of the MRE is much slower beyond the 4th mode.

In the cases where the MRE is small and almost the same (or the same within 1 decimal place) when 2 or 4 modes are used, it is clear that the signal can be reconstructed using the first 2 modes only as discussed also by Kevlahan *et al.* (1994). In those cases the contribution of small scale fluctuations is negligible in the sense that the amount of energy contained in the first 2 eigenfunctions representing the larger scales of motion is sufficient to reproduce the signal fairly well.

The mean relative error (MRE) computed over all replications, between observed and reconstructed signal is shown in Table 2.

In Figure 3, the reconstructed signal from the first four modes is plotted together with the original signal, portraying the good approximation expected from the theory.

4.5. *The distribution of the expansion coefficients*

It is also of interest to discuss the distribution of the coefficients (eq.(19)) in the expansion (18) of the concentration field, and particularly the distribution of a_1 used in the prediction model presented in section 5.

In Figure 4, histograms (with estimated density lines) of the first expansion coefficients are shown in all cases with R_i=0. The sample means are almost zero and the sample variances are almost equal to the respective eigenvalues. The small differences observed between these statistics and the theoretical values of the respective parameters provided by the theory should be attributed to the fact that a finite sample of gas releases was used to produce these results.

Next, the hypothesis that the expansion coefficients come from a normal distribution is tested. The reason behind this particular choice is that normality might be a desirable property related to the practical application of the model, although no such property is prescribed by the theory.

The setting is:
Null hypothesis
H_0:True PDF of coefficients is normal
vs.
Alternative hypothesis
H_1:True PDF of coefficients is not normal.

To carry out this test, an appropriate test statistic and a test criterion must be chosen.

Kolmogorov and Smirnov have developed a normality test in which estimates of the parameters of the normal distribution from the data are used. This test is rather conservative though, in the sense that lower p-values than those computed by the Kolmogorov-Smirnov test can be achieved (Lilliefors, 1967). A more powerful normality test has been proposed by Shapiro and Wilk (Shapiro and Wilk, 1965), based on a variance ratio, that will reject the null hypothesis more often than the more conservative Kolmogorov-Smirnov test.

Carrying out both tests it was found that the more powerful Shapiro-Wilk test rejected the null hypothesis concerning the first expansion coefficient a_1 in the following cases:R_i=0, station X3 (p-value=0.0094); and R_i=0.5, station X1 (p-value=0.0082), while the Kolmogorov-Smirnov test never did.

Further, regarding the question of independence of the first expansion coefficient α_1, Pearson's χ^2 - criterion was employed in order to carry out two hypothesis tests according to the following setting:

(a) Null hypothesis

H_0: α_1 independent with respect to measuring position vs.

Alternative hypothesis H_1: α_1 not independent with respect to measuring position;

(b) Null hypothesis

H0: α_1 independent with respect to Richardson number vs.

Alternative hypothesis H1: α_1 not independent of Richardson number.

In all cases the tests rejected the null hypotheses of independence, thus providing evidence that the first expansion coefficients are not independent with respect to measuring position or Richardson number.

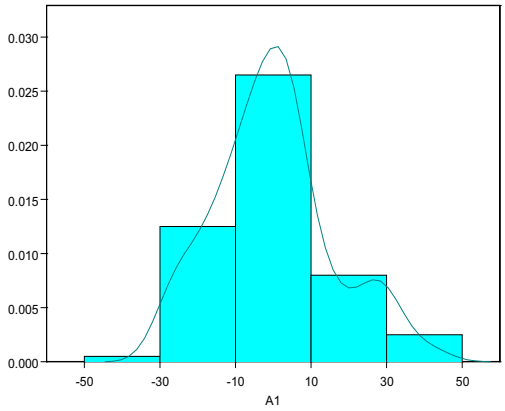

Fig. 4. Histogram with estimated density line of first expansion coefficients α1 for Ri=0 at position X1.

5. Predictive model validation

In this section, a predictive model for the concentration fluctuation field is proposed, based on the results found in the previous sections.

The model is built based on the assumption that the true mean-corrected concentration curve is closely approximated by the first eigenfunction of concentration, given that the first eigenfunction carries most of the variability.

Furthermore, the universal shape of the first eigenfunction obtained in section 4.3 suggests the feasibility of constructing a POD–based prediction model for the case of neutrally buoyant gases (Ri=0 and Ri=0.5).

5.1. The predictive model

The proposed model is

$$F'(t) = a_1 '\phi_1 '(t) \qquad (34)$$

or, from the definition (eq.(11)) of the concentration fluctuation *F(t),*

$$C(t) = \mu(t) + a_1\phi_1(t) + \varepsilon(t) \quad (35)$$

where dashes denote predictors and

$C(t)$ is the concentration field,

$\mu(t)$= E[$C(t)$] is the mean concentration field,

$\varphi_1(t)$ is the normalised first (most energetic) eigenfunction of concentration,

a_1 is the first expansion coefficient chosen from the PDF of coefficients given and discussed in section 4.5, and

$\varepsilon(t)$ is the POD model error.

Now, since $C(t)$ is a random variable, because of the nature of the phenomenon of turbulent dispersion, any prediction of it would be a prediction of one member of the ensemble of realisations considered. It is more sensible therefore to predict the mean and standard deviation of the ensemble instead, as discussed in the introduction of this work, where the standard deviation is approximated by the square root of $\lambda_1\varphi_1^2(t)$.

Given that the collapse of the first concentration eigenfunction occurred at all measuring stations for R_i=0, and at X1 and X2 for R_i=0.5 (see section 4.3), predictions for the mean concentration (and dose) field can be obtained by means of linear interpolation between values of the varied parameters.

As discussed in the introduction (section 1.1), the sample mean curve is the *best unbiased* estimator of the true mean concentration curve.

The sample mean concentration defined by eq. (6) at a measuring position is computed as

$$\overline{C}(t) = \frac{1}{n}\sum_{i=1}^{n} C_i(t) \qquad (36)$$

where n is the number of repetitions of the experiment.

To assess convergence of the statistic, the data sets were split in half and the sample mean was recalculated with $n/2$ observations. The resulting mean curves were indistinguishable from the ones found using all n observations in agreement with Schopflocher and Sullivan (1997), thus providing evidence that convergence of the statistic was attained.

Furthermore, the proposed model predicts the mean concentration when the *best* estimator of the coefficient a_1 is chosen to be the expected value of the distribution of coefficients - which is found to be zero in agreement with the theory (see section 2.2).

To predict the variance of concentration now, the first eigenvalues and squared first eigenfunctions of concentration must be predicted, so as to form their product as the concentration variance predictor. The method of linear interpolation can be used again as discussed above.

Regarding the question of estimating the first eigenfunction independently of the data, the Chatwin-Sullivan collapse theory (eq. (9)) in combination with the first comments in section 4.1 suggests that the mean concentration (which is easily modelled with a Gaussian) can be used as the predictor of φ_1, thus providing a theoretical predictive model.

5.2 Model uncertainty

As far as the predictive uncertainty is concerned, this is due to three factors:

One is the data input and observations factor, which is not dealt with in this work; the second factor is the mean estimation error and the third one is stochastic or random atmospheric (turbulent) fluctuations characterising the natural phenomenon of dispersion of a pollutant substance (Mole *et al*, 1993).

It is mandatory that any predictive model claiming reliability and completeness must incorporate its uncertainty in the output so as to assist end users in the decision making process.

Since the second and third uncertainty components are uncorrelated by virtue of eq. (20), the total uncertainty of the model is given by the following equation (*mean squared error*; Papoulis, 1965), when data input and observations errors are ignored (Hanna and Drivas, 1987):

$$E\{[C(t)-C'(t)]^2\} = E\{[\mu(t)-\mu'(t)]^2\}$$
$$+E\{[C(t)-\mu(t)]^2\}$$
$$= E\{[\mu(t)-\mu'(t)]^2\} + E\{[a_1\varphi_1(t)]^2\} + E\{[\varepsilon(t)]^2\}$$
$$= [\mathrm{l}(t)/2]^2 + \lambda_1\varphi_1^2(t) + Var\{\varepsilon(t)\} \quad (37)$$

where

$l(t)$ is the expected length of the confidence interval for the mean, λ_1 the first eigenvalue,

$\phi_1(t)$ the first normalised concentration eigenfunction and

$\varepsilon(t)$ the POD model error.

The total model uncertainty is reduced when

(a) the expected length of the confidence interval for the mean is minimised;

(b) the first (normalised) eigenvalue approaches 1 so that $\varepsilon(t)$ approaches 0 (see section 4.4).

Regarding (a), the classical approach of the construction of a large sample confidence interval (see for example Arnold, 1990; Roussas, 1973), gives large expected lengths. However, a non-parametric approach that makes no assumptions about prescribed form distributions of the concentration mean, described by Frangos and Antypas (2001), leads to a dramatic reduction of the confidence interval expected lengths.

Regarding the POD model error now, $\varepsilon(t)$, using the definition of the relative error (RE) in the L^2 space introduced in section 4.4 (eq. (32), where C_t is the observed concentration signal and Cr_t the predicted concentration signal now from the POD model at sampling time t, it is possible to obtain a measure of $\mathrm{var}\{\varepsilon(t)\}$ by computing the variance of *RE* when Cr_t is "predicted" by eq. (35) for values of the parameters for which C_t has been observed:

$$VRE = \frac{1}{n-1}\sum_{i=1}^{n}(RE_i - MRE)^2 \qquad (38)$$

where *MRE* is given by eq. (33).

Here, of course, the model employs only the first (most energetic) principal component.

The following table (Table 3) summarises the results concerning the model predictors variance in all cases exhibiting common features, that is, the cases with corresponding collapsing first eigenfunctions (R_i=0, all stations and R_i=0.5, low-height stations – see section 4.3).

Table 3: Variance of POD model predictors a_1' *and* $F'(t)=a_1'\varphi_1'(t)$.

R_i	Station	Variance of a_1'	*VRE* of $F'(t)=a_1'\varphi_1'(t)$
0	X1	0.438	0.037
	X2	0.220	0.014
	X3	0.397	0.033
	X4	0.213	0.013
0.5	X1	0.458	0.031
	X3	0.481	0.039

6. Discussion of results and conclusions

In this paper is provided a model for the description and prediction of concentration fluctuations of gases dispersing in the atmosphere. The model describes turbulence fluctuations that charactrerise the physical phenomenon of dispersion of contaminants in the atmosphere (Chatwin, 1981, 1982). Initially it was hoped that by applying the POD method of analysis, a model unifying the common features of neutrally buoyant and heavy gases would be obtained. The research provided in the previous sections did not prove to be all that fruitful, nevertheless:

The proposed model claims *universality* for neutrally buoyant gases with respect to downwind distance and height, as well as for gases close to the threshold between neutral buoyancy and heaviness, with respect to downwind distance.

For these gases predictions can be made, that is, the end user of the model is in position predict the mean concentration with the associated POD model uncertainty will be at a fixed time instant, downwind location, height and Richardson number, the varied parameters taken into account by the model. As far as heavier gases are concerned, physical modelling (box models, as discussed by Andronopoulos (1992) can be employed so as to provide predictive results

extending thus the proposed model for the neutrally buoyant gases case.

An important advantage of the model is its *simplicity* in the sense that only a few modes (sometimes only one) suffice so as to describe and predict the concentration field. This is in agreement with the findings of Kevlahan *et al*. (1994).

Among the limitations of the model, except for its inability to incorporate heavier gases, one notes that predictions are limited within the range of the values of the varied parameters because of their small sample sizes; extrapolation beyond these values being thus performed via physical modelling as mentioned above.

Regarding now the estimation of population parameters in the model, the sample mean lends itself as a reliable estimator of the true mean being *unbiased*, while the first eigenfunction's shape can be estimated by the mean concentration. Rice and Silverman (1991) propose smoothing the first eigenfunction via the choice of a smoothing constant by the method of cross-validation, thus counteracting its tendency to track high frequency components not present in the population eigenfunction. This may result in a predictor with smaller expected mean squared error. No attempt was made in this work to apply the smoothing method, for it could be argued that smoothing the first concentration eigenfunction might lead to loss of valuable information concerning turbulence phenomena.

Acknowledgements

This work was supported partially by a grant from the University of Sheffield, UK. The authors would like to thank N. Mole and C. Anderson for useful discussions and comments. The work is dedicated to the memory of P. Chatwin and D. Zafiropoulos.

References

Andronopoulos S. (1992): "A review of vapour cloud dispersion models" Report EUR 143 29 EN
Arnold S.F. (1990): "Mathematical Statistics". Prentice – Hall.
Batchelor G.K. (1953): "The theory of homogeneous turbulence". Cambridge University Press.
Berkooz G., Elezgaray J., Holmes P., Lumley J., Poje A. (1994): "The proper orthogonal decomposition, wavelets and modal approaches to the dynamics of coherent structures". Applied Scientific Research, 53, 321-338.
Chatwin P.C. (1981). The Statistical Description of the Dispersion of Heavy Gas Clouds. Report on Contract No 1189/01.01, Health and Safety Executive, Sheffield, UK.
Chatwin P.C. (1982). The use of statistics in describing and predicting the effects of dispersing gas clouds. J Haz Mat 6:213-230.
Chatwin P.C., Sullivan P.J. (1990): "A simple and unifying physical interpretation of scalar fluctuation measurements from many turbulent shear flows". J. Fluid Mech. 212, 533-556.
Chatwin P. (1990): "Statistical methods for assessing hazards due to dispersing gases" Environmetrics, 1(2): 143-162.
Frangos C., Antypas A. (2001): "Jackknife interval estimation of air pollutant concentrations from road traffic". Proc. 14th Conference of Hellenic Operational Research Soc.
Griffiths R.F. (1991): "The use of probit expressions in the assessment of acute population impact of toxic releases" J. Loss Prevention in Process Industries, 4, 49-57.

Hall D.J., Waters R.A., Marsland G.W., Upton S.L., EMMOTT M.A. (1991): "Repeat variability in instantaneously released heavy gas clouds – some wind tunnel experiments". Report LR 804 (PA), Warren Spring Laboratory, Department of Trade and Industry, U.K.

Hanna S.R. and Drivas P.J. (1987): "Guidelines for use of vapour cloud dispersion models". Center for Chemical Process Safety, AIChE, N.Y.

Holmes P., Lumley J.L., Berkooz G. (1996): "Turbulence, Coherent structures, Dynamical systems and Symmetry". Cambridge University Press.

Hunt J.C.R., Rothman J.W. and Britter R.E. (1983): "Some physical processes involved in the dispersion of dense gases". IUTAM Symposium on Atmospheric Dispersion of Heavy Gases and Small Particles, Deift, The Netherlands, August 1983.

Kevlahan N.K.–R., Hunt J.C.R., Vassilicos J.C. (1994): "A comparison of different analytical techniques for identifying structures in turbulence". Applied Scientific Research 53, 339-355.

Lilliefors H.W. (1967): "On the Kolmogorov – Smirnov test for normality with mean and variance unknown". American Statistical Journal,6, 99-402.

Makridakis S., Wheelright S.C., McGee V.E. (1978): "Forecasting: Methods and Applications". J. Wiley & Sons.

Mole N., Chatwin P.C., Sullivan P.J. (1993): "Modelling concentration fluctuations in air pollution". Modelling Change in Environmental Systems (ed. Jakeman, Beck and McAleer), J. Wiley & Sons Ltd.

Mole N. (1995): "The α-β model for concentration moments in turbulent flows". Environmetrics, 6, 559-569.

Mole N., Clarke E.D., Rietzler E. (1997): "Modelling concentration fluctuation moments for spherically symmetric mean concentration". Il Nuovo Cimento, 20c, N.3

Mole N. (2001): "The large time behaviour in a model for concentration fluctuations in turbulent dispersion". Atmospheric Environment, 35, 833-844.

Mole N., Schopflocher T.P., Sullivan P.J. (2007): "High concentrations of a passive scalar in turbulent dispersion" under consideration for publication in JFM.

Monin A.S., Yaglom A.M. (1965), "Statistical Fluid Mechanics: Mechanics of turbulence" Volume 1, The MIT Press.

Moseley D.J. (1991) M.Sc. dissertation, University of Western Ontario.

Papoulis A. (1965): "Probability, random variables, and stochastic processes". McGraw – Hill.

Rice J.A., Silverman B.W. (1991): "Estimating the mean and covariance structure non-parametrically when the data are curves" J. R. Statist. Soc. B 53, No 1 pp.233-243.

Robinson C. (1996): "An overview of gas dispersion with fences". University of Sheffield.

Roussas G.G. (1973): "A first course in Mathematical Statistics". Addison – Wesley.

Schopflocher T.P., Sullivan P.J. (1997): "Spectral analysis of heavy contaminant clouds". Environmetrics, 8, 603-612.

Shapiro S.S., Wilk M.B. (1965): "An analysis of variance test for normality (complete samples)". Biometrika, 52, 591-611.

Sullivan P.J. (1990): "Physical modelling of contaminant diffusion in environmental flows". Environmetrics 1(2), 163-177.

Tennekes H., Lumley J.L. (1972): "A first course in turbulence". MIT Press.

Venables W.N., Ripley B.D. (1999): "Modern applied Statistics with S-PLUS" Springer.

Zimmerman W.B., Chatwin P.C. (1995) Boundary Layer Meteorology, 75, 321.

СОД Модель для Концентрации Флуктуаций Газов Мгновенно Выпущенных в Атмосферу

Анастасиос ANTYPAS (соотвествующий автор)

Тел. и факс: (210)8991372, e-mail: t_antypas@yahoo.com

Адрес: 4, Ул. Пригипос Петру 16673 Вула, Атика, Греция

Джон BARTZIS

Адрес: Университет Западной Македонии, Департамент Машиностроения,

Ул. Салвера & Бакола, 50100 Козани, Греция

Резюме: В этой работе, разлагается сигнал концентрации повторяющихся мгновенных выбросов опасных газов в атмосферу на свои основные компоненты с помощью Собственной Ортогональной Декомпозицией (СОД) с двойной целью, его восстановления от первых (самых энергичных) компонентов и его прогнозирования в случаях имеющих общие особенности со случаями из которых ети особенности были извлечены с применением СОД. Анализ показал, что модель может составляться интерполяцией в диапазоне варьируемых параметров. Важный вопрос о неопределенности, связанной с предлагаемой модели СОД, обсуждается и предоставляется масштаб неопределенности.

Ключевые слова: Концентрация, дисперсия, прогноз.