International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Model Fitting and Optimal Design of Atmospheric Tracer Experiments: Part I

Fedorov, V. and Pitovranov, S.E.

IIASA Working Paper

WP-88-065

July 1988

WORKING PAPER

MODEL FITTING AND OPTIMAL DESIGN OF ATMOSPHERIC TRACER EXPERIMENTS: PART I.

V.V. Fedorov
S.E. Pitovranov

July 1988
WP-88-65

IIASA
International Institute
for Applied Systems Analysis

MODEL FITTING AND OPTIMAL DESIGN OF ATMOSPHERIC TRACER EXPERIMENTS: PART I.

V.V. Fedorov
S.E. Pitovranov

July 1988
WP-88-65

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

Foreword

Dr. M. Dickerson, Deputy Division Leader of Lawrence Livermore National Laboratory (USA) visited IIASA briefly in the autumn of 1987. In his discussions with the authors of this paper, Dr. Dickerson suggested that some of the statistical methods developed at IIASA and in the USSR might be useful in the analysis of atmospheric tracer data. Subsequently, Dr. Dickerson provided tapes of data collected during a series of field studies at the Savannah River Laboratory (USA).

The results of a preliminary analysis of these data were reported by S. Pitovranov at a workshop on Optimal Design of Environmental Networks, organized by the Electric Power Research Institute (Palo Alto, California, May 1988) and by V. Fedorov and S. Pitovranov at a subsequent seminar at Lawrence Livermore National Laboratory.

In my view, the results are quite remarkable and deserve rapid publication. In particular, traditional approaches to locating sampling stations downwind of a point source are shown to be inefficient, resulting in ill-conditioned problems of parameter estimation.

I agree with Dr. Dickerson who summarized as follows the benefits to be derived from this study:

- Better design of field tracer experiments used to evaluate models;
- Better objective estimates of accident parameters, e.g., height of release and source strength;
- Better testing of the sensitivity of model parameters to determine those that are most crucial to providing dose estimates;
- Better placement of samplers following an accidental release of material.

R.E. Munn
Head, Environment Program

Abstract

The results presented in this paper are that part of IIASA's activity related to the application of the statistical methods in the optimization of monitoring networks (see, for instance, Fedorov et al. 1987; Fedorov and Mueller, 1988; Mueller, 1988).

The main approach is based on the optimal experimental design theory. Two things are essential for this approach: an experimenter must have a model, or set of competitive models, which describe the observed process appropriately and he must formulate quantitatively the objective of the experiments. In the forthcoming Part II, the monitoring network design problem will be considered for cases which include prior uncertain inputs, i.e., weather conditions during a designed experiment. Some corresponding theoretical results have been reported by Atkinson and Fedorov (1988).

MODEL FITTING AND OPTIMAL DESIGN OF ATMOSPHERIC TRACER EXPERIMENTS: PART I.

V.V. Fedorov and S.E. Pitovranov

Introduction

The tracer experiments performed by Savannah River Laboratory (SRL) in 1983, as part of its Mesoscale Atmospheric Transport Studies (MATS), were used for examining the skill of the MATHEW/ADPIC coupled model in the prediction of the pollutant spatial distribution downwind from a point source (Rodriguez and Rosen, 1984).

A comparison of predicted concentrations and the observed data made by Rodriguez and Rosen (1984) showed considerable discrepancies between model predictions and the observed pollutant spatial concentration distributions in some of the tracer experiments. It was recognized that improvement of the simulation performance needs a better experimental design for model evaluation including location of the receptor sites and the choice of meteorological conditions surrounding the releases.

In this work we concentrate our attention on the former aspect, namely the determination of the number and location of sampling sites. For the sake of clarity, a simple Gaussian model has been chosen as a test for the proposed methodology of the experimental design.

MATHEW supplies a three-dimensional array of winds to ADPIC which sums the advective and diffusive components of the velocities to describe the movement of Lagrangian particles in a Eulerian framework.

1. A short description of the SRL tracer experiment

The SRL records contain the results of the 14 tracer experiments in which SF_6 was released at the Savannah River Plant at a controlled rate during 15-minute periods, from a height of 62 m.

Prior to any release, a meteorologist predicted the most probable path of the effluent cloud. This information was used to guide a field operation team in the deployment of samplers. The sampling interval lasted 20-minutes and the source receptor distance was 30 km. The separation distance between samplers was approximately 1 to 1.5 km. Figure 1 shows a typical sampler experimental layout.
From Figure 1 it can be seen that the samplers are located along the arc (which in fact coincides with a road) in relation to the source, marked by the letter S (for details see Rodriguez and Rosen, 1984).

The data base also includes meteorological parameters. Wind speed and direction observations were analyzed to obtain 15 minute averages as well as the standard deviations of the wind direction fluctuations. The locations for a subset of the stations are shown in single letters in Figure 1. At about 30 km north of the source, a 304 m. television tower was instrumented at seven levels to obtain turbulent, wind and temperature data which are also averaged every 15 minutes. All 14 experiments were conducted during daytime, between the hours of 14:00 and 16:30. The wind speed in all experiments was in the range from 2 to 5 m/sec.

2. Model description

The simplest and most extensively used model for local scale dispersion is the Gaussian model. The concentration distribution from a single release is

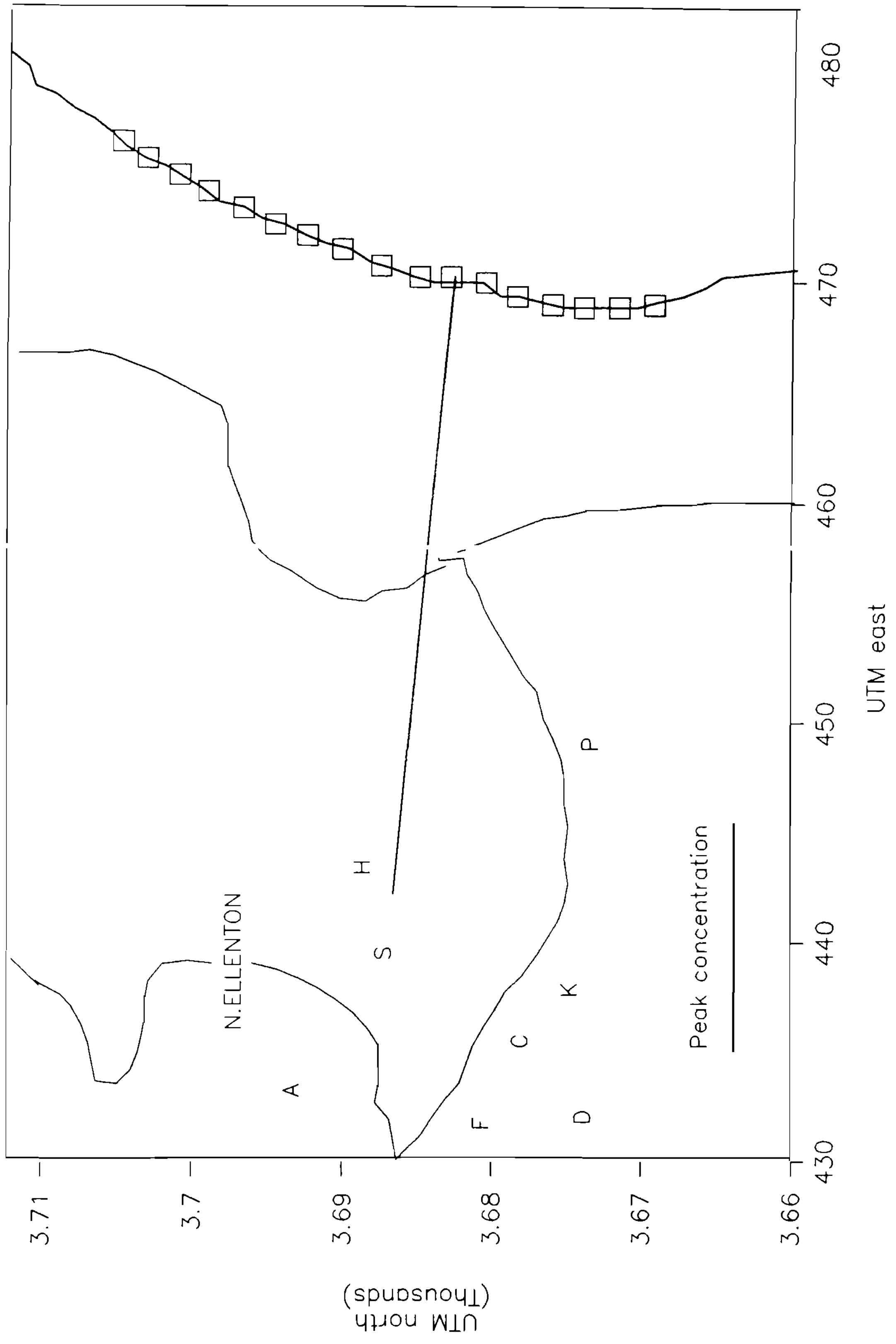

Figure 1: **Location of the source (S), 62m meteorological towers (single letters) and the samplers (small squares) during the experiment at SRL (MATS 8, 22 July, 1983).**

$$\eta_{puff}(x,t,\vartheta) = [(2\pi)^{3/2}\sigma_x\,\sigma_y\,\sigma_z]^{-1}\vartheta_1 \exp\{-(x_1-\vartheta_5-\vartheta_7 t)^2/2\sigma_x^2-(x_2-\vartheta_6-\vartheta_8 t)^2/2\sigma_x^2\}$$

$$\left\{\exp\left[-\frac{(x_3-\vartheta_2)^2}{2\sigma_z^2}\right]+\exp\left[-\frac{(x_3+\vartheta_2)^2}{2\sigma_z^2}\right]\right\} \tag{1}$$

where η_{puff} is a pollutant concentration; x_1, x_2, x_3 are the coordinates of a sampler; t is the travel time; ϑ_7 and ϑ_8 are the mean wind speed along horizontal components x_1 and x_2 correspondingly; ϑ_1 is the total amount of material released at time $t=0$; ϑ_2 is the effective height of release; ϑ_5 and ϑ_6 are the coordinates of source.

Variances σ_x, σ_y and σ_z are functions of travel time (see e.g., Doury, 1976). A simple hypothesis is that:

$$\sigma_x = \sigma_y = \vartheta_3\, t,\ \sigma_z = \vartheta_4\, t^{1/2}. \tag{2}$$

More sophisticated functions can be used in (2) but it is not very crucial for our considerations. The instantaneous surface concentration is defined by the integral

$$\eta_{inst}(x,t,\vartheta) = \int_0^{\vartheta_9} \eta_{puff}(x,t-\tau,\vartheta_1,\ldots,\vartheta_8)d\tau , \tag{3}$$

where ϑ_9 is the duration of the release.

Each sampling interval lasts $t_j - t_{j-1}$ = 20 min., $j=1,\ldots,k$, and the measured value is

$$\eta(x,t_j,\vartheta) = \int_{t_{j-1}}^{t_j} \eta_{inst}(x,\tau,\vartheta)d\tau . \tag{4}$$

It was assumed that the observations contained an additive "error":

$$y_{ij} = \eta(x_i,t_j,\vartheta) + \varepsilon_{ij} . \tag{5}$$

The term ε_{ij} comprises observational errors, random turbulence of atmospheric flow, deviation of the model from the "true" behavior, irregularity of terrain, etc. In what follows it is assumed that ε_{ij} are random values with zero mean $E(\varepsilon_{ij}) = 0$, independently identically distributed with finite variance σ^2. More complicated assumptions on the variance of ε_{ij} (for instance $\sigma_{ij}^2 \sim \eta^2(x_i, t_j, \vartheta)$) deserve to be considered, and this will be done in subsequent publications where more realistic $\eta(x, t, \vartheta)$ will be analyzed.

3. Model fitting.

Let us consider $\vartheta = (\vartheta_1, \ldots, \vartheta_9)^T$ as unknown parameters which should be identified on the basis of the observations y_{ij}. The data of a tracer experiment MATS-8 (22 July, 1983) has been chosen as a pattern for the model fitting, see Rodriguez and Rosen (1984). Under the abovementioned assumption on ε_{ij} it is reasonable from the statistical point of view to use the least square estimator (l.s.e.) for identification of unknown parameters.

The iterative second-order algorithm without calculation of derivatives has been applied (see Fedorov and Vereskov, 1985). This algorithm is based on ideas developed by Peckham (1970) and Ralston and Jennrich (1978). Though the algorithm demands rather extensive intrinsic calculation, it applies only once at every iteration to the subroutine where the square residuals sum

$$v^2(\vartheta) = \sum_{i,j} [y_{ij} - \eta(x_i, t_j, \vartheta)]^2 \tag{6}$$

is calculated. The majority of other methods either use this subroutine at least $m+1$ times (m is the number of unknown parameters) or demand the calculation of m derivatives $\partial v^2(\vartheta) / \partial \vartheta$ at every iteration.

The computation shows that the problem of simultaneous estimation of all unknown parameters is ill-posed. The variance-covariance matrix (or more accurately its first order approximation) of estimated parameters is essentially ill-conditioned. This fact indicates that the organization of the experiment (the design of the experiment) is not appropriate for the stated problem. Therefore, several reduced estimation problems were considered, each including only part of the nine abovementioned parameters.

Rather reasonable results were found when parameters ϑ_3, ϑ_4, ϑ_7, ϑ_8 were estimated (see Table 1 (a.b)).

Table 1: The estimates of dispersion parameters and wind speed components $\vartheta_3, \vartheta_4, \vartheta_7, \vartheta_8$

(a) $v^2(\hat{\vartheta}) = 0.37 \cdot 10^5$

$\hat{\vartheta}_3$ m/sec	$\hat{\vartheta}_4$ $m/\sqrt{sec}$	$\hat{\vartheta}_7$ m/sec	$\hat{\vartheta}_8$ m/sec
0.26 ± 0.011	7.4 ± 0.38	3.7±0.014	-0.5 ± 0.013

(b) Variance-covariance and correlation matrices

$$10^{-3}\begin{bmatrix} 0.122 & & & \\ -3.18 & 142 & & \\ -0.003 & 0.322 & 0.196 & \\ -0.006 & 0.256 & -0.002 & 0.165 \end{bmatrix}, \quad \begin{bmatrix} 1 & & & \\ -0.76 & 1 & & \\ -0.02 & 0.06 & 1 & \\ -0.04 & 0.05 & -0.01 & 1 \end{bmatrix}$$

A comparison of observed and computed concentrations for different sampling times can be seen in Figure 2. The comparison shows that the computed results are in agreement with the observed data.

Frequently, the estimation of power (ϑ_1) and time (ϑ_9) of pollutant release is an important problem for practitioners, for instance, when the characteristics of an accident are evaluated.

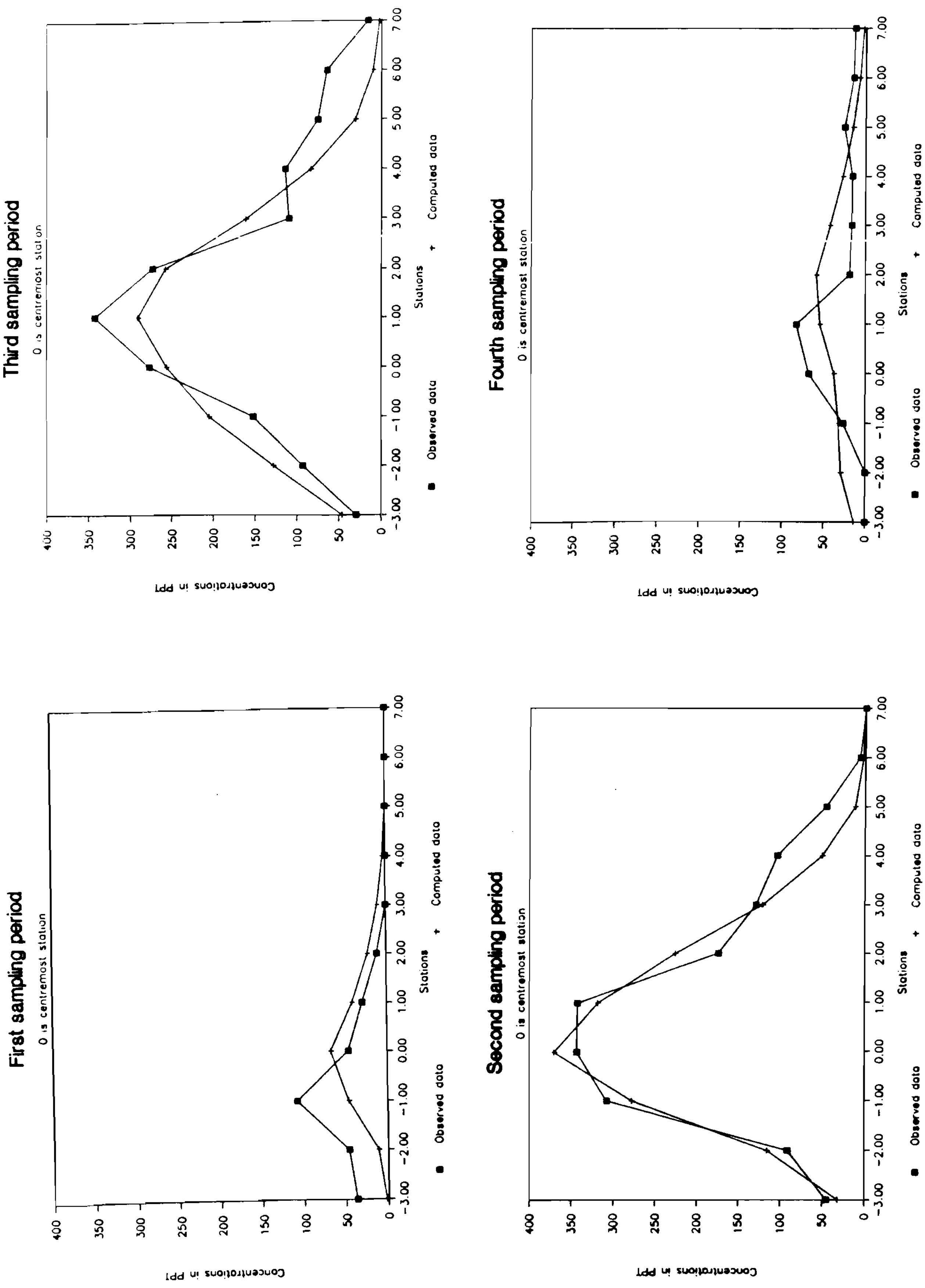

Figure 2: Observed and computed concentrations for various sampling times.

In the tracer experiments under consideration, these parameters were directly controlled and therefore known rather accurately.

For both parameters the prior values (which are initial for iterative least squares procedure) were chosen with 100% errors, i.e., $\vartheta_1 = 120$ g/sec instead of the true value 66.7 g/sec and $\vartheta_9 = 1800$ sec instead of 900 sec. Only ϑ_1 and ϑ_9 were estimated and all other parameters were fixed ($\vartheta_2 = 72m$, $\vartheta_3 = 0.26m/sec$, $\vartheta_4 = 7.4m/\sqrt{sec}$, $\vartheta_5 = 0.0m$, $\vartheta_6 = 0.0m$, $\vartheta_7 = 3.7m/sec$, $\vartheta_8 = -0.5m/sec$).

Table 2: Estimates of the power and duration of release.

(a) $v^2(\hat{\vartheta}) = 0.38 \cdot 10^5$

$\hat{\vartheta}_1$ g/sec	$\hat{\vartheta}_9$ sec
67, 1 ± 4.1	889 ± 51

(b) Variance-covariance and correlation matrices

$$10^2 \begin{Bmatrix} 0.168 & \\ -.78 & 26.0 \end{Bmatrix} , \begin{Bmatrix} 1.0 & \\ -.855 & 1.0 \end{Bmatrix}$$

Similar numerical experiments were made with the estimation of the location of the source. The computations showed that the least squares procedure also allows one to identify the coordinates of release with sufficient accuracy. (The actual location was at the origin of the coordinate system.)

Table 3: Estimates of the coordinates of release.

(a) $v^2(\hat{\vartheta}) = 0.38 \cdot 10^5$

$\hat{\vartheta}_5$ *m*	ϑ_6 *m*
-300 ± 104	60 ± 98

(b) Variance-covariance and correlation matrices

$$10^2 \begin{Bmatrix} .108 & \\ -.874 & 95 \end{Bmatrix} \cdot \begin{Bmatrix} 1.0 & \\ -.009 & 1.0 \end{Bmatrix}$$

It was found impossible to identify the effective height of release from the SRL tracer experiment sampler locations. Approximately the same residuals were for ϑ_2 equal to 210 meters $v^2(\hat{\vartheta})$ $(0.37 \cdot 10^5)$ as for 10 meters $v^2(\hat{\vartheta})$ $(0.39 \cdot 10^5)$. The standard deviation of assessment of the effective height is equal to 127 m.

4. Optimal design of experiment.

4.1. *Uncontrolled sampling intervals.*

The set $\xi = \{p_i, x_i\}_{i=1}^{n}$ is usually called (see, for instance, Fedorov 1972) a design, where "weights" p_i could be the duration, frequency or the precision of the observation which has to be made at a point x_i (this is called the "supporting point"). Searching for optimal design ξ providing in the sense of some objective function, the best estimations of unknown parameters of a regression model (in our case, model (5)) is a traditional problem in optimal design theory (see Fedorov, 1972; Silvey, 1980). The non-standard element in regression model (5) is the dependence of the model upon t_j which are known, but cannot be controlled.

The asymptotic information matrix in this case has the following structure (for details see Fedorov and Atkinson, 1988):

$$M(\xi_n) = n^{-1} T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{T} f(x_i, t_j, \vartheta_0)\, f^T(x_i, t_j, \vartheta_0) = n^{-1} \sum_{i=1}^{n} m(x_i)\,, \qquad (7)$$

where $f(x,t,\vartheta) = \partial\eta / \partial\vartheta$ and ϑ_0 are the prior values of the estimated parameters.

The optimal exact design ξ_n^* is a solution of the following minimization problem:

$$\xi_n^* = Arg \min_{\xi_n} \Psi[n^{-1} \sum_{i=1}^{n} m(x_i)] . \tag{8}$$

It has to be pointed out that (8) admits repeated observations at some points x_i. In the sampler location problem, this could cause some difficulties: one cannot locate two or more samplers at the same site. Of course, x_i^* can be considered as the central point of some relatively small area, where all these samplers can be neighbors. For more information, see Section 5.

It is crucial that (7) has an additive structure and therefore the traditional results and algorithms (see for example Fedorov, 1972; Fedorov et al, 1987) can be applied. The D-criterion (i.e., $\Psi(D) = \ln|M|$, where $D = M^{-1}(\xi_n)$ or some part of it) was used as the optimality criterion in this study.

4.2. *Experiments admitting different "weights".*

The first series of experiments were done to seek an optimal design for the estimation of parameters ϑ_3 and ϑ_4 with various travel times. To avoid the calculation (which can be very extensive for a more sophisticated model) of the partial derivatives $f(x,t,\vartheta)$ at every iteration, they were computed and stored on a regular grid with a mesh spatial scale 1.0 × 1.0 km, with the help of the auxiliary program before starting the iterative procedure. The same idea was used by Gribik *et al.* (1976) in one of the first attempts to optimize regional air pollution monitor networks. The computed designs for several travel times can be seen in Figure 3. The optimal number of supporting points is either 3 or 5. One station is on the plume centreline (in all exeriments it is assumed that the wind speed in x_2 direc-

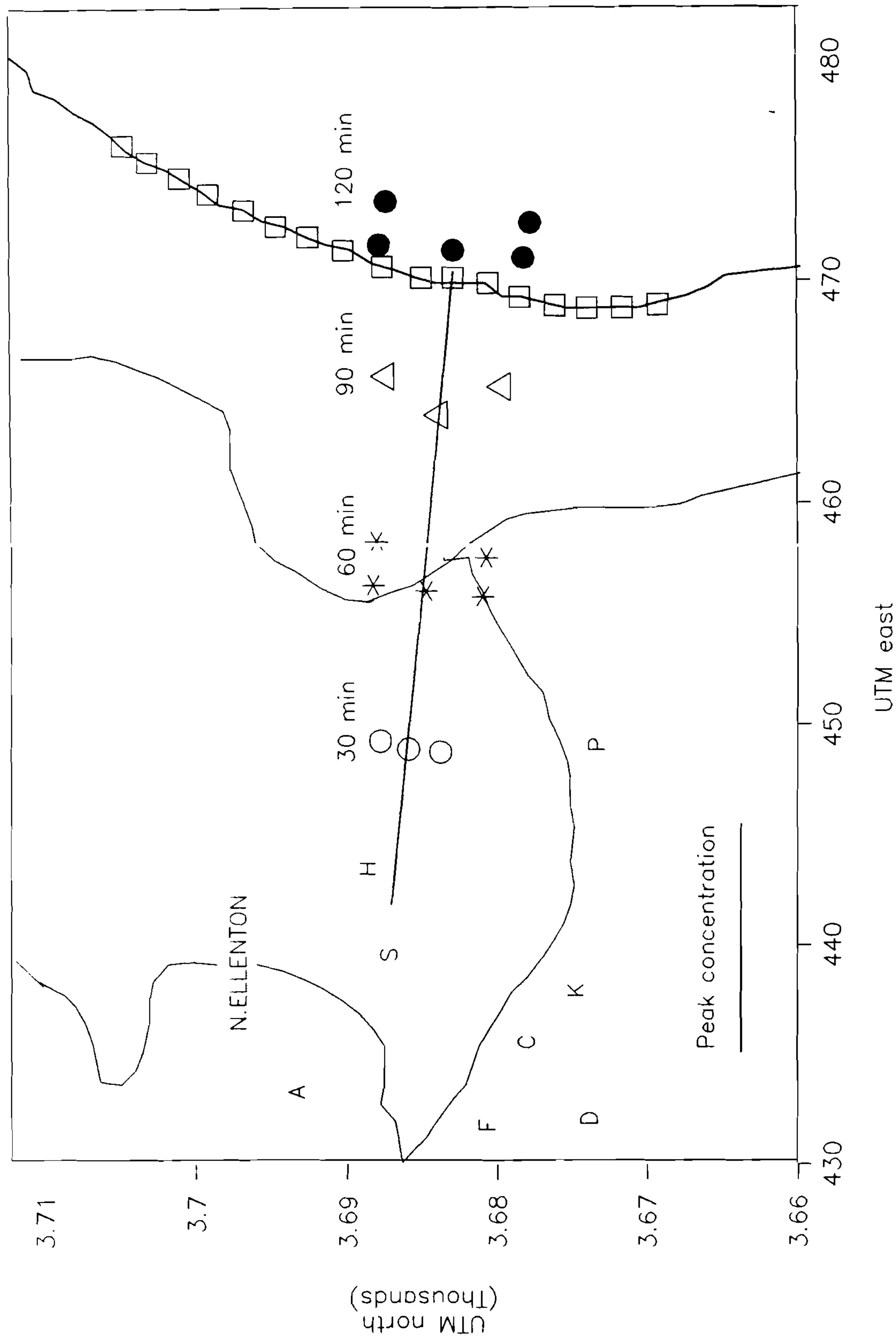

Figure 3: **Optimal sampler location for estimation of dispersion parameters for various travel times.**

tion is equal to zero), and others are allocated symmetrically. The distance between the centerline station and other stations increases from 2 to 3 km when we increase the travel time. The value of the determinant of the information matrix is the characteristic of informativeness of the observing network. From Table 4 one can see that the value of the determinant changes considerably as a function of the travel time.

Table 4: Determinants of information matrix for various travel times.

Travel time, min	30	45	60	75	90	105	120
D	58.4	7.9	1.5	0.5	0.17	0.06	0.03

Assessment of an accidental release needs knowledge of three main parameters: the source strength (ϑ_1), time of release (ϑ_9), and effective height of release (ϑ_5). Therefore, the corresponding optimal designs were computed for the model under consideration. The median wind direction during an accidental non-elevated release is assumed to be known. The problem is to allocate samplers to estimate parameters ϑ_1, ϑ_5 and ϑ_9 as precisely as possible. Optimal design for this case contains three supporting points allocated along the centerline direction (Figure 4). The dependence of the determinant of variance-covariance matrix of estimated parameters ϑ_1, ϑ_5 and ϑ_9 for various travel times is given in Table 5.

Table 5: Determinants of information matrix of estimates of release power, release time, and release height for various travel times.

Travel time, min	45	90	120
D	36	0.003	0.0001

It can be seen that the determinant decreases sharply when travel time increases from 45 min to 90 min. The standard deviation of the effective height of the release

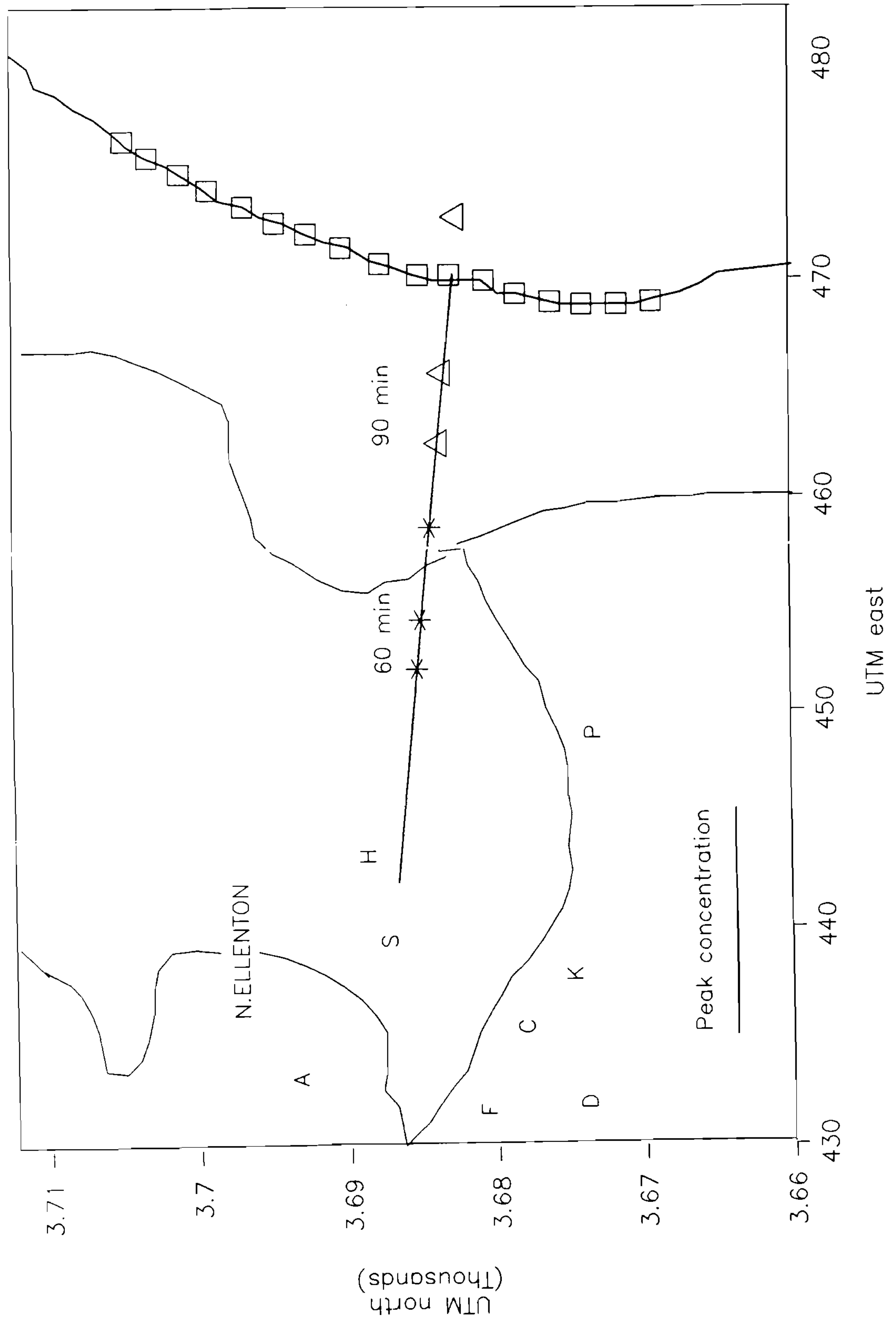

Figure 4: Optimal sampler locations for estimation of source strength, effective height, and release duration for various travel times.

estimate increases from 6 to 127 meters correspondingly. This effect is related to the fact that at some distance from the source, the released material becomes well-mixed due to vertical dispersion. Therefore, at a distance of 20 km from the source it is impossible to identify the height of the release.

5. Experiments with the prescribed number of samplers.

Assume that a number of samplers is available. The problem is to allocate these samplers in an optimal way, i.e., to find the solution of problem (8), under the constraint that supporting points x_i^*, $i = \overline{1,n}$, cannot coincide and that all weights p_i are equal n^{-1}. To find the corresponding solution one can apply the exchange type algorithm developed by Fedorov, 1986.

The comparison of optimal allocation of 20 stations for the estimation of dispersion parameters and SRL samplers allocation design can be seen in Figure 5.

The samplers should be allocated rather close to the source and their distribution over the region should reflect the shape of the pollutant cloud. The determinant of information matrix of estimated parameters for such an allocation is $D = 0.1$, while the determinant for the SRL tracer experiment is equal to 0.002.

6. A remark on the empirical design of sampler locations.

It is evident that to some extent any serious physical experiment is designed to make it sensitive to the parameters of interest. Possibly this was done when the original sampler allocation (see Figure 1) was chosen.

For the model considered in this paper even the *empirical approach* leads nevertheless to different sampler allocations. This fact emphasizes the evident statement that the optimal allocation essentially depends upon the model.

Suppose one wishes to evaluate the dispersion coefficients in model (1). Very roughly, the empirical procedure for the construction of the optimal sampler allo-

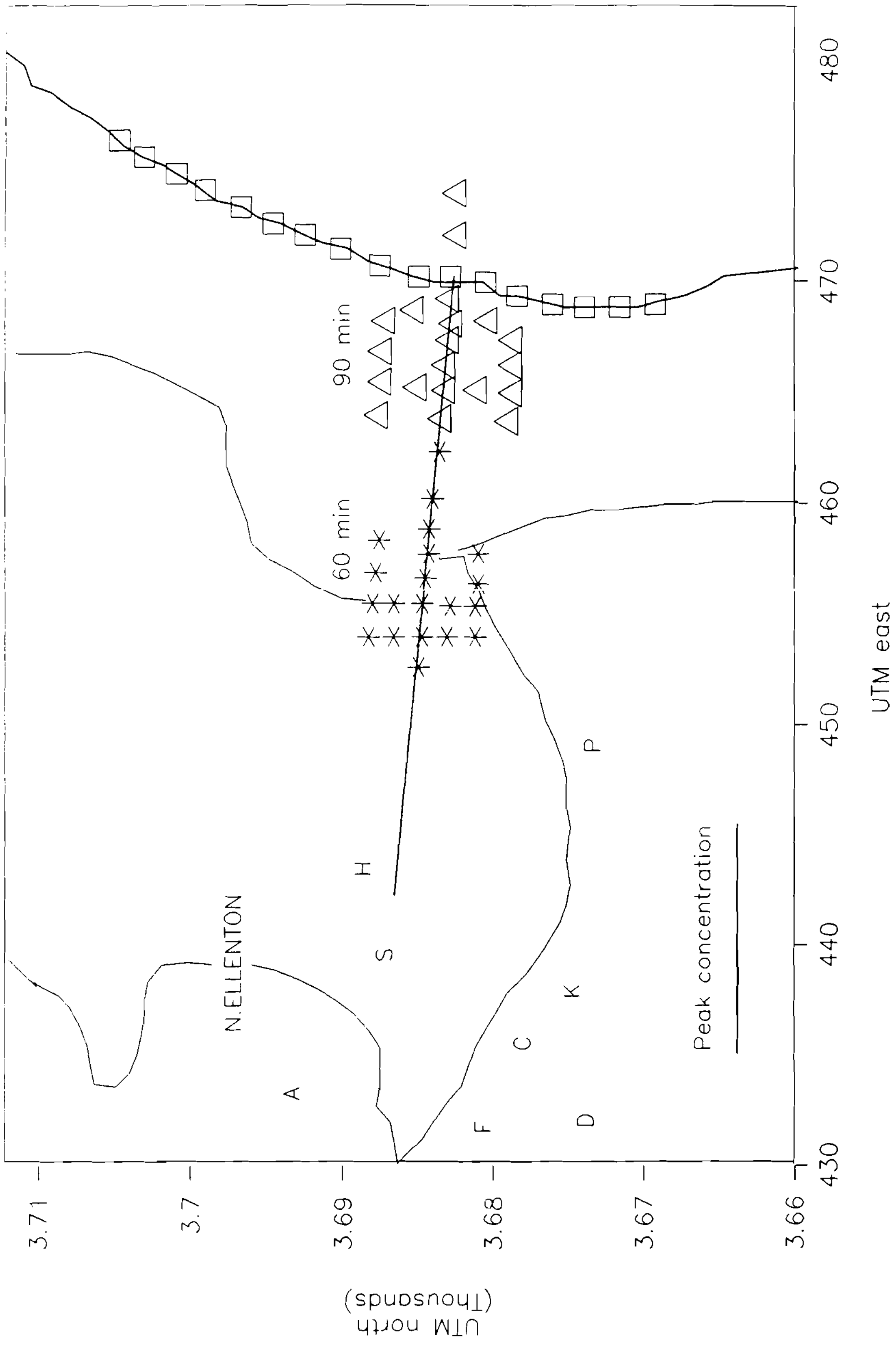

Figure 5: **The optimal allocation of 20 stations for the determination of dispersion parameters for various travel times.**

cation can be described in our case as follows:

- If the terrain is uniform, there has to be a symmetry in the sampler locations because of the symmetry of the considered model.
- A number of samplers have to be located along the centerline direction to measure the peak concentration and the rest of the available samplers have to be symmetrically remote from them. The distance between the source of emission and the samplers at the centerline direction is mainly defined by the wind speed. Samplers have to be located where the ground peak concentration is sufficiently high (maybe the highest) to be reliably measured.
- The other samplers have to be located at points where it is possible to observe the gradient of tracer concentration but, nevertheless, the signal/noise ratio has to be sufficiently high to obtain reliable observations.

Manipulation with the known wind speed (~ 5m/sec) and the most probable values of the dispersion coefficients for the given type of weather conditions leads to samplers allocation similar to Figures 3 and 5.

It is clear that in spite of the use of some mathematics, our answer has a partly qualitative character. At the same time the methods considered in the previous sections allow one to put the solution of the design problem on a well-formalized basis, converting the optimal design of sampler allocation into a routine computing operation.

Probably for more sophisticated models than (1) and more complicated experimental situations, one has to combine both approaches.

Conclusions

1. The allocation of all available samplers along one arc (see Figure 1) results in ill-conditioned problems of parameter estimation.

2. The tracer experiment should be designed with the linkage of the parameter estimation problem for certain air-pollutant transport model (or models, when one has to choose between them). The optimal location of samplers is sensitive to the structure of this model (or models).

3. The optimal design depends upon the objective function, which has to reflect the experimenter's needs expressed in qualitative form.

References

Doury, A. (1976) Une Méthode de Calcul Pratique et général pour le prévision numérique des pollutions véniculees par l'atmosphère, Centre d'études nucléaries de Seclay, Commissariat à l'énergie atomique, Rapport CEA-R-4280 (Rév.1).

Draxler, R. (1984) Diffusion and Tracer Experiments. pp 367-422 Ch.9. in R. Randerson (ed.), Atmospheric Sciences and Power Production. Tech. Inform. Center, Office of Scientific and Tech.Inform. US DOE Science and Power Production.

Fedorov, V. (1972) Theory of optimal experiments, N.Y., Academic Press, pp.292.

Fedorov, V. (1986) The experimental design of an observation network: optimization algorithms of the exchange type, Working Paper WP-86-62, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Fedorov, V. and Vereskov A. (1985) Numerical aspects of some nonstandard regression problems, Working Paper WP-85-32, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Fedorov, V., Leonov, S. and Pitovranov, S. (1988) Experimental Design Technique in the Optimization of a Monitoring Network, pp.165-175 in V. Fedorov and W. Laüter (eds.) *Model-Oriented Data Analysis. Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag.

Fedorov, V. and Mueller, W. (1988) Two Approaches in Optimization of Observing Networks, pp.239-256 in Y. Dodge et al. (eds.) *Optimal Design and Analysis of Experiments*. North-Holland.

Fedorov, V. and Atkinson, A.C. (1988) The optimum design of experiments in the presence of uncontrolled variability and prior information, pp.327-344 in Y. Dodge et al. (eds.) *Optimal Design and Analysis of Experiments*, North-Holland.

Mueller, W. (1988) Design of an Air-Pollution Monitoring Network. Working Paper WP-88-64, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Ralson, M.L. and Jennrich, R.J. (1978) DUD, a derivative-free algorithm for nonlinear least squares, Technometrics, **20**, p.7-14.

Rodriguez, D.J. and Rosen, L.C. (1984) An evaluation of a series of SF_6 tracer releases using the MATHEW/ADPIC model. Lawrence Livermore National Laboratory, Rept. UCRL-91854.

Silvey S.D. (1980) Optimal Design, London, Chapman and Hall, pp.86.

Acknowledgements

The authors are grateful to M. Dickerson, R. Harvey and P. Gudicksen from Livermore Laboratory, and to M. Antonovsky and R. Munn, for their constructive comments.