Uploaded to VFC Website

~ October 2012 ~

This Document has been provided to you courtesy of Veterans-For-Change!

Feel free to pass to any veteran who might be able to use this information!

For thousands more files like this and hundreds of links to useful information, and hundreds of "Frequently Asked Questions, please go to:

Veterans-For-Change

Veterans-For-Change is a 501(c)(3) Non-Profit Corporation
Tax ID #27-3820181

If Veteran's don't help Veteran's, who will?

We appreciate all donations to continue to provide information and services to Veterans and their families.

https://www.paypal.com/cgi-bin/webscr?cmd=_s-xclick&hosted_button_id=WGT2M5UTB9A78

Note: VFC is not liable for source information in this document, it is merely provided as a courtesy to our members.

Item ID Number	02374
Author	Belli, G.
Corporate Author	
Report/Article Title	Typescript: Statistical Interpolation Model for the Description of Ground Pollution Due to the TCDD Produced in the 1976 Chemical Accident at Seveso in the Heavily Contaminated Zone "A"
Journal/Book Title	
Year	1983
Month/Day	November
Color	☐
Number of Images	45
Description Notes	

STATISTICAL INTERPOLATION MODEL FOR THE DESCRIPTION OF GROUND POLLUTION DUE TO THE TCDD PRODUCED IN THE 1976 CHEMICAL ACCIDENT AT SEVESO IN THE HEAVILY CONTAMINATED ZONE "A".

G.Belli[2], S.Cerlesi[1], E.Milani[3] and S.Ratti[3,4].

1) Regione Lombardia - Ufficio Speciale Seveso (Mi)
2) Istituto Tecnico Industriale Pavia
3) Istituto di Fisica Nucleare Università Pavia
4) Istituto Nazionale di Fisica Nucleare Sezione Pavia

Novembre 1983

IFNUP/RL 14/83

INDEX

1. INTRODUCTION

Following the results of ref (1) and what anticipated in the last paragraph of that paper, the present report is dealing with the rigorously mathematical and statistical problem of finding the best methodological procedure to determine in a way as unbiased as possible the quantity of TCDD deposited on the ground within the limits of the so-called "zone-A" around the Icmesa Factory in Seveso (Italy).

The main features of the problem have been anticipated in ref. (1) here, in Section 2 we shall define our data sample while in Section 3 we shall go through all major mathematical aspects in order to prepare the actual procedure applied in Section 4. Finally in Section 5 the graphycal result of the interpolation is presented.

2. DEFINITION OF THE DATA SAMPLE

It is important to qualify and certify the data used in the integration process: the data used are those obtained in the 1976 Campaign (December '76) limited to zone A (for the reasons given in ref. (1)).

The topographical distribution of the coordinates belonging to the points in which the ground sample has been collected is shown in fig. 1.

Zone A has an extention of about 1Km in the East-West direction and of about 2 Km in the North-South direction and the collecting points follow an approximately regular grid of about 50 m x 50 m.

The values of TCDD concentration on the ground vary from a minimum of 0.75 $\mu g/m^2$ (corresponding to the detectable limit of the analitical measurements-denoted as N.V.) and a maximum of 5477 $\mu g/m^2$.

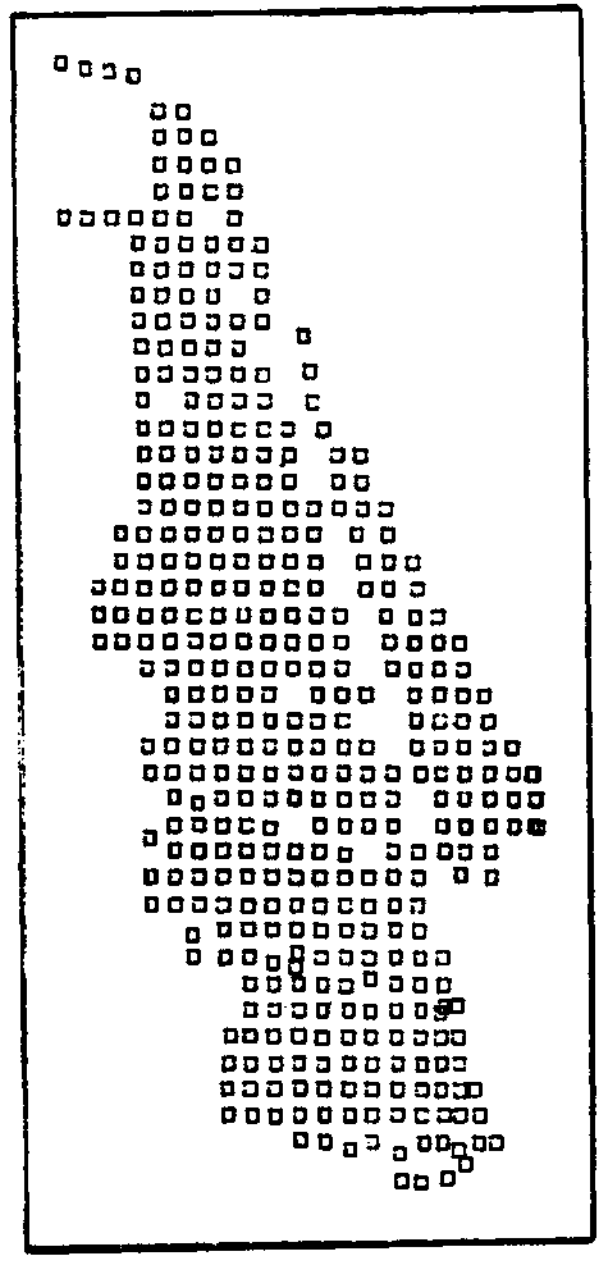

Fig. 2.1: Every squares is the thopographycal rapresentation of the single sample of the 1976 campaign.

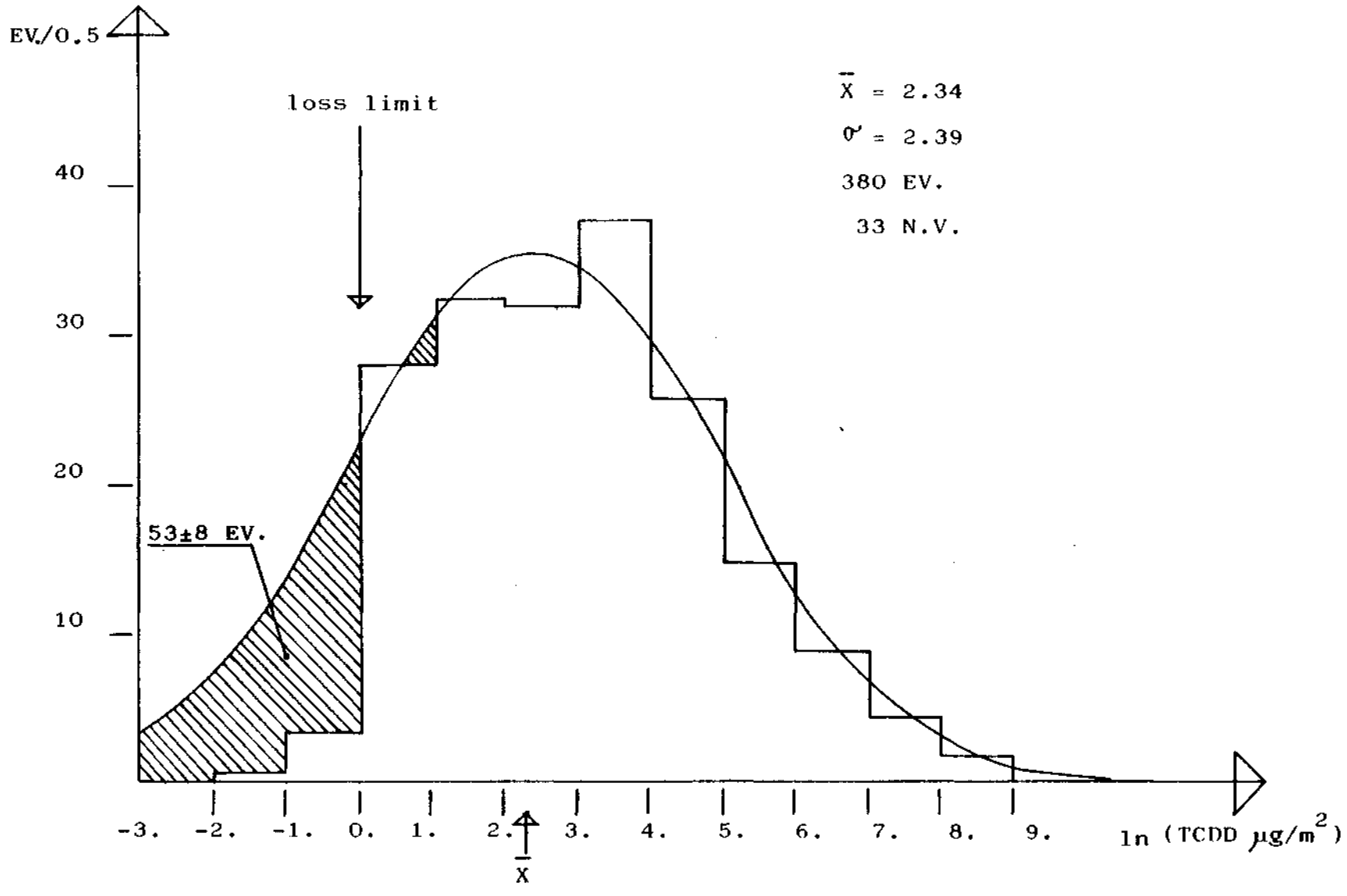

Fig. 2.2: Histogram of all samples and gaussian curve fitted using only the right part of data starting from the "loss limit"; numerical results of procedures fit.

In order to take into account the main indications of a previous analysis of the overall TCDD distribution[2], in the present report we use as a quantification parameter the logarithm of the TCDD density since this is the quantity showing a gaussian distribution (see fig. 2.2) thus being the most suitable variable to be used in an optimization process.

With this variable, the contaminant ranges from a minimum value of - 0.287 $\ln(\mu g/m^2)$ and a maximum value of 8.608 $\ln(\mu g/m^2)$.

3. THEORY OF THE MATHEMATICAL METHOD

3.1 Introduction

In this Section we define the problem in its general aspects and propose the algorithms which have been included in the program used for the numerical solution of it.

Let D be the quantity of TCDD under consideration which depends upon the values directly measured of the location (x,y) which define the coordinate vector (hereafter referred to as vector $\underline{x} = (x,y)$).

Given a set of data providing the values D_j of the quantity $D(\underline{x})$ in several points $\underline{x}_i = (x_i, y_i)$ in a 2-dimensional space, our target will be to find a$_n$ explicit function of the independent vector-coordinate variable

$$y = f(\underline{x})$$

To this end a program[3] will be used to find by means of a least square fit a reasonable approximation, of the type

$$D = \sum_{1}^{n} c_i \, f_i(\underline{x}) \qquad (3.1.1)$$

approximating the measured values D_i distributed within the set of measured coordinates (x,y).

In (3.1.1) c_i are constant coefficients and $f_i(x)$ are proper polynomial functions of the vector variable $\underline{x}_i$.

The following Sections shall be devoted to the study of the mathematical functions, to their choice and to the choice (and definition) of the constant coefficients c_i.

For sake of semplicity we shall treat the case of a single independent variable x, which can be easily extended to a more general multidimensional space $\underline{x}$.

3.2 Approximation of an arbitrary function by means of a set of given functions.

We propose the general problem of representing an arbitrary function f(x), by means of a finite number of functions chosen with a certain degree of arbitrarity, for instance polynomial of increasing power such as:

$$p_o(x),\ p_1(x),\ p_2(x),\ldots\ldots p_n(x) \qquad (3.2.1)$$

or rather by means of arbitrary linear combinations of the type:

$$S_n(x)=a_o p_o(x)+a_1 p_1(x)+a_2 p_2(x)+\ldots\ldots+a_n p_n(x) \qquad (3.2.2)$$

where the coefficients a_k can be choosen, within the linear combination, in such a way that the difference

$$d_n(x) = f(x)-S_n(x) \qquad (3.2.3)$$

be "the smallest possible" (in the most general sense still to be discussed).

It is more than clear that the criterion of "small error" as given by (3.2.3) is largely arbitrary and driven by the principal interest of using the approximation $S_n(x)$ instead of the function f(x) (which might well be unknown).

Among the most frequently used criteria there is the least-square criterion which requires to minimize the average quadratic error in the range (a,b) for x

$$\bar{d}_n^{\,2} = \frac{1}{b-a} \int_a^b d_n(x)^{\,2}\, dx \qquad (3.2.4)$$

that is to determine a_k by means of a least square fit.

The use of a quadratic form is largely considered an optimal application to the practical calculations. A least square fit approximation may give locally large discrepancies but it gives in general a rather accurate global representation of the function f(x).

3.3 Conditions on the coefficents a_k

The most obvious condition to be imposed is that $\bar{d}_n^{\,2}$ be the minimum possible as a consequence of the choice made on a_k; being $\bar{d}_n^{\,2}$ a positive polynomial in the variables $a_o, a_1, .. a_n$ it indeed admits a minimum.

Thus zeroing the first derivatives with respect to a_k we shall obtain an unique solution corresponding to a minimum. Thus:

$$\frac{\partial \bar{d}_n^{\,2}}{\partial a_k} = 0 \qquad k = 0,1,2,\ldots, n \qquad (3.3.1)$$

From (3.2.4) and (3.2.3) we get, by deriving with respect to

a_k within the integral:

$$\frac{\partial \bar{d}_n^{\,2}}{\partial a_k} = \frac{1}{b-a} \int_a^b 2\left[f(x)-S_n(x)\right] \frac{\partial}{\partial a_k}\left[f(x)-S_n(x)\right] dx =$$

$$= \frac{2a}{b-a} \int_a^b -p_k(x)\left[f(x)-S_n(x)\right] dx \qquad (3.3.2)$$

$$k = 0,1,2....n$$

where we take advantage of the fact that the arbitrary function $f(x)$ is independent of a_k, while S_n for the (3.2.2) depends linearly on a_k through the choosen polynomials $p_k(x)$.

Formula (3.3.1) thus becomes:

$$\int_a^b p_k(x)S_n(x)dx = \int_a^b p_k(x)f(x)dx \quad k=0,1,2,...n \qquad (3.3.3)$$

and, given the (3.3.2) for $S_n(x)$, formula (3.3.3) assumes the shape:

$$\int_a^b p_k(x)f(x)dx = a_o\int_a^b p_k(x)p_o(x)dx + a_1\int_a^b p_k(x)p_1(x)dx + ... +$$

$$+ a_n \int_a^b p_k(x)p_n(x)dx \qquad k=0,1,2,..,n \qquad (3.3.4)$$

or rather in a compact form, for any k :

$$\int_a^b p_k(x)f(x)dx = \sum_o^n{}_j a_j \int_a^b p_k(x)p_j(x)dx \qquad (3.3.5$$

We have thus obtained a linear system of (n+1) equations in the (n+1) unknowns a_j.

The solution of such a system is uniquely determined provided the

$$
\begin{aligned}
f_1 &= c_{00}\, a_0 \\
f_2 &= c_{10}\, a_0 + c_{11}\, a_1 \\
f_3 &= c_{20}\, a_0 + c_{21}\, a_1 + c_{22}\, a_2 \\
&\cdots\cdots\cdots \\
&\cdots\cdots\cdots \\
f_n &= c_{n0}\, a_2 + c_{n1}\, a_1 + a_{n2}\, a_2 + \ldots + c_{nn} a_n
\end{aligned}
\qquad (3.4.1)
$$

this selections allows us to add the significant conditions

$$
\int_a^b p_k(x) p_j(x)\,dx = 0 \quad j>k, j,k=0,1,2,\ldots n \qquad (3.4.2)
$$

It is important to note at this point that, as j and k are two indices restricted by the relation k<j, but otherwise arbitrary integers, eq. (3.4.2) must hold also when k and j are interchanged, i.e. k>j so that (3.4.2) implies a more restricted condition:

$$
\int_a^b p_k(x) p_j(x)\,dx = 0 \quad j \neq k \qquad \begin{matrix}\text{ORTHOGONALITY} \\ \text{CONDITIONS}\end{matrix} \qquad (3.4.3)
$$

Thus imposing the triangularization of the linear system we authomatically obtain its diagonalyzation[3]; this is due to the fact that the coefficient matrix $|C_{jk}|$ (3.3.7) is symmetric as it immediately appears by inspecting the definition (3.3.6).

If the orthogonality conditions are verified by the system of the $p_k(x)$ functions chosen as a basis for the approximation, then the system (3.3.5) is now reduced to the form, for any k:

determinant of the coefficient matrix

$$C_{kj} = \int_a^b p_k(x)p_j(x)dx \qquad (3.3.6)$$

be different from zero; i.e.:

$$\mathrm{Det}||C_{kj}|| \neq 0 \qquad (3.3.7)$$

It has to be clear that, if n is large, the expression to solve the system (3.3.5) which can be rewritten as:

$$\sum_{o}^{n}{}_j C_{kj}a_j = F_k \qquad (3.3.8)$$

where $F_k = \int_a^b p_k(x)f(x)dx$, may be anything but a simple problem. In such a case the use of the approximation $S_n(x)$ would be a bad and unconfortable choice.(4)

3.4 Choice of the functions $p_k(x)$

In order to overcome the difficulty we have to perform a suitable choice on all possible functions $\{p_k(x)\}$ in such a way that the solution of the system (3.3.5) be as simplified as possible.

Among the most interesting choices is a selection of the functions $p_k(x)$ such that the system (3.3.5) be reduced to a triangular form, without lack of generality. That is a selection on the basis of which in the K-th equation only the unknowns a_j with $j \leq k$ appear, i.e.:

$$a_k \int_a^b p_k(x)^2 dx = \int_a^b p_k(x) f(x) dx \qquad (3.4.4)$$

since only the integral having j=k turns out to be different from zero in the left-hand side of (3.3.5).

It follows immediately then that a_k is simply given by

$$a_k = \frac{1}{N_K^2} \int_a^b p_k(x) f(x) dx \quad \text{for any } k \qquad (3.4.5)$$

where

$$N_K^2 = \int_a^b p_k(x)^2 dx \qquad (3.4.6)$$

is a positive normalization constant; i.e. a positive number which depends exclusively upon the functions $p_k(x)$ but not upon the function f(x) to be approximated.

In this subsection we have thus reached the following result: if the functions $p_k(x)$ are orthogonal (and it is always possible to select a set of well known orthogonal functions) the problem of approximating an arbitrary function f(x) is solved by (3.4.4).

3.5 Legendre Functions and Tchebychev Functions

In the previous Sections we have introduced the problem of the least-square approximation and clearly underlined the need to select a system of functions having the requisite of obeying the orthogonality conditions since in such a case the solution of (3.3.5) is particularly simplified.

The ortogonalization process suggested by Gram-Schmidt[4] provides a tool to select from among a finite or an infinite set of linearly independent functions defined in the range (a,b) a set of functions which are orthonormal in (a,b).

In this section we briefly propose two important and well known orthogonal systems which will constitute the alternative basis of our data handling.

The first system is provided by the Legendre polynomial[5,6] functions. They are defined by the formula:

$$p_n(x) = \frac{(2n-1)(2n-3)\ldots 1}{n!}\; x^n - \frac{n(n-1)}{2(2n-1)}x^{n-2} + \qquad (3.5.1)$$

$$+ \ldots\ldots\ldots\ldots\ldots$$

and alternatively by the recurrence formula

$$p_{(n+1)}(x) = \frac{2n+1}{n+1}\; xp_n(x) - \frac{n}{n+1}\; p_{n-1}(x) \qquad (3.5.2)$$

from which the very first components are easily derived:

$$\begin{aligned} p_o(x) &= 1 \\ p_1(x) &= x \\ p_2(x) &= \tfrac{1}{2}(3x^2-1) \\ p_3(x) &= \tfrac{1}{2}(5x^3-3x) \\ &\ldots\ldots\ldots \end{aligned} \qquad (3.5.3)$$

It is of fundamental importance the result:

$$\int_{-1}^{+1} p_m(x)p_n(x)dx = o \qquad m \neq n \qquad (3.5.4)$$

which is a consequence of the orthogonality conditions. It implies that the Legendre Polynomial Functions are orthogonal in the range (-1,+1).

The second example of orthogonal system is provided by the Tchebychev polynomial functions[6] which are defined by the formula

$$t_n(x) = \cos\ (n \cos^{-1} x) \qquad (3.5.5)$$

From (3.5.5), by using the De Moivre Theorem and the Theorem on the binomial's power we can rewrite:

$$T_n(x)=x^n-\binom{n}{2}x^{n-2}(1-x)^2+\binom{n}{4}x^{n-4}(1-x)^4+\ldots. \qquad (3.5.6)$$

and alternatively write the recurrence formula:

$$T_{n+1}(x) = 2x\ T_n(x) - T_{n-1} \qquad (3.5.7)$$

from which the very first components can be easily derived

$$\begin{aligned} T_o(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2-1 \\ T_3(x) &= 4x^3-3x \end{aligned} \qquad (3.5.8)$$

The Tchebychev polynomial functions are orthogonal in the range (-1,+1) since for $m \neq n$

$$\int_{-1}^{+1} \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}\, dx = 0 \qquad (3.5.9)$$

As a general matter, in a non-orthogonal model, in order to estimate the value of a coefficient a_k one has to know all the other preceeding coefficients $a_j (j<k)$ as well as all the following coefficents $a_i (i>k)$ as one has to invert the matrix (3.3.6). In an orthogonal model, on the contrary, all coeffi-

cients can be separately evaluated since we are dealing with a diagonal matrix.

However an orthogonal model has to be preferred not only for a matter of convenience and mathematical semplicity; the method appears to be very sure against the dangerous round-off errors which so often become relevant in computer working.

We conclude this section by stating that the Tchebychev polynomal functions are practical to minimize the MAXIMUM ERROR, while the Legendre polynomial functions are practical to minimize the ERROR OF THE MEAN.

3.6 Multidimensionl parametrization.

In this Section we indicate the most natural procedure to extend the proceeding algorithm to an n-dimensional space, limiting however one computation to the case under consideration which requires a 2-dimensional space.

In our case, then, for the approximation of a two dimensional function D(x,y) by means of a linear combination of Tchebychev polynomial functions we can write:

$$D(x,y) = \sum_m \sum_n C_{mn}\, T_n(x)\, T_m(y) \qquad (3.6.1)$$

Let

$$a_m(x) = \sum_n C_{mn}\, T_n(x) \qquad (3.6.2)$$

Then we can write

$$D(x,y) = \sum_m{}' a_m(x)\, T_m(y) \qquad (3.6.3)$$

Clearly in the two dimensional case a strategy for the selection of the most convenient functions is needed, since all possible

combinations $T_k(x) \cdot T_j(y)$ constitute a very large variety of possibilities.

An obvious procedure, which incidentally economizes on computer time, suggests to consider first the lower order functions; in addition, for each given function taken into account prove if its contribution to the reduction of the squares of the residuals is large enough (step by step improvement).

To prove this latter point, even using a non orthogonal model, it is not strictly necessary to invert the matrix but an orthogonalization can be reached by applying the Modified Gram-Schmidt procedure.(4)

In this way an orthogonal model is easily built in which the potential reduction in the sum of the squared residuals ΔS_i^2 is easily evaluated for the different choices of the functions f_i, in the available measured values of the representative function $D(\underline{x})$ to be approximated $D_i(x_i, y_i)$.

In summary we can handle our multidimensional approximation problem if we can reduce the large number of possible arbitrary functions to a few dozens and if we can perform an orthogonal transformation able to reduce the approximation to the one-dimensional case, which implies a single inversion of a triangular matrix, at the most.

4. PSEUDOMEASUREMENTS.

4.1 Introduction

The data sample is provided by the 1976 campaign during which carots have been collected every $\sim$50 m.

However in order to perform a two-dimensional fit the starting sample has to be increased by producing new pseudo-measurements in a much denser grid. In fact for a convergent ap-

proximation of the function D(x,y) described in Capt. 3 at least 50 measurements per coefficient is desirable.

Then our target here is to construct a new data sample which preserves the structure and the characteristics of the original one, by properly interpolating the TCDD values in intermediate points.

The starting experimental data are not sufficient to guarantee the applicability of the model; from them the results obtainable in a straight forward way are those of Ref. 1.

4.2 General Comments on the Interpolated data.

The interpolation needed to guarantee a sufficient number of starting measurements in order to perform the approximation of D(x,y) is based on the following procedure:

given in a limited region of the x,y plane a number of measurements $D_i(x_i,y_i)$ in the points (x_i,y_i), look for a function g(x,y) able to designate a reasonable value of D in any arbitrary point (x,y).

The domain in which our experimental function is defined has a non-geometrical form (since the limits of zone A are irregular as well as the distribution of the original measurement points (x_i,y_i)).Then we replace the contour with a rectangle subdivided into a regular grid both in x and y. The crossing points of the new grid are the reference points in which we intend to evaluate the new interpolating function g(x,y).

Two basic methods provide a solution of the problem:
- the first one consists in building a function g which interpolates exactly the measured values, i.e.:

$$g(x_i, y_i) = D_i \qquad i=1,2,\ldots,n \qquad (4.2.1)$$

This method gives excellent results only when the values D_i are known with high accuracy;

- The second method consists in building a function g as a weighted average of the experimental observations and it is desirable when the starting experimental data are likely to be subject to inaccuracies and large unavoidable fluctuations. Due to the nature of our data we choose the weighted interpolation, suggested by D.Shepard[7a] and recently used by others.[7b,c,d]

Essentially the weighted interpolation of sparse data irregurarly scattered can be represented by the following formula:

$$g(x_i, y_j) = \frac{\sum_{k=1}^{n} D_{kk} W(r_k)}{W(r_k)} \qquad i,j = 1,2,\ldots n \qquad (4.2.2.)$$

where:

D_{kk} is the measured value in the point (x_k, y_k)

$W(r_k)$ is a proper weighting function

r_k is the distance between the points (x_k, y_k) and (x_i, y_j)

Note that the value $g(x_i, y_j)$ represents the weighted average of the observations of the entire sample in the case of a "global interpolation"; it represents the weighted average value of the sorroundings observations for a "local" interpolation.

4.3 Shepard's Method

The Shepard's interpolation method,[7a] in its general form is applicable to measurements arbitrarily scattered.

Given in the plane a point (x,y), let r_i be the distance between (x,y) and the n points (x_i,y_i) in which measurements have been made, for any i = 1,2,...n.

The Shepard's interpolation formula reads:

$$g(x,y) = \frac{\sum_{1}^{n} {}_i \; D(x_i,y_i)\; W\;(r_i)}{\sum_{1}^{n} {}_i \; W\;(r_i)} \qquad \text{if } r_i \neq 0$$

$$g(x,y) = D\;(x_i,y_i) \qquad \text{if } r_i = 0 \qquad (4.3.1)$$

Note that (4.3.1) is defined in all points of the plane R^2 and that it interpolates exactly the values D_i in the given points (x_i,y_i), while the value g(x,y) in the "new points" is given as the weighted average of all given measurements. The contribution of the i-th measurement is weighted as a function of the distance between the point (x,y) under consideration and the given points (x_i,y_i).

It is obviously inconvenient to use this method when n is very large; however in such a case the method would not be needed.

Furthermore the method increases in selectivity when the interpolation is performed in local form.

Let us fix a radius R>0 and define a weighting function

$$W(r) = W(r) \qquad \text{if } r \leq R$$
$$W(r) = 0 \qquad \text{if } r > R \qquad (4.3.2)$$

Thus in the local form formula (4.3.1) becomes:

$$g(x,y) = \frac{\sum_{1}^{n} D(x_i,y_i) \; W(r_i)}{\sum_{1}^{n} W(r_i)} \qquad \text{if } r_i \leq R \qquad (4.3.3)$$

$$g(x,y) = 0 \qquad \text{if } r > R$$

Formula (4.3.3) is still defined in every point of the plane, but now the value of the function in the point (x,y) is given by the weighted average of the measurements $D(x_i,y_i)$ only in the neighbouring circle of radius R.

Therefore the problems is now lying in a reasonable choice of the values for the cut-off radius R in such a way that for any point (x,y) of the plane there is an adeguate number of measurements included in a circle of radius R, so as to compensate fluctuations.

This second method opens up the new possibility of choosing different values of R, in different regions of the domain within which D(x,y) is defined.

Theoretically the choice of R, in this kind of procedure depends upon the statistical sample under consideration.

In our particular case we want to define a variable R depending upon the distance from the ICMESA Factory having in mind both the topographycal distribution of the measurement points and the TCDD concentration. Infact maximal TCDD concentration is

found immediately around ICMESA and, due to the transport phenomenon caused by the wind, along a maximum concentration bound[8] in the south-east direction, while, perpendicular to such a line and away from it the TCDD concentration values are significantly decreased.

Therefore the choice for R has been done in order to maintain these particular characteristics.

4.4 Choice of the Weighting functions W(r)

In the practical case with which we are dealing we wish to introduce weighting factors W(r) able to preserve statistically the same characteristics of the original sample, a point which has been clearly pointed out from the beginning.

The choice is suggested by the successfull use, (found in the literature) in metheorology[7] to solve analogous problems such as, for instance, the distribution of the ozone concentration in the bay around Los Angeles (U.S.A.). A reference to the papers by Gustafson, Kortanek, Sweigart[7b], Goodin, McRae, Seinfeld[7c] and by Glahn[7a] is imperative.

To perform the calculations of Sect.5 we have selected three weighting functions.

The first choice is:

$$W(r) = \frac{R^2-r^2}{R^2+r^2} \qquad (4.4.1)$$

The second is:

$$W(r) = \left[1+T(r)\right] S^2(r) \qquad (4.4.2)$$

where

$$S(r) = \frac{1}{r} \quad \text{if} \quad 0<r\leqslant\frac{R}{3}$$

$$S(r) = \frac{27}{4R}\left(\frac{R-r}{R}\right)^2 \quad \text{if} \quad \frac{R}{3}<r\leqslant R$$

having defined

$$T(r) = \frac{\sum_{\ell=1}^{m} S_\ell(r)(1-\cos\alpha)}{\sum_{1}^{m} S_\ell(r)} \qquad (4.4.3)$$

in which:

- m is the number of measurement points lying within the disk of radius R
- α is the angle defined by the segment (see fig. 4.1)

$\overline{(x_k,y_k)-(x_i,y_j)}$ and $\overline{(x_i,y_j)-(x_1,y_1)}$

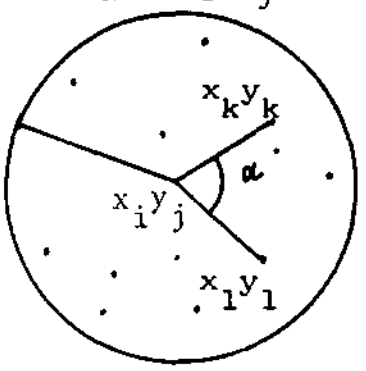

Fig. 4.1

the third choice is

$$W(r) = (1+T(r))\, S^2(r)$$

$$S(r) = \frac{R^2-r^2}{R^2+r^2} \qquad (4.4.4)$$

$T(r)$ as defined by (4.4.3)

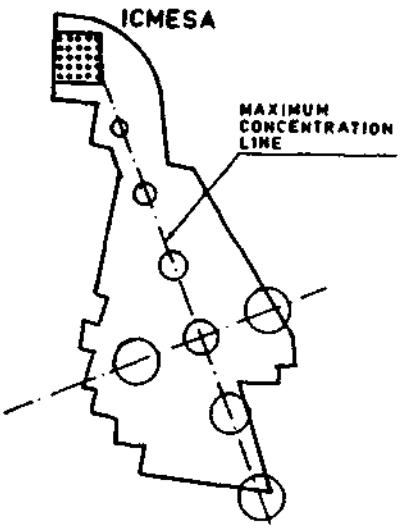

Fig. 4.2

In the first choice, formula (4.4.1), the weighting factor depends only on the distance between the point (x,y) in which we want to construct a pseudo-measurement and the original measurement-points (x_i,y_i) falling within the disk of radius R.

In the other two methods a directional dependence is also included by (4.4.3).

In all 3 methods R is variable in the plane according to the increase in the width of the contaminating cone in the wind direction along the maximum contamination line.[8] (see fig.4.2)

All the three formulae (4.4.1); (4.4.2) and (4.4.4) give final samples which are well compared in their global characteristics with the real data.

As an example fig.4.3 and fig.4.4, obtained with the program HBOOK[9], give the scatter plots of the original data while fig.4.5 and fig.4.6 give the scatter plots of the enriched sample using (4.4.2). These are not topographycal maps; in fig.4.3 ln(TCDD) is plotted versus x and in fig.4.4 ln(TCDD) is plotted versus y for the original data sample.

Comparing fig.4.3 with fig.4.5 the density of points are different but the structure of the two data samples is the same; points with large TCDD values around 5000 $\mu g/m^2 \approx \exp(8.5)$ $\mu g/m^2$ are very few and have small abscissa x (see fig.2.I for the definition of the reference frame) while the majority of the points have values between 2.7 $\mu g/m^2 \approx \exp(1.0)$ and 150 $\mu g/m^2 \approx \exp(5.0)$.

Comparing fig.4.4 with fig.4.6, the points with large TCDD values are fewer and located at large y coordinates (close to the Icmesa Factory in the reference frame of fig.2.I). Furthermore one can notice that the granularity of the information is increased but that in the regions where there was no data in fig.4.4, no pseudo-data have been invented in fig.4.6; an observation supporting the adeguateness of the Shepard's method to our goal.

In conclusion we wish to point out that (4.4.4) contains the maximum "a priori" information that can be extracted from the original data sample as a guide-line to the finding of the approximating function D(x,y).

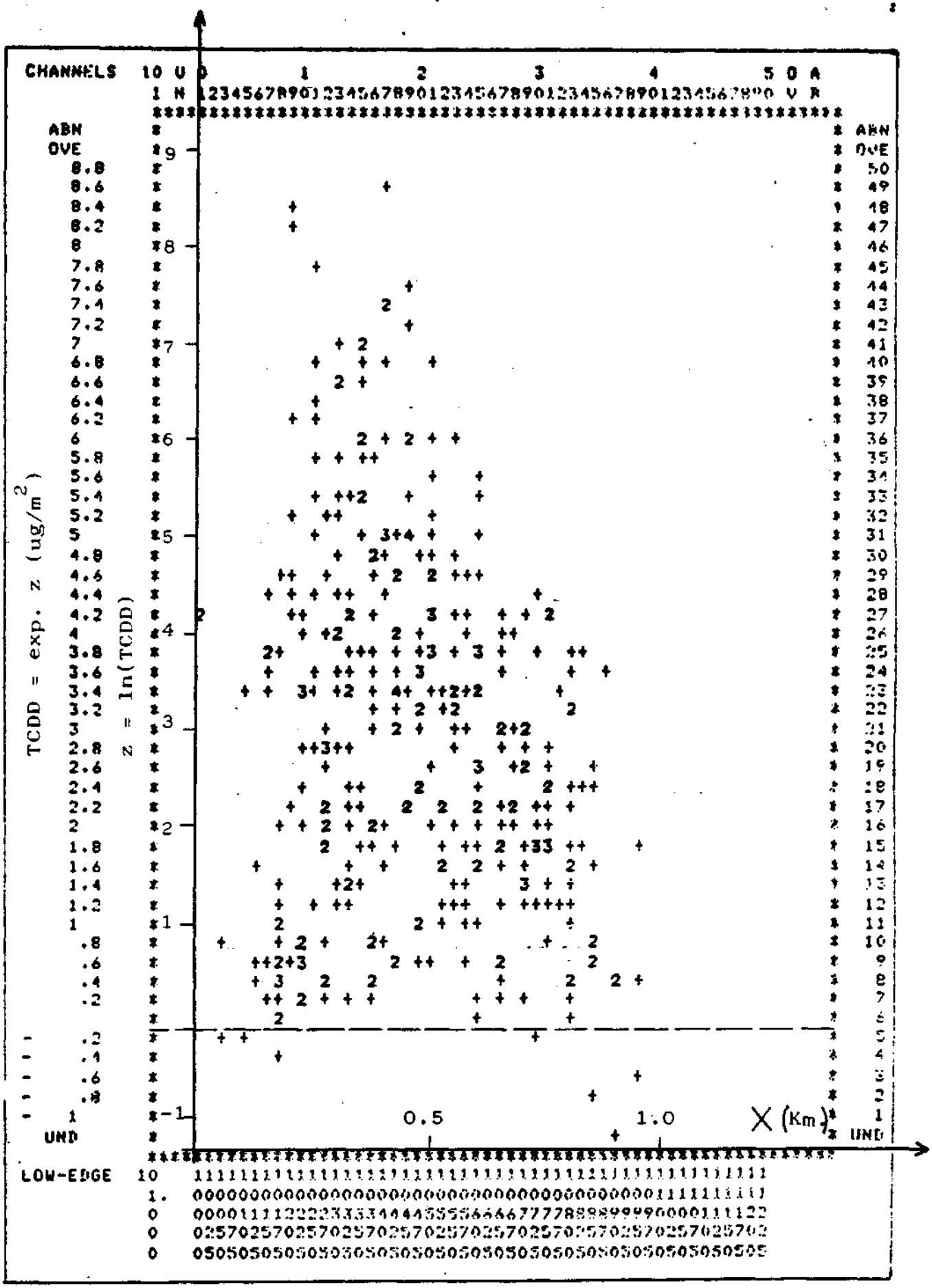

Fig. 4.3

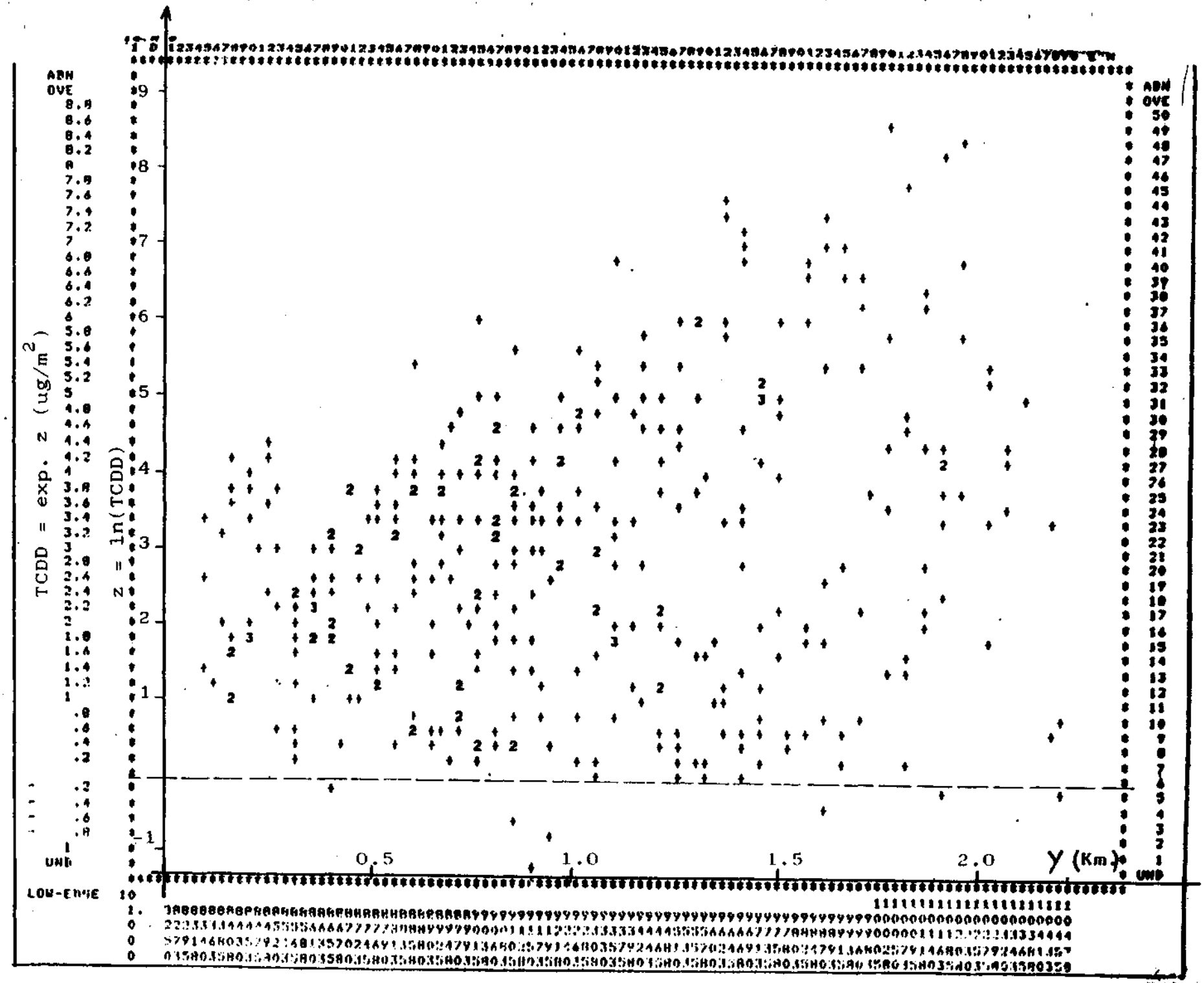

Fig. 4.4

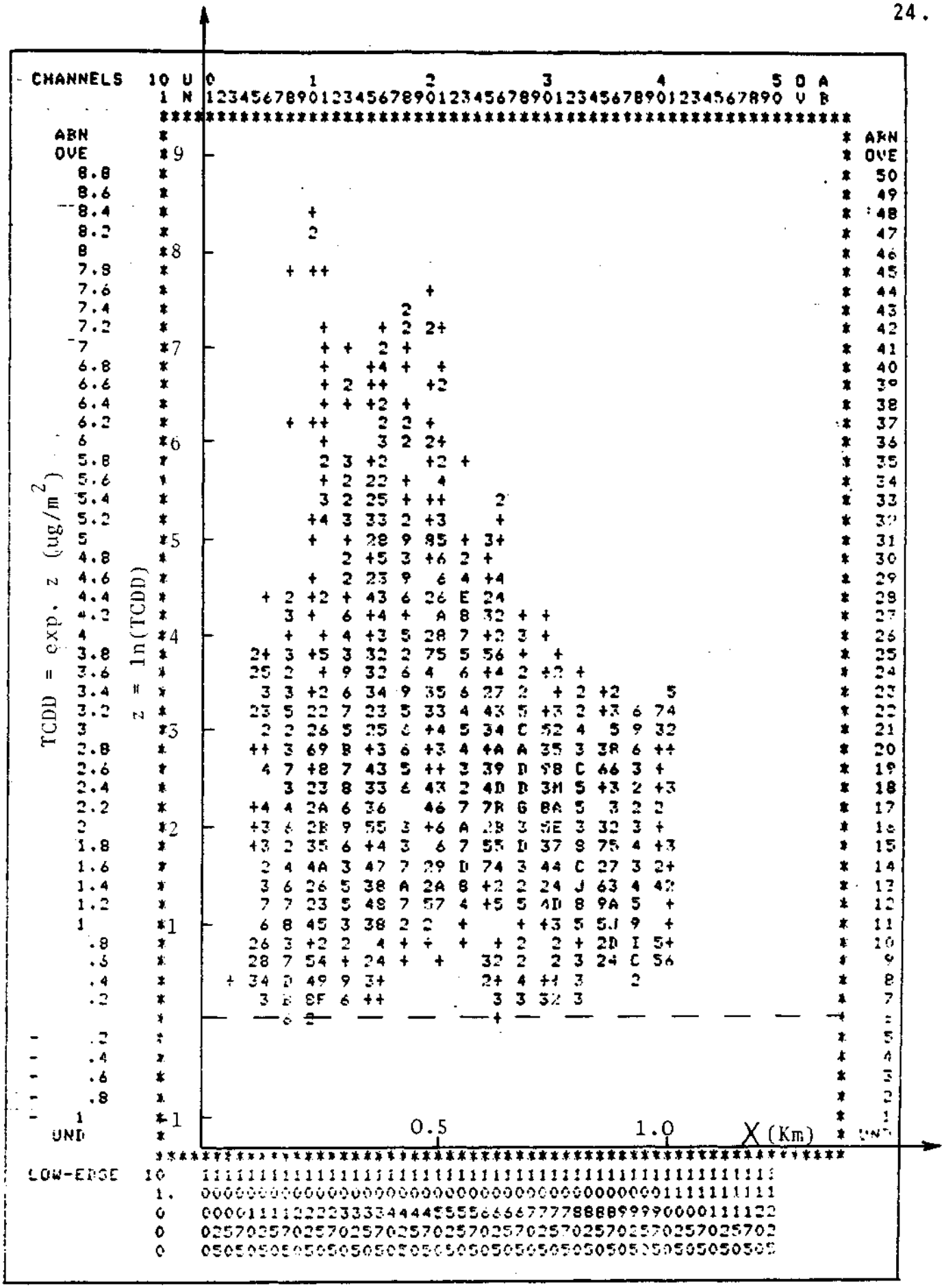
CHANNELS
ABN OVE
UND
LOW-ERGE
TCDD = exp. z (ug/m²)
z = ln(TCDD)
0.5
1.0
X (Km)

Fig. 4.5

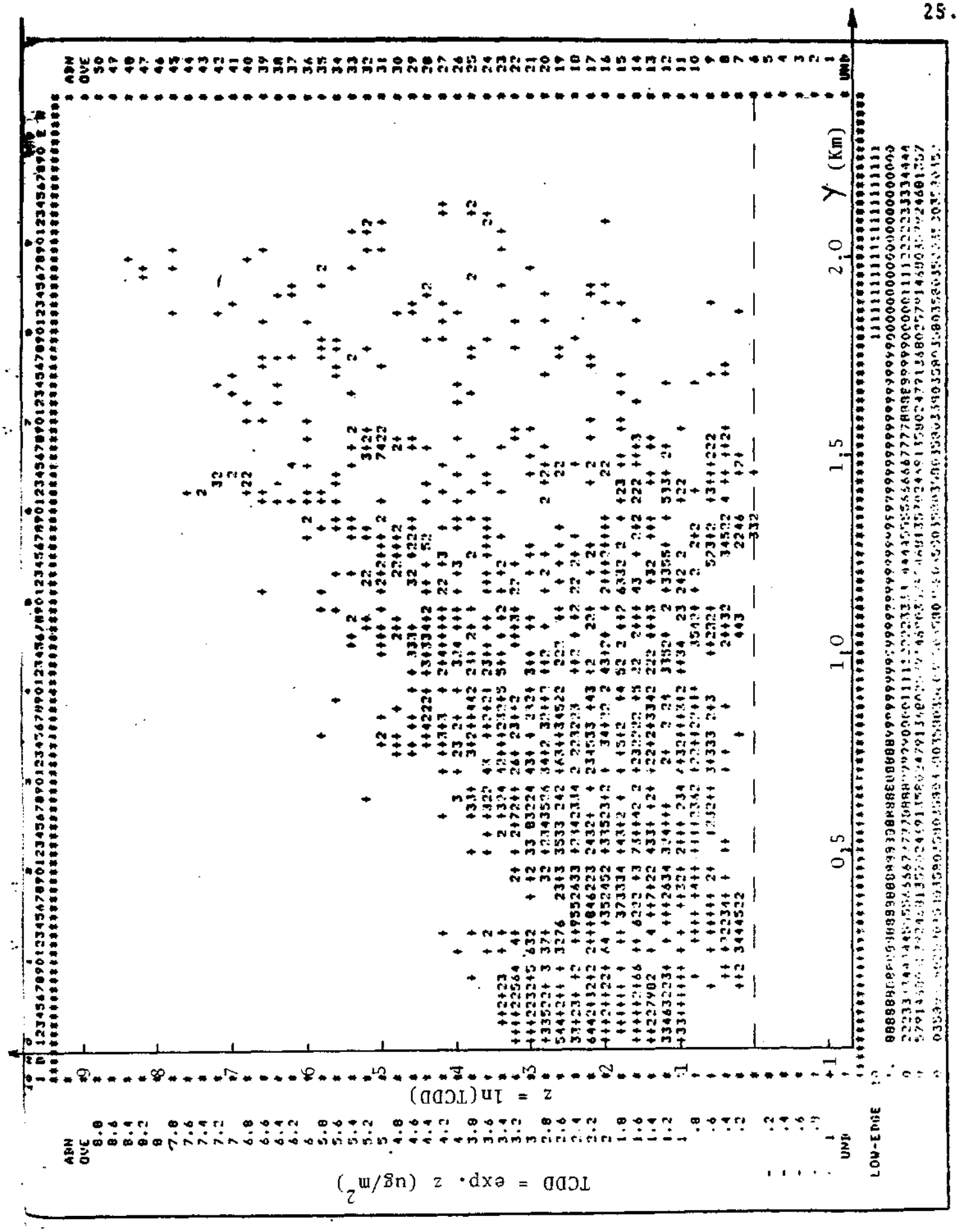

Fig. 4.6

5. GRAPHICAL REPRESENTATION OF THE RESULTS

5.1 Introduction

In this final Section we shall present the results of our investigation by showing the quantitative solution of the approximation of the contaminant distribution in zone A by means of the analytical function built with a number of Tchebychev polynomials.

Two different presentations of our results will be given:

1 - a quantitative presentation of the coefficient for the different polynomials and the confidence parameters

2 - a graphycal drawing of a 3-dimensional surface as seen from different perspective points describing the TCDD distribution in a rectangle containing zone A, using the program SURFAC[10].

We have checked the goodness of the model adopted and compared our results with those obtained in previous investigations [8,11-14]. The model is adequate and several characteristics, already pointed out by others, are verified.

5.2 Numerical results of the approximation

The data constructed with the interpolation methods described in Sect.4 have been used as input to the program MUDIFI (Multi-Dimensional-Fitting). The principal algorithms have been outlined in Section 3.

The program allows to fix the number of coefficients a_k that we want to introduce in the final form (3.1.1) and gives the possibility to specify if the approximation has to be per

formed with functions product of simple monomials in the two variables x and y or rather with orthogonal polynomial forms such as, for instance, the Tchebychev functions of Sect.3.5.

We have, thanks to the Shepard method, a relevant number of input points; thus the goal of a rather accurate representation of the unknown function $y(x,y)$ measured in n points $y_i=D_i(x_i y_i)$ can be reached by using as many as 30 free parameters (coefficients a_k) to build the approximant function $D(x,y)$.

We show in the present paper only the results obtained by the use of the Tchebychev orthogonal polynomial (3.5.5) (3.5.6) (3.5.7) and (3.5.8) mentioned in Sect.3, and on the interpolation method, mentioned in Sect.4, using only the formula (4.4.4).[14]

The all procedure however has been applied also using the Legendre polynomial[I4].

Due to the particular nature of the experimental data which often show large fluctuations even between very close points, and due to the very large ratio between the "area of zone A" and the "total area of zone A submitted to the contamination analysis", the results of our approximation can be considered as sufficiently good.

In Table 5.1 the values of the residuals for each of the coefficients are collected and the corresponding reductions together with the value of the multiple correlation coefficient C are given.

Table 5.2 collects in the first column the values obtained for the 30 coefficients a_k, in the second column the variance and in the third column the degree of the Tchebychev polynomials to which they refer.

As an explicative example, the second line quotes the coefficient a_1(col.1)=(0.759 ± 0.140) related to the combination

(col.3) 01. This means that a_1 refers to the product $T_o(x)T_1(y)$. The second coefficient (line 3) is related to the combination 20 which means that a_2 refers to the product $T_2(x)\ T_o(y)$. Explicitely:

$$D(x,y) = (2{,}61 \pm .183)+(.759 \pm .14)T_o(x)T_1(y)- \\ - (0.268 \pm .101)T_2(x)T_o(y)-....+a_{ij}T_i(x)T_j(y) \qquad (5.2.1)$$

and from (3.5.8):

$$D(x,y) = (2.61 \pm .183)+(.759 \pm .140)y-(0.268 \pm .101)\cdot \\ \cdot (2^2-1)+ ..... \qquad (5.2.2)$$

The parameter C of Table 5.1 gives an indication for the goodness of the approximation. The closer C is to unity, the better the fit can be considered.

As one can notice in Table 5.1, colum 3, the sum of the residuals is reduced by about a factor 10 per coefficient; the multiple correlation coefficient is $\sim$0.87 close enough to 1.0 for our purposes. Finally in Table 5.2 the errors on the different coefficients are very reasonable.

As a matter of principle the "mathematical" result could improve for instance by allowing a larger number of coefficents, which would require a larger number of data points. Alternatively we could "eliminate" some "bad point" giving a value of TCDD drastically different from the nearby values and reflecting an anomalous large fluctuation.

In this paper however by all means we do want to give an unbiased interpolated description without any arbitrary elimination of any value.

Therefore we claim that the result presented is the best possible in the given circumstances.

COEFF NO	SUM OF SQUARES OF RESIDUALS	REDUCTION OF OF SQUARES
1	0.3536308594E+04	0.3754131775E+03
2	0.2996910889E+04	0.5393977661E+03
3	0.2429095947E+04	0.5678148804E+03
4	0.2265266357E+04	0.1638295898E+03
5	0.2099861084E+04	0.1654051971E+03
6	0.1831718872E+04	0.2681422729E+03
7	0.1688964111E+04	0.1427547302E+03
8	0.1568103638E+04	0.1208605042E+03
9	0.1356953613E+04	0.2111499786E+03
10	0.1051628662E+04	0.3053250122E+03
11	0.9939321899E+03	0.5769648743E+02
12	0.9610200806E+03	0.3291210556E+02
13	0.8918918457E+03	0.6912826538E+02
14	0.8129942017E+03	0.7889766693E+02
15	0.7688598633E+03	0.4413433838E+02
16	0.7501573486E+03	0.1870249748E+02
17	0.7405953979E+03	0.9561931610E+01
18	0.7146628418E+03	0.2593254089E+02
19	0.6897740479E+03	0.2488878632E+02
20	0.6724725342E+03	0.1730154228E+02
21	0.6638720093E+03	0.8600533485E+01
22	0.6075493774E+03	0.5632264328E+02
23	0.5990665894E+03	0.8482768059E+01
24	0.5884318848E+03	0.1063469791E+02
25	0.5616769409E+03	0.2675496292E+02
26	0.5530006714E+03	0.8676253319E+01
27	0.5437915039E+03	0.9209181786E+01
28	0.5315605469E+03	0.1223097420E+02
29	0.5212128906E+03	0.1034767246E+02
30	0.5128872681E+03	0.8325633049E+01

MULTIPLE CORRELATION COEFFICIENT = 0.868885E+00

Tab. 5.1: Results of the function weight

$$W(r) = [S(r)]^2 (1+T(r))$$

$$S(r) = \frac{R^2 - r^2}{R^2 + r^2}$$

COEFFICIENTS	VALUE	VARIANCE	POWERS OF VARIABLES IN MONOMIAL	
0	0.2611816406E+01		0	0
1	0.7591783404E+00	0.182943E+00	0	1
2	-0.2681191862E+00	0.140330E+00	2	0
3	-0.8827306628E+00	0.100968E+00	4	2
4	-0.4400894046E-01	0.479781E+00	1	1
5	0.5425338745E+00	0.116808E+00	3	0
6	0.6950988173E+00	0.242769E+00	3	1
7	0.6278941631E+00	0.129216E+00	1	5
8	0.5745822191E+00	0.132157E+00	5	2
9	0.2406100035E+00	0.974676E-01	3	5
10	0.1857645333E+00	0.138151E+00	5	3
11	-0.3786304593E+00	0.566410E-01	0	5
12	0.3746468425E+00	0.667838E-01	2	6
13	0.2251543403E+00	0.998844E-01	5	4
14	0.6250208616E+00	0.917136E-01	7	5
15	0.3926346600E+00	0.762477E-01	7	6
16	0.2824673951E+00	0.259199E+00	1	0
17	0.4311328530E+00	0.149761E+00	0	2
18	0.6149656773E+00	0.224754E+00	1	2
19	0.6629428864E+00	0.199875E+00	2	2
20	-0.4905254394E-01	0.532743E-01	6	0
21	-0.5814285874E+00	0.121924E+00	3	3
22	-0.6244755387E+00	0.993173E-01	1	6
23	-0.6606221199E-01	0.462425E-01	0	8
24	0.2131620049E+00	0.700691E-01	4	5
25	-0.3567886055E+00	0.765648E-01	7	2
26	-0.1992565244E+00	0.877238E-01	7	3
27	0.5482710898E-02	0.753213E-01	8	4
28	-0.2291990817E+00	0.666651E-01	8	5
29	0.1528140604E+00	0.782392E-01	7	7
30	-0.1678609997E+00	0.760031E-01	8	6

Tab. 5.2: Results of the function weight $W(r)=[S(r)]^2(1+T(r))$; $S(r)=\frac{R^2-r^2}{R^2+r^2}$

5.3 Graphycal Results of the approximation.

Having obtained the analytical form of the function D(x,y) completing (5.2.2) with all the terms indicated in Table 5.2 (that is having obtained the mathematical description of the TCDD distribution on the ground),we can graphycally visualize the result so as to give a direct check of the overall properties of the distribution function and of the approximation procedure adopted. (The graphycal visualization could also suggest general comments on the topographycal distribution in comparison with the equal density line description given in ref.8).

To solve the graphycal problem,we have used the program SURFAC[10]. It is a multipurpose program which produces a prospectic view of a function given in cartesian coordinates.

The resulting figure draws the intersections of the surface with parallel planes orthogonal to the axes. To obtain the visualization of possible "hidden points" of the surface it is possible to rotate the entire figure by a variable angle (which can be properly chosen) from -90° and +90°.

In fig.s 5.1, 5.2,5.3,5.4 the function D(x,y) given by (5.2.2) is shown from different points of view under different angles. The vertical axis (ln TCDD concentration) is not relevant here and is reported only on fig. 5.1.

Let us comment on fig: 5.1 which shows the function rotated by 30° clockwise with respect to the North-South direction.

We can clearly notice that there are two pronounced peaks in the vicinity of the ICMESA Factory; then the function decreases along the y axis maintaining however large TCDD values

along the well defined band of maximum concentration mentioned in Sect. 4.4 and determined in ref. 8 .

Note that zone A is included in the rectangle (fig.5.1); the contours of the function are not always slowly degrading but show secondary maxima. This is due to the fact that the analysis has been limited to zone A where the TCDD concentration is maximal, but which does not cover all the domain of the contaminated region. Nonetheless, the secondary maxima are not very relevant and coincide with those shown in ref. 8 .

Fig. 5.3 seen from,an angle of 60° clockwise, shows the surface from a direction almost perpendicular to the maximum concentration line.

Fig. 5.4, seen from an angle of 60° counterclockwise, shows the surface almost along the maximum concentration line clearly showing the decrease of the TCDD concentration with the distance from the ICMESA Factory.

It is important to point out that the weak points of the analysis is concentrated at the boundaries. There the number of measured points is small.

It is however interesting to note that,in spite of this the surface is satisfacorily reproducing the known overall characterics of the distribution (even if the knowledge is relatively scanty).

A quick subdivision of zone A into small rectangles gives the average TCDD concentrations reported in Table 5.3.

The graphycal description reproduces also quantitatively the numerical description.

We can thus conclude that the empirical model proposed in this paper is adeguate to reproduce the reality with sufficient accu

racy and proves that the approach used may be interesting for applications to similar problems.

5.4 Concluding remarks

The analysis performed in this paper is one of the possible investigations which can be performed in connection with the ICMESA accident. Although with a much lower significance one could extend the analysis to all the contaminated region taking nontrivial risks, but providing a tentative complete mathematical description of the phenomenon.

We certainly believe that the procedures used could be applied to similar cases since, by using relatively simple and handy mathematical formulae it is possible to build a descriptive function over a given geographycal extension of an area interested by a measured phenomenon, putting in evidence both the global and local characteristics of the measurement.

As of the value of the integral we prefer to be as cautious as possible.

As explicitely stated in ref.8 and in ref.1 the multiplication factor between analyzed area and contaminated area is $R=3.04\times10^{5}$ for zone A.

This imposes a priori an enormous incertainty on any possible result. The numerical integration performed here is however the best mathematical calculation which can be performed, given the original measurements.

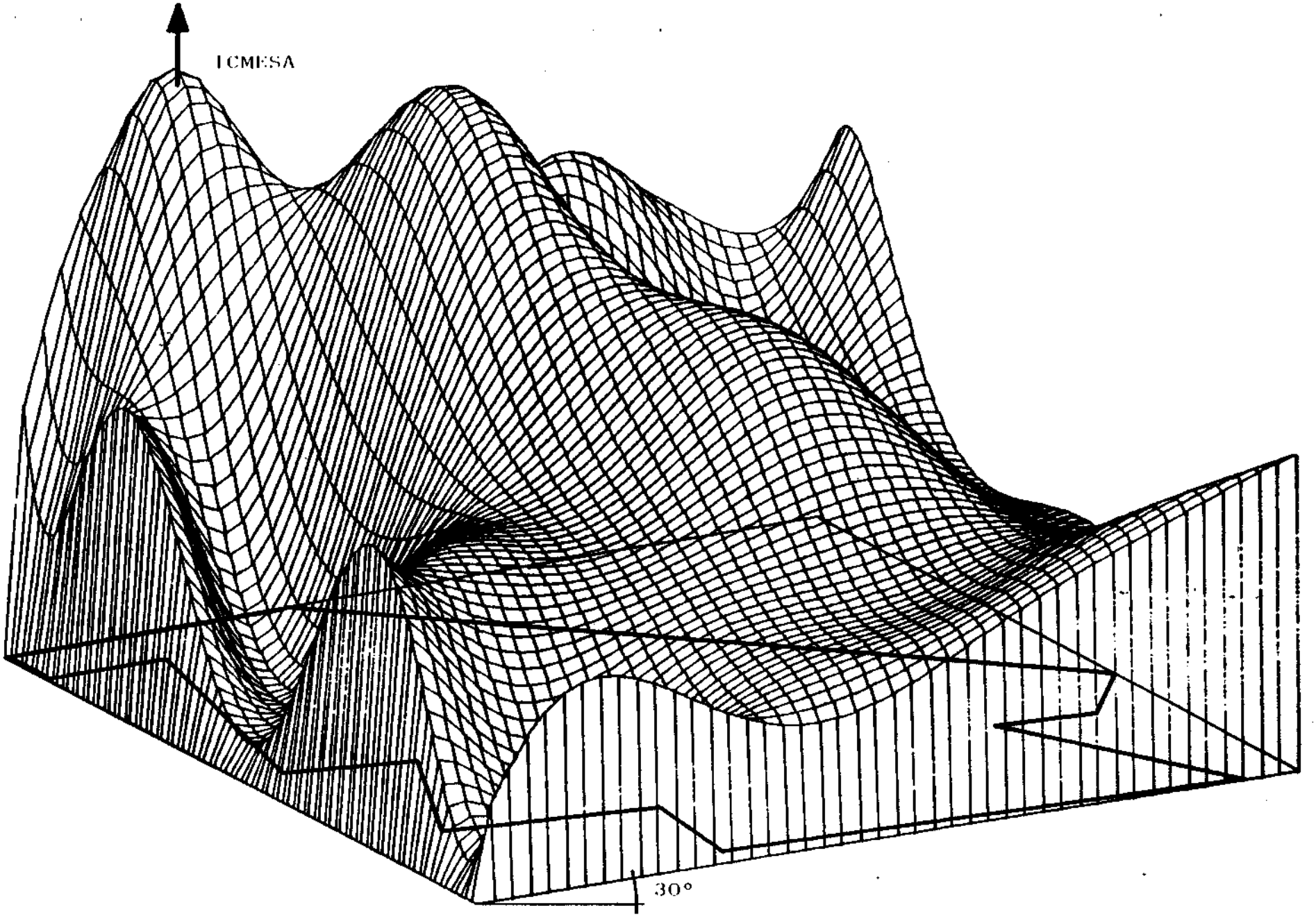

Fig. 5.1 : Rotation angle + 30°

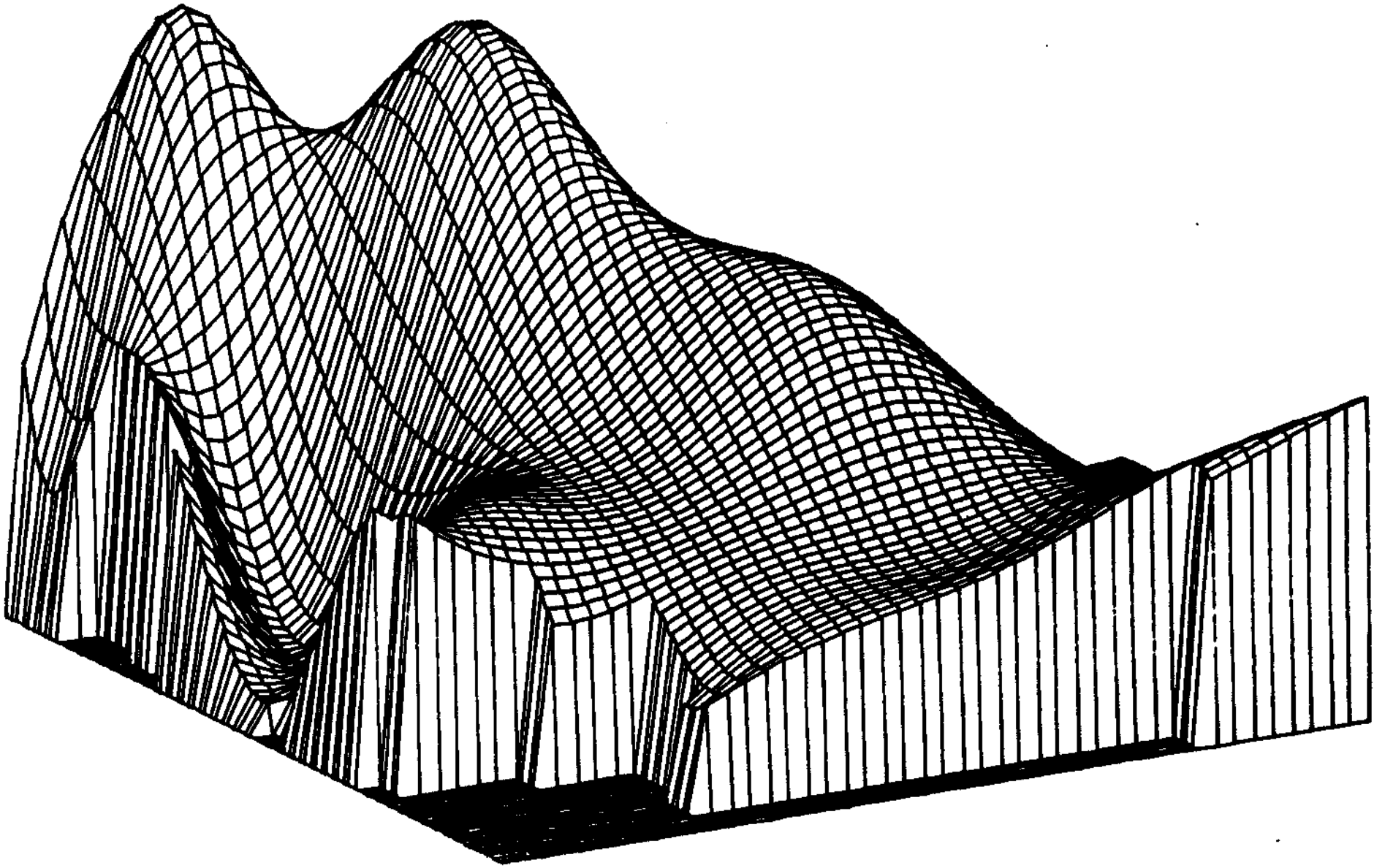

Fig.5.1 : Rotation angle + 30°
This grhafh is limited to A zone borders

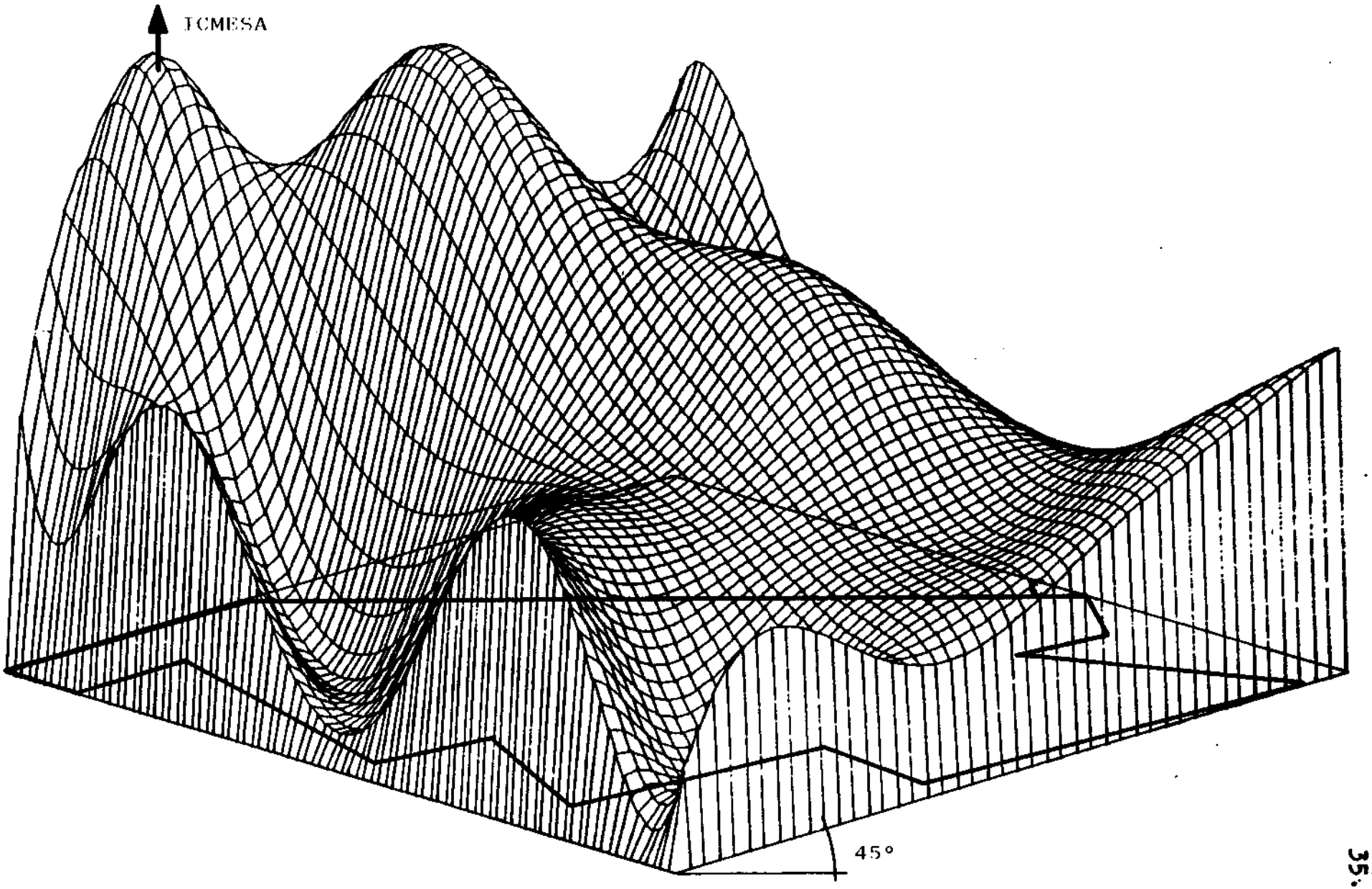

Fig. 5.2: rotation angle 45°

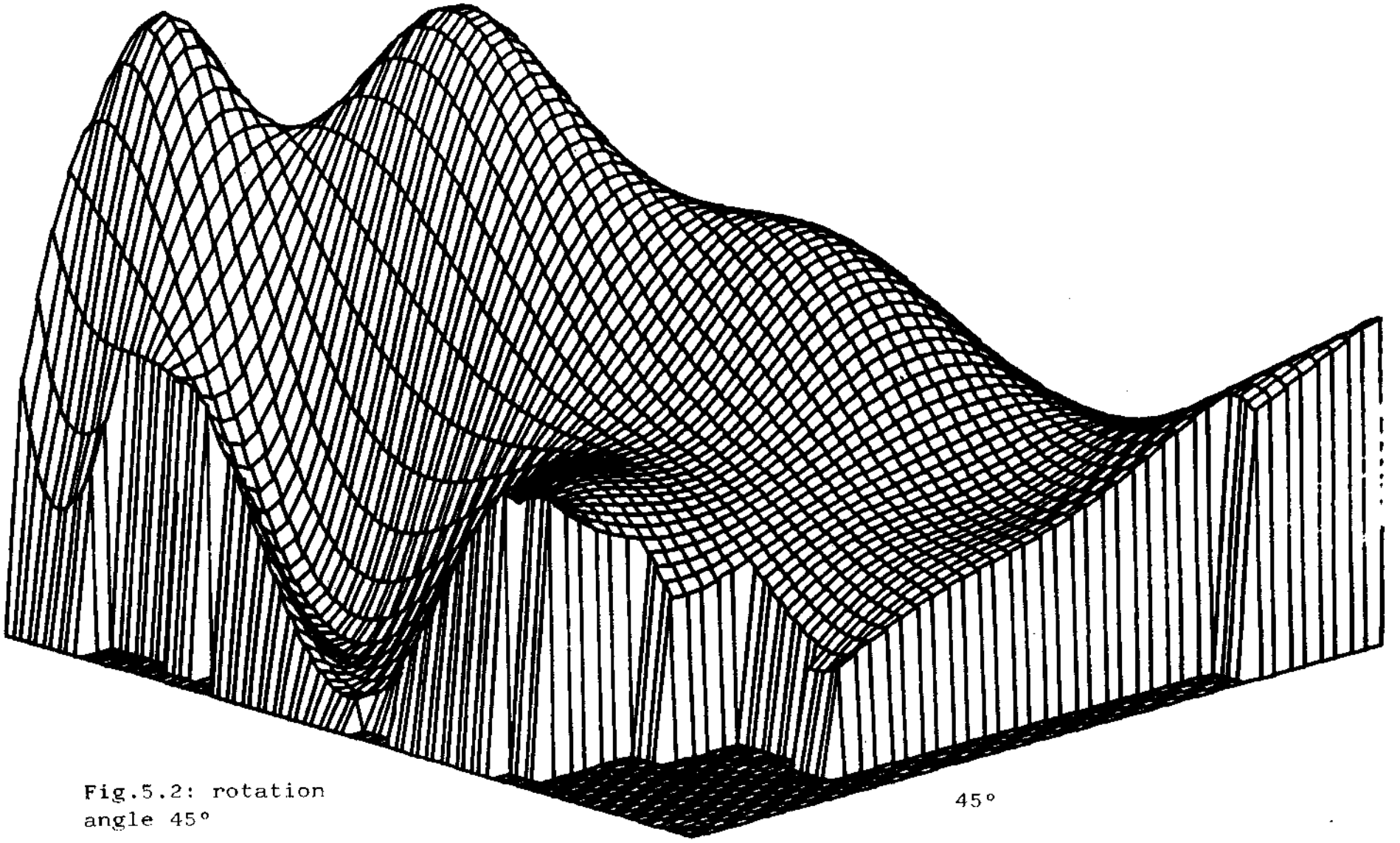

Fig.5.2: rotation
angle 45°

This grafh is limited to A zone bordens

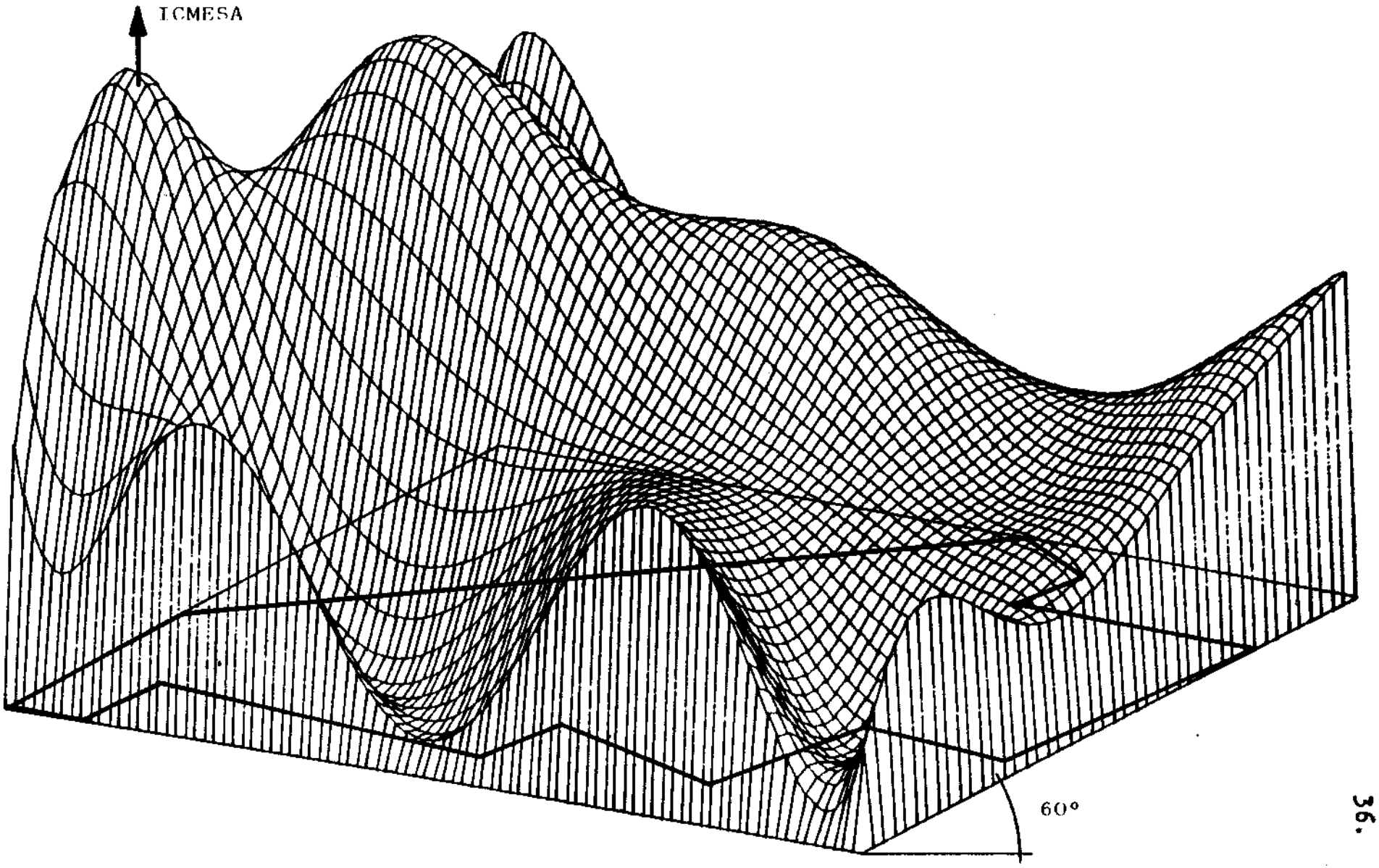

Fig. 5.3 : rotation angle 60°

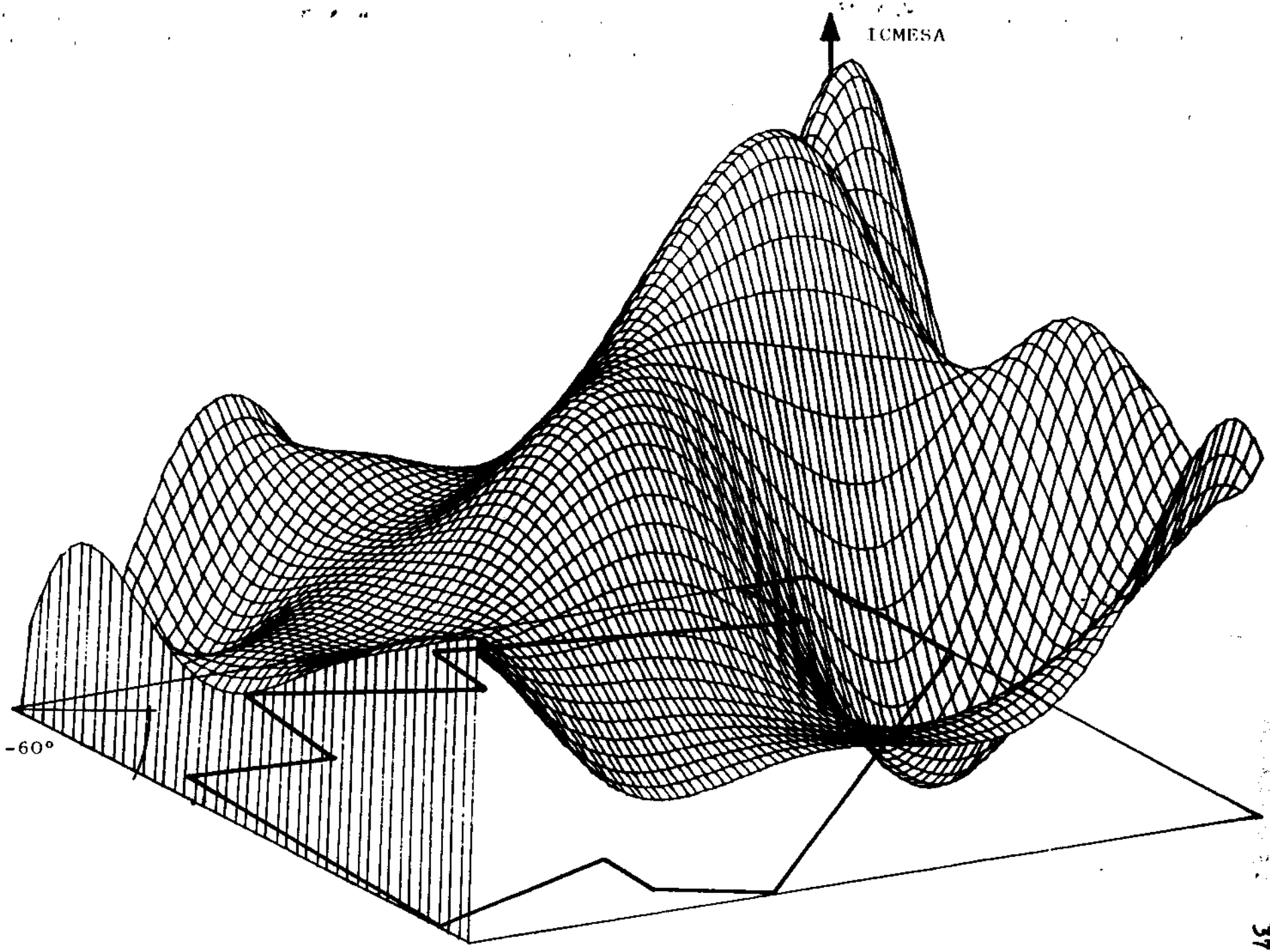

Fig. 5.4: rotation angle -60°

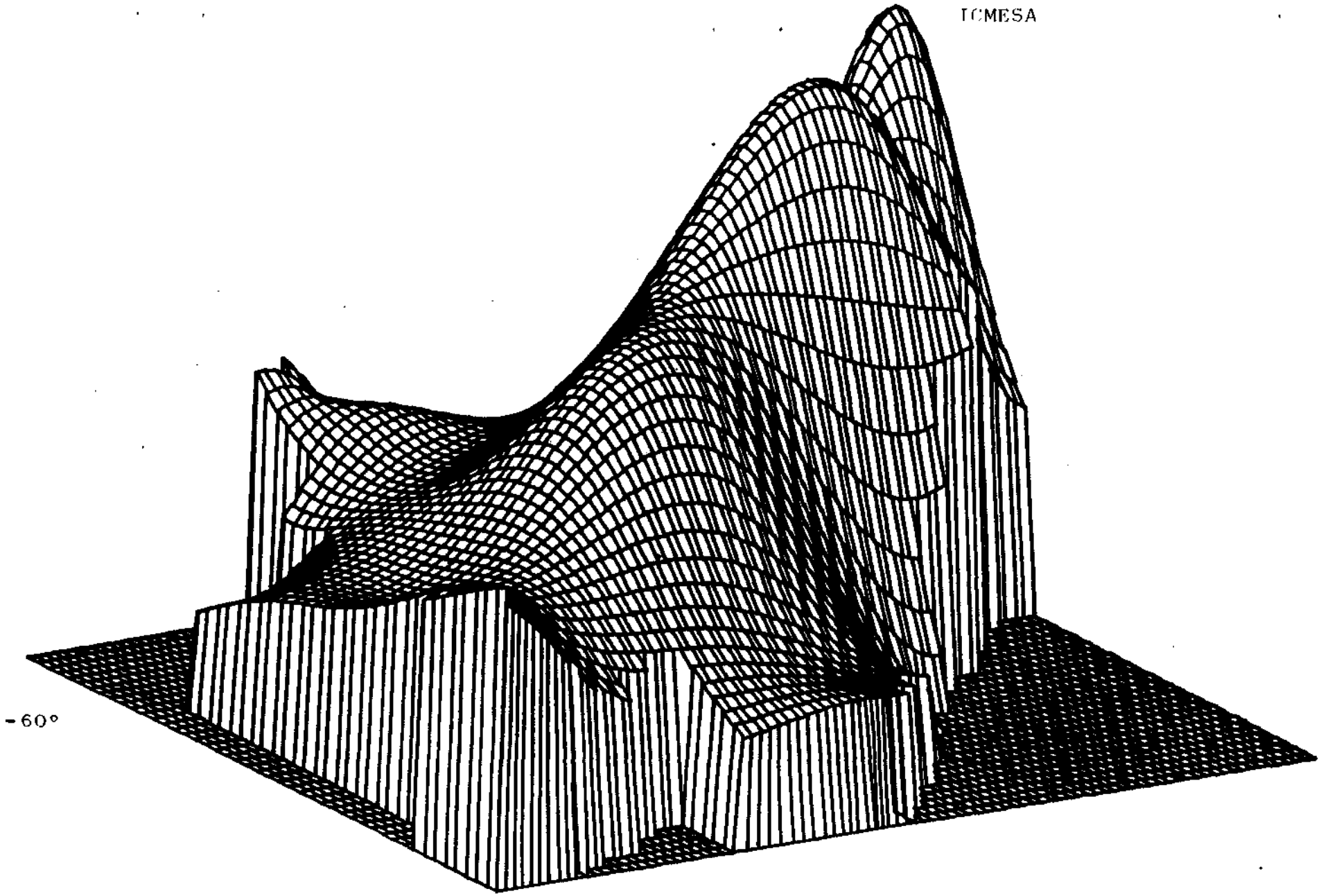

Fig.5.4: rotation angle - 60°

This gr~aph is limited to A zone borders

TCDD= 3.720 ICMESA	TCDD= 4.104		
TCDD= 1.741	TCDD= 5.081	TCDD= 2.102	TCDD= 0.312
TCDD= 0.684	TCDD= 4.116	TCDD= 3.589	TCDD= 1.185
TCDD= 2.502	TCDD= 2.911	TCDD= 3.811	TCDD= 1.702
TCDD= 2.890	TCDD= 2.233	TCDD= 2.657	TCDD= 2.167
	TCDD= 2.015	TCDD= 2.371	TCDD= 2.752

Table 5.3: Average concentrations of the logarithm of TCDD

6. ACKNOWLEDGEMENTS

We thank all the personnel of the Institute of Nuclear Physics - University of Pavia - and of INFN - Sezione di Pavia for their collaboration.

In particular Mr. F.Bossi and G.Fumagalli for the excellent support in the use of the computers and of the program libraries.

REFERENCES

1) G.Belli, S.Cerlesi, S.P.Ratti: "Valutazione della quantità di Diossina depositata al suolo entro la zona A" IFNUP/RL 14/81

2) G.Belli, G.Bressi, S.Cerlesi, S.P.Ratti: "The chemical accident at Seveso (Italy):statistical analysis in regions of low contamination". (CHEMOSPHERE 12, 517-521 1983).

3) R.Brun, M.Mansroul, H.Wind: "MUDIFI - Multi Dimensional FIT Program".
CERN Report DATA Handling Division

4) K.S.Kunz: "Numerical Analysis" McGraw-Hill, N.York (1957)

5) Ch.Jordan: "Calculus of Finite Differences"
Chelsea, N.York (1950)

6) Monografie di Matematica Applicata CNR
G.Vitali, G.Sansone: "Moderna teoria delle funzioni di variabile reale" Zanichelli
Parte II: "Sviluppi in serie di funzioni ortogonali"
Sansone 1946

7) a - D.Shepard: "Proc. XXIIIrd A.C.M. National Conf. Las Vegas" (1968) p.517
b - S.A.Gustafson, K.O.Kortanek, J.R.Sweigart
J.Appl.Meteor. 16, 1243 (1977)

c - W.R.Goodin, G.J.McRae, J.H.Seinfeld
J.Appl.Meteor. 18, 761 (1979)
d - H.R.Glahn
J.Appl.Meteor. 20, 88 (1980)

8) G.Belli, G.Bressi, E.Calligarich, S.Cerlesi, S.P.Ratti: in "Chlorinated Dioxin & related compounds: impact on environment" (Ed. O.Hutzinger et al. Pergamon Press, Oxford, N.York) 1982 p. 155

9) R.Brun, I.Ivanchenko, P.Palazzi: "Hbook User's Guide, version 3.0" CERN Report DATA Handling Division

10) Harold V.McIntosh: "SURFAC"

11) S.Cerlesi: "10 luglio 1976: incidente ICMESA" (Thesis Univ. of Pavia 1979).

12) P.Sartori: "Analisi statistica della distribuzione di diossina fuoriuscita dal reattore chimico dell'Icmesa" (Thesis Univ. of Milano 1980 Unpublished).

13) G.Belli, G.Bressi, E.Calligarich, S.Cerlesi, S.P.Ratti: in "Chlorinated Dioxins & related compounds: impact on environment" (Ed. O.Hutzinger et al. Pergamon Press, Oxford, N.York) 1982 p. 137

14) E.Milani: "Distribuzione di diossina sul terreno inquinato attorno a Seveso" (Thesis Univ. of Pavia 1981 Unpublished).