Item ID Number	02421
Author	Belli, G.
Corporate Author	
Report/Article Title	Typescript: The Chemical Accident at Seveso (Italy): Statisitical Analysis in Regions of Low Contamination
Journal/Book Title	
Year	0000
Month/Day	
Color	☐
Number of Images	8
Descripton Notes	

THE CHEMICAL ACCIDENT AT SEVESO (ITALY):

STATISTICAL ANALYSIS IN REGIONS OF LOW CONTAMINATION

G.Belli[1], G.Bressi[2], S.Cerlesi[3], S.P.Ratti[2,4].

1) Istituto Tecnico Industriale Pavia, Italy
2) Istituto Nazionale di Fisica Nucleare, Sezione Pavia, Italy
3) Ufficio Speciale della Regione Lombardia, Seveso (Mi), Italy
4) Istituto di Fisica Nucleare dell'Università di Pavia, Italy

1. Introduction

In this report we provide a normalized methodology to allow a significant statistical evaluation of the contaminant distribution in lightly polluted regions; in particular the 1980 sampling campaign in zone R is studied with an adeguate rationale(1).
The analysis of the pollution in regions of low contamination is more sensitive to the detailed mechanism responsible for the deposition of the contaminant in a given point and is most strongly affected by the sensitivity of the instruments used for the analytical treatment.

2. Feature of the contaminant's distributions.

In a previous paper[2] we have shown that a Poisson distribution is not inadeguate to describe the "local" behaviour of the contaminant.
On the other hand the most general mechanism responsible for the falling of TCDD[3] droplets on the ground is bound by the following conditions:

A - The amount of TCDD in a sample (having an area $S=70\ cm^2$) depends upon the deposition on S of a finite number n_i of droplets, each of which has a variable content in TCDD. Thus, the average amount X_i of TCDD in the sample i depends on :
  1 - the number n_i of droplets in that sample of area S_i
  2 - the "TCDD content" x_j of that single droplet j: i.e.

$$X_i = \sum_{j=i,n_i} x_j \qquad (1)$$

B - The values of n_i and x_i depend on the "TOTAL AMOUNT OF TCDD" included in the local contaminating cloud causing the fall-out.

C - The value X_i measured is to be considered as the cumulative value at infinite time $t = \infty$; i.e. the sum (or the integral) of very many tiny contributions ζ_i fallen between $t = 0$ and $t = \infty$.

Discussing the reasons why different droplets may have different TCDD content, as well as different dimensions is not a matter of concern in the present paper. Nonetheless the mathematical consequences of the above hypotheses are clear: the observed value X_i of TCDD is the sum of finite number of contributions ζ_k:

$$X_i = \Sigma_k \zeta_k \qquad (2)$$

where each ζ_k is the amount of TCDD deposited in a given finite time interval Δt_k. Here the ζ_k's are not independent as they are proportional to the total TCDD content of the local cloud, content which is decreasing obviously with time. Therefore:

$$\zeta_i = A - \alpha X(t_i) \qquad (3)$$

where $X(t_i)$ is the amount of TCDD in the sample at time t_i; A is the TCDD amount at t = 0 and α a constant. Due to the linear relationship (3) the Central Limit Theorem[4] implies that X follows a logonormal distribution i.e. Y = ln(X) (X = TCDD in $\mu g/m^2$) is gaussian distributed. This theorem has general validity, however,in regions of low contamination,the effects of wild possible fluctuations are more apparent than in highly polluted regions where, being large the amount of TCDD present anyhow,the fluctuations are more limited. It is worth underlying that adopting a "distribution in ln x" implies a "loss in resolution" of the analysis since a value (x_1 = 1000 x_o), one thousand time larger than the average value has a log value (ln x_1) only 7 units away from the average (ln x_o). In this report, therefore,we adopt a phenomenological description using the behaviour in the heavily contaminated area (zone A) as a calibration of the poorer information in the low-contamination area (zone B and zone R).

3. Check of the distribution functions

Fig.s 1 show the experimental distributions of the TCDD content (in $\mu g/m^2$) in zone A (1976/77 fig. 1a) and zone R (1980 fig. 1b). The two distributions display some common features although they come from different data samples: one from a highly polluted zone investigated immediately after the accident during 1976 and 1977; the other from a slightly polluted zone investigated in 1980, four years later. The distributions shown can be reproduced by a common curve f (z), function of the variable Z = x/x_{av} defined as the ratio between the observed contamination x and its average value x_{av} (in spite of the fact that all data of fig. 1b would be well contained in the first bin of fig. 1a!). The function f(z) is of the type:

$$f(z) = A(\exp(-\alpha z) + B\exp(-\beta z)) \qquad (4)$$

where: A = 4.27, B = 0.17, α = 18.07, β = 0.95.

The curve gives a good fit even for the data from zone R, except for x values close to zero. During the different surveys,the level of TCDD content (in $\mu g/m^2$ in this case) may vary giving rise to a threshold below which its value cannot be neither measured nor detected. The sample found in such a condition is classified as N.V. i.e. as "not visible", "not measurable" or "not detectable", the reason for that being twofold: either because that sample contains no contaminant or because the contaminant is contained below the instrument's threshold. As a matter of principle the "detection threshold" is decreasing with the time passing and with the improvement of the analytical methods.

In the histogram of fig. 1b the loss due to the N.V. samples is clearly seen.

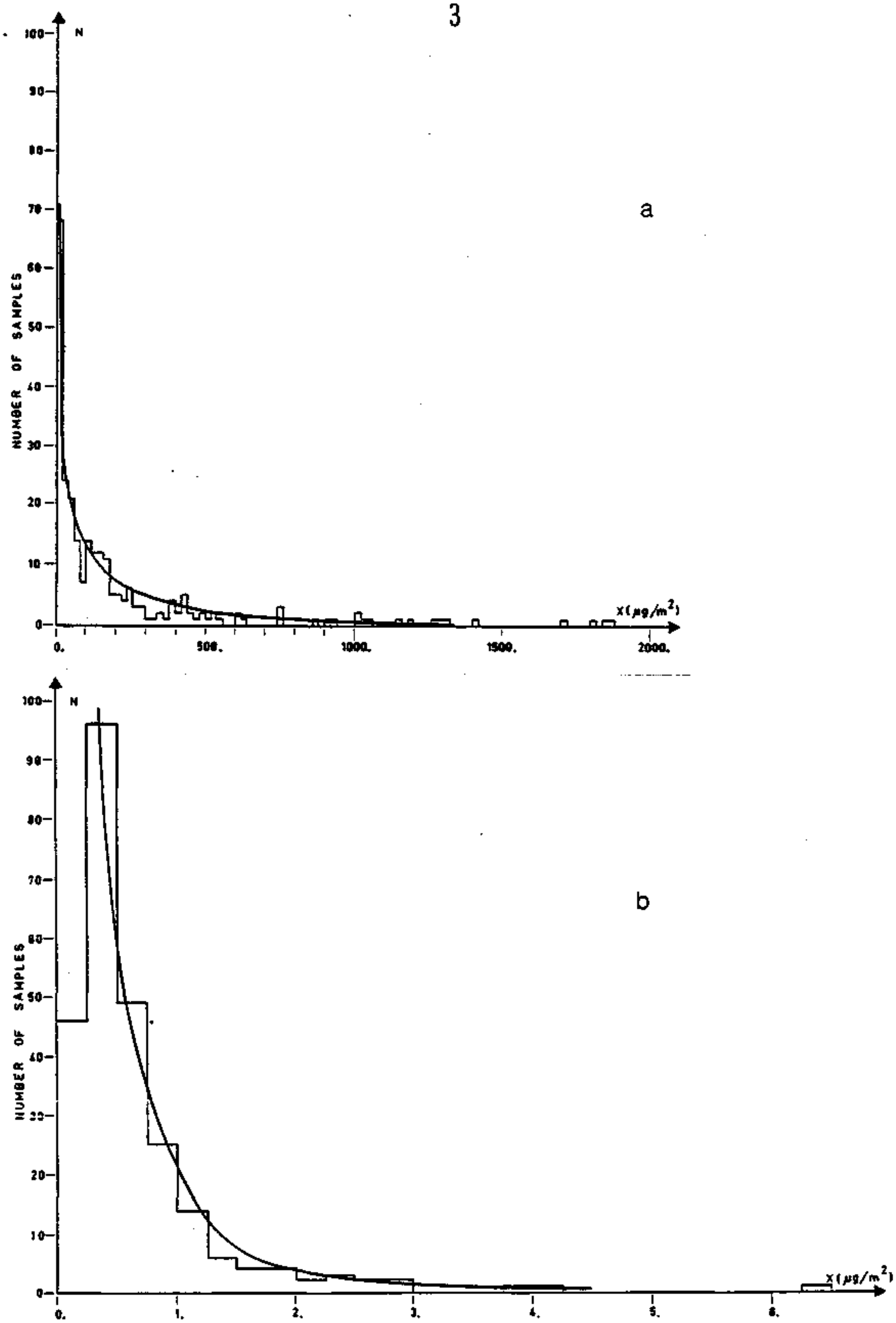

Fig. 1: Distributions of two contaminated samples:
a) HIGH contamination zone A 1976/77;
b) LOW contamination zone R 1980.
The full curves rapresent f(z) given by formula (4)

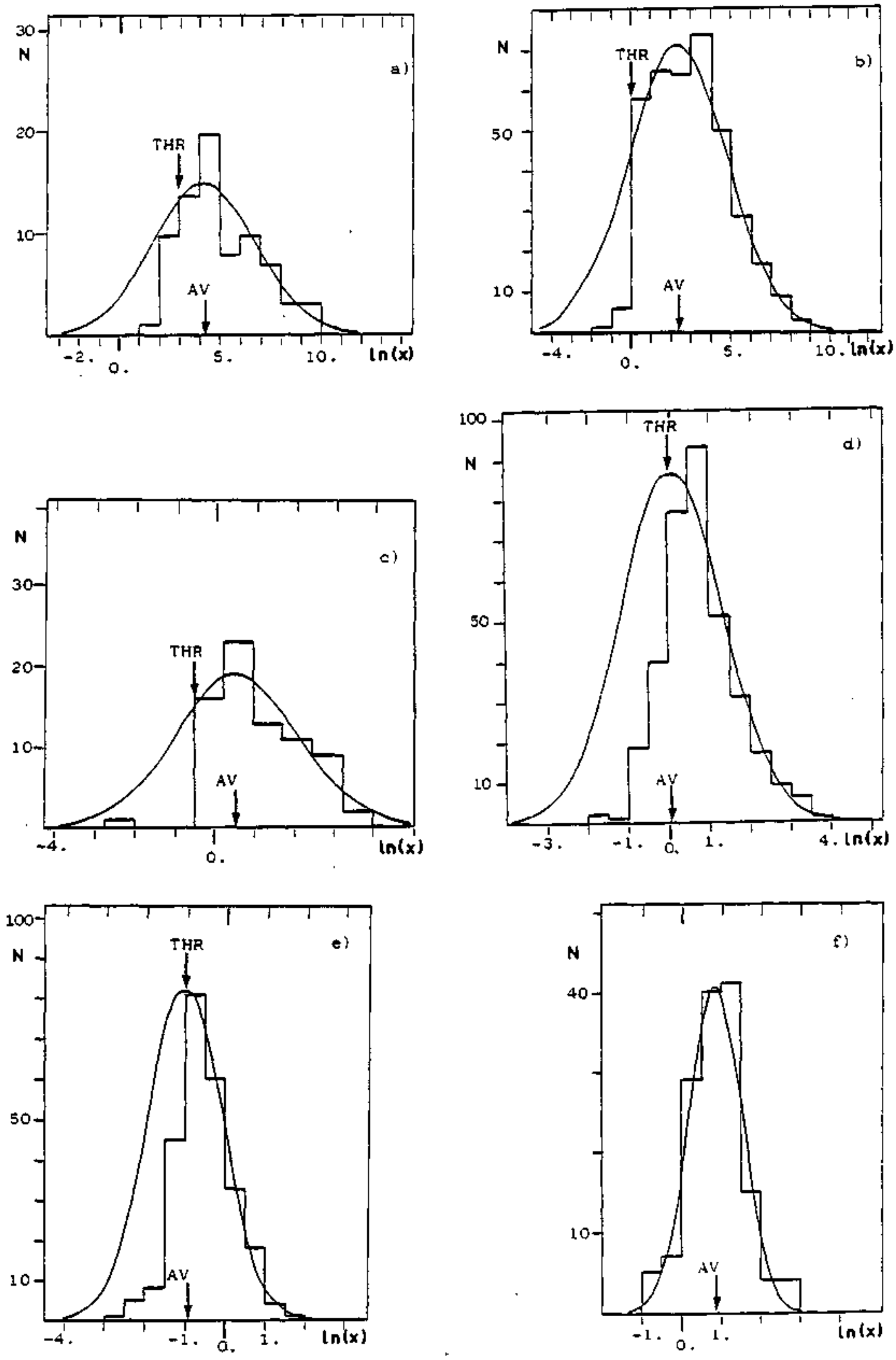

Fig.2: Distribution of Y=ln(TCDD) in different zones:a) A 1976; b) A 1976/77; c)POLO 1977; d)(B+R)1976/77; e)R 1980; f)B 1980/81.Full lines: the gaussian logonormal distribution with parameters riported in Tab.I. Average values indicated by arrows on the abscissa and threshold values by arrows on gaussian curves.

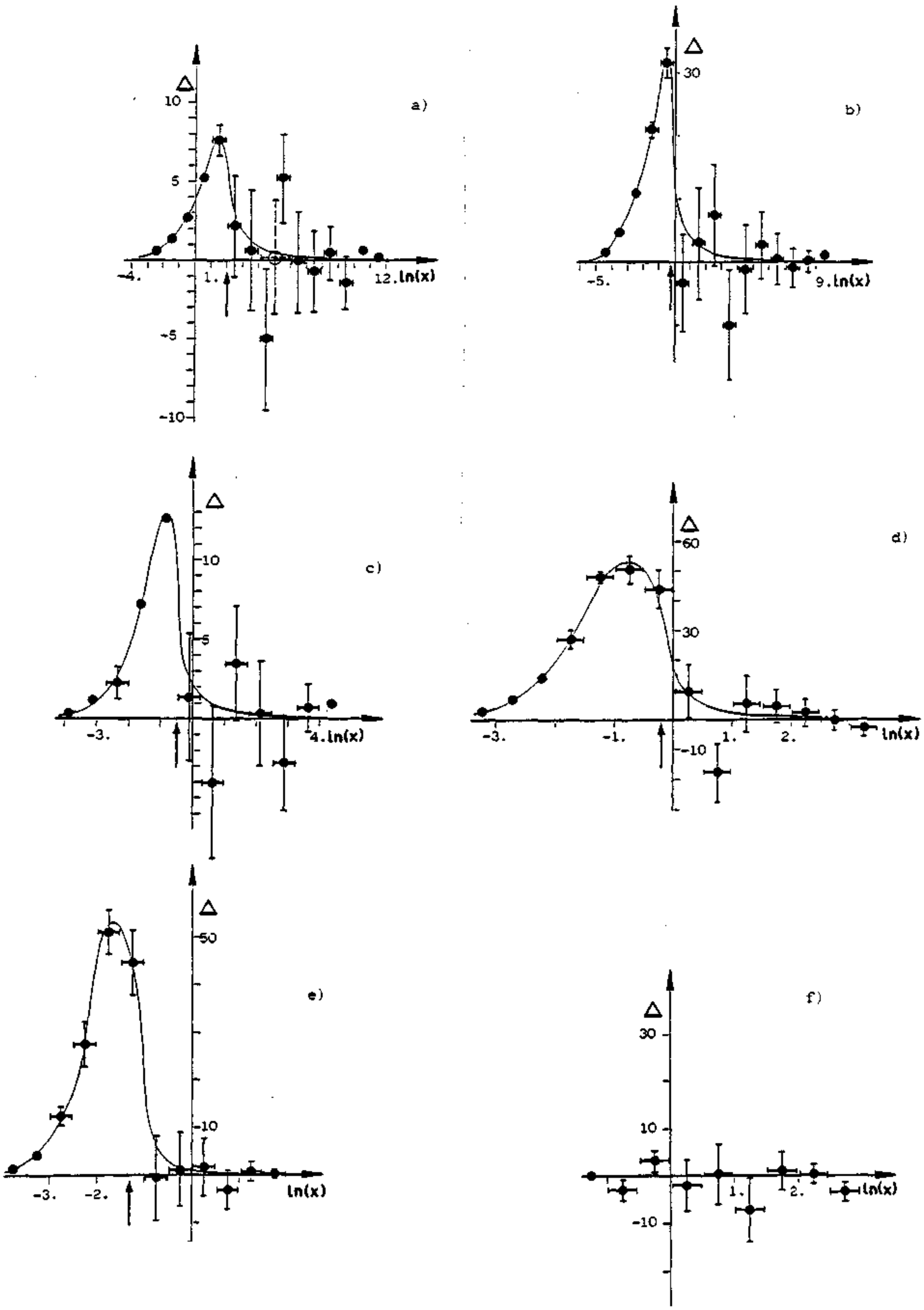

Fig.3: Distributions of the fluctuations around the logonormal curves, for the same samples of
... best estimate of the threshold value $\ln(x_{TH})$

4. The logonormal distributions.

An additional fact visible both in fig.s.1a and 1b is that the function f(z)given by (4) does not account properly for the tails of large TCDD values.This is due to the boundary conditions discussed in the previous section and forces the use of the log value to recover an important reproducibility of the different data.
Fig.s2 give the distributions of the variable Y = ln(X) (X = TCDD $\mu g/m^2$): they are clearly gaussian. In fig.s2 it is clear that in the lower part of the distributions (low TCDD values) there is a definite loss of events,due to the effect introduced by "N.V."events, which cannot obviously be plotted; indeed any measuring device has a threshold which fluctuates within the resolution of the instrument. This implies that, for a given threshold X_{th} of the instrument, a sample with TCDD content X_b greater then X_{th} has a non-zero probability of being classified amoung the "N.V." events, mostly when X_b is fairly close to X_{th}.
Since we know that the distributions are logonormal, we fit the gaussian curves to the distributions of fig.s 2 using only Y values larger than any possible "threshold limit" easy to find. In fig.s 2 the limits adopted to fit the data are indicated (top arrows) and the gaussian curves fitted are drawn. The numerical results obtained are collected in Table 1.

Table 1: FIT WITH GAUSSIAN DISTRIBUTIONS TO ln(X)

ZONE	YEAR	AVERAGE VALUE of Y = lnX	STANDARD DEVIATION σ (y)	$P(\chi^2)$
A	1976	4.18	2.5	25%
A	1976/77	2.32	2.47	85%
B + R	1976/77	0.08	1.23	20%
POLO	1977	0.54	1.54	50%
R	1980	- 0.98	0.92	90%
B	1980/81	0.85	0.64	40%

Interesting enough,in fig. 2f reproducing the most recent investigation using the most advanced instrument in a region rather contaminated (zone B, 1980/81) no threshold is clearly visible and the gaussian gives a very good fit to all data. To clearly localize the "detection threshold" we search for the Y value at which the discrepancy between the theoretical logonormal distribution and the real data becomes systematically relevant.
In fig.s 3 the values of Δ are:

$$\Delta = Y_{TH} - Y_{EXP}$$

the differences between theoretical and experimental values are plotted for the samples of fig.s 2.

The values of Δ become clearly positive to the left of the threshold limit, and compatible with zero to the right of it. In fig. 3b referring to the 1976/77 data in zone A, Δ drops around the value ln (TCDD) = -0.25 or X (TCDD) = 0.78 $\mu g/m^2$.
In fig. 3e (1980 data in zone R) the drop is less prominent and occurs at ln (TCDD) = -1.25 or X (TCDD) = 0.3 $\mu g/m^2$.
These two values recovered by the present method are exactly the sensitivity limits declared by the lavoratories which performed the chemical analyses.

5. Uncontaminated samples in regions of low contamination.

In fig.s 2 the areas between the experimental histograms and the theoretical curves give the best estimate of the number of events lost because the contaminant was present below the threshold limit. If the number of N.V. events is larger than the above, the remaining N.V. events can be considered as statistically uncontaminated.
Table II summarizes the numerical values of this investigation.

Table II

ZONE	THRESHOLD LIMIT		NUMBER OF EVENTS			3 STANDARD DEVIAT. VALUE	
	ln X	X	LOST	N.V.	NON CONTAM.	ln X	X
A 1976	3.0	20.08	19 ± 4	32	13 ± 5	0.43	1.53
A 1976/77	0.0	1.0	60 ± 8	37		-1.38	0.25
POLO 1977	-0.5	0.6	26 ± 5	60	34 ± 5	-1.77	0.17
B+R	0.0	1.0	195 ± 14	391	196 ± 14	-1.76	0.17
R 1980	-1.0	0.37	125 ± 11	503	378 ± 11	-2.36	0.09
B 1980/81				8	8	-0.11	0.89

What we learn from fig.s 2 and Table II is very instructive.
Each gaussian curve may give a 3 Standard Deviation (99.96% Confidence Level) indication that the statistically uncontaminated samples contain less that the amount of contaminant listed in the last column of Table II.

6. Conclusions.

As expected from the statements of Section 2 and from the Poisson-like origin of the distributions, the average values and the standard deviations are correlated to one another. We can therefore use all the statistical information available in fig.4 to search for the regression line between the two variables, by fitting a linear equation:

$$\sigma(X) = \alpha X_{av} + \beta \qquad (5)$$

(with: α = 0.379 and β = 1.178).
From fig.4 we learn that we cannot do better than σ = 0 (i.e. zero fluctuation!) Therefore the value ln(x)=-3.106 or x = 0.045 $\mu g/m^2$ is the minimum possible detectable amount of contaminant with the methods presently in use as well as the maximal possible sensitivity of the instrument.

Furthermore knowing approximately the average value of X_{av} in a given region, fig.4 allow to infer the corresponding value of $\sigma(X_{av})$ i.e. to know all the parameters of the logonormal distribution. From this information, risk limits may be set. Such subject however cannot be deald with in this short report.

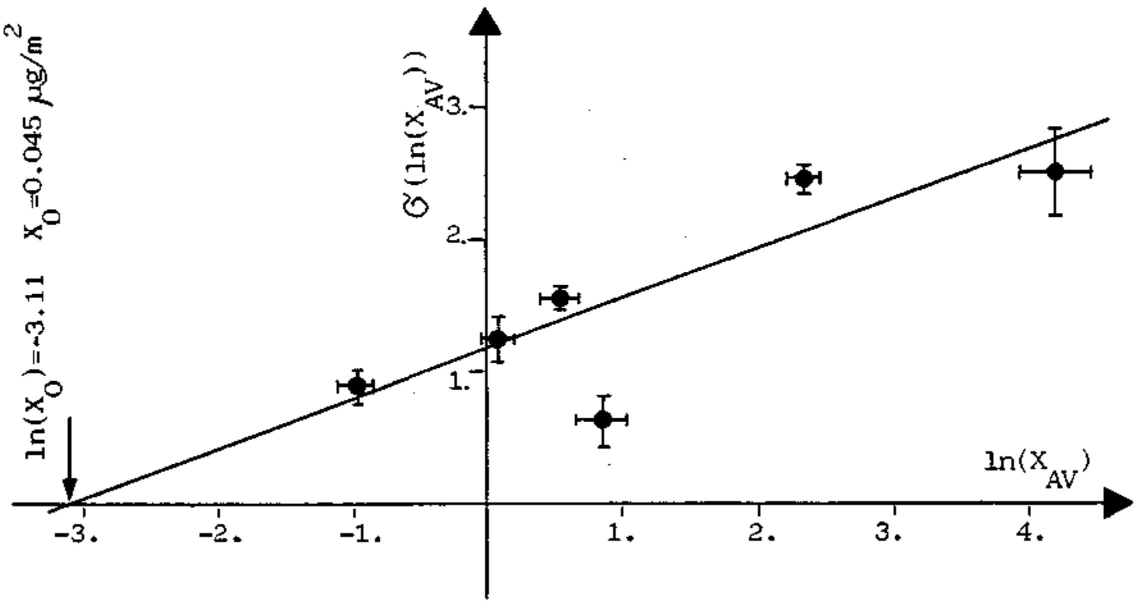

Fig.4: Summary of the maximum statistical available information: dispersion $\sigma(\ln(x)$ vs. $\ln(x_{av})$ for all samples. The straight line is the regression line given by formula(5). The intercept $\ln(x_o)$ gives the maximum sensitivity obtainable at present in analytical methods.

References

1. G.Bressi, S.Cerlesi, S.Ratti: "A possible analysis of the TCDD distribution in a low polluted region" University of Pavia Report IFNUP/RL 07/81.
2. G.Belli, G.Bressi, E.Calligarich, S.Cerlesi, S.Ratti in "Chlorinated Dioxins & related compounds: impact on environment" (Ed.O.Hutzinger et al. Pergamon Press, Oxford, N.York) 1982 p.137 and 155.
3. TCDD stand obviously for. 2,3,7,8 - Tetrachlorodibenzo para dioxin.
4. H.Cramer: Mathematical Methods of Statistics, Princeton University Press (1974) p.213, p.218.