ZJE — 264 1982

J. Liška, M. Světlík, K. Sláma

THE USE OF CLUSTER ANALYSIS METHOD FOR THE LOCALIZATION OF ACOUSTIC EMISSION SOURCES DETECTED DURING THE HYDROTEST OF PWR PRESSURE VESSELS

ŠKODA WORKS
Nuclear Power Construction Division, Information Centre
PLZEŇ, CZECHOSLOVAKIA

ZJE - 264 1982

J. Liška, M. Světlík, K. Sláma

THE USE OF CLUSTER ANALYSIS METHOD FOR THE LOCALIZATION OF ACOUSTIC EMISSION SOURCES DETECTED DURING THE HYDROTEST OF PWR PRESSURE VESSELS

Reg. č. ÚVTEI 73 307

ŠKODA WORKS
Nuclear Power Construction Division, Information Centre
PLZEŇ, CZECHOSLOVAKIA

ABSTRACT

The acoustic emission method is a promising tool for checking reactor pressure vessel integrity. Localization of emission sources is the first and the most important step in processing the emission signals. The paper describes the emission sources localization method which is based on cluster analysis of a set of points depicting the emission events in the plane of coordinates of their occurrence. The method is based on constructing the minimum spanning tree on this set of points and its partition into fragments corresponding to clusters of points. Furthermore, the paper considers the laws of probability distribution of the minimum spanning tree edge length for one as well as several clusters with the aim of finding the optimum length of the critical edge for the partition of the tree. Practical application of the method is demonstrated on localizing the emission sources detected during hydrotest of a pressure vessel used for testing the reactor pressure vessel covers.

1. INTRODUCTION

The method of acoustic emission appears to be a promising tool for checking the integrity of large pressurized devices as, e.g. reactor pressure vessels. The method has been developed so far that it is now possible to use the method in industrial scale. In the case of using the acoustic emission method for the testing of reactor pressure vessel integrity, the following areas of application come into practical consideration:

- testing the vessel during the hydrotest performed in the framework of post-manufacturing tests in the testing room of the manufacturing plant
- testing the vessel during the hydrotest performed in the framework of reactor in-service testing in the nuclear power plant
- continuous testing of the vessel during normal operation of the reactor.

A very important problem involved in processing the data obtained during the acoustic emission test is localizing the emission sources, i.e. determining the localities of their activity. It may be said that proper localization of the emission sources underlies further processing and the way of its performing impacts fundamentally the quality of the whole work.

The report presents the method of localizing the emission sources as it is applied in evaluating the acoustic emission tests performed in the framework of post-manufacturing tests of reactor pressure vessels in our country. The method belongs into the group of cluster analysis methods and is based on a partition of the minimum spanning tree created on a set of points.

2. THE METHOD OF LOCALIZATION OF EMISSION SOURCES

The emission events detected during the test may be depicted as points in the x - y coordinate plane of events occurrences on the developed surface of the vessel. These events result from the activity of the sources of either useful or disturbing signal.

These points may be a priori assumed to create clusters which correspond to emission sources. Consequently, the task of finding the localities of action of these sources is equivalent to the task of finding the clusters.

2.1. THE METHOD OF CLUSTER ANALYSIS BASED ON PARTITION OF THE MINIMUM SPANNING TREE CREATED ON A SET OF POINTS

The concept of minimum spanning tree is known from the theory of graphs. It is a tree having the minimum weight of all the trees made on a set of points, the weight being expressed by means of the lengths of edges connecting the individual points. The minimum spanning tree is a suitable tool for characterizing the structure of a set in a symbolic way. It is possible to transform the minimum spanning tree into the shape of a dendrogram in which are recorded the points of the set as well as the lengths of the edges connecting these points to the minimum spanning tree. Fig. 1 presents a simple example of a minimum spanning tree and of a corresponding dendrogram.

The principle of determining the clusters consists in finding a limiting value of the edge of the tree d_k (so called critical edge). By breaking all edges d for which d exceeds d_k it is possible to divide the minimum spanning tree into fragments, the points contained in individual fragments being the points of found clusters. Partition of the minimum spanning tree on a level of selected critical edge is also shown in fig. 1.

The problems of finding a suitable criterion for determining the length of the critical edge, d_k, and of an adequate technique of dividing the minimum spanning tree are dealt with hereinafter.

2.2. THE PROBABILITY DISTRIBUTION OF THE MINIMUM SPANNING TREE EDGE LENGTHS FOR ONE CLUSTER

Let us assume that the coordinates of the nodes of the minimum spanning tree are random variables with a uniform probability distribution. Therefore, the probability that in an arbitrary region of the plane there will be just k nodes of the tree, is given by Poisson probability distribution

$$P(k) = \frac{(\mu S)^k}{k!} e^{-\mu S}, \qquad /1/$$

where μ is a parameter of Poisson distribution determining an average amount of the number of nodes which correspond to unit area, and S denotes the size of the region.

From one node of the minimum spanning tree there cannot stem more than 6 edges. Let us have a tree node A_i and find the probability that an edge stemming therefrom will not exceed d (definition of the distribution function). Let us further define a set of points whose distance from the point A_i is d. This can be recorded in the manner as follows: $\{A_{id}: \varrho(A_{id}, A_i) = d\}$. If any point of this set were joined to the tree in the point A_i and if, moreover, the conditions for the minimum spanning tree were fulfilled, then it may be said that in no point A satisfying simultaneously relations $\varrho(A, A_i) \leq d$, and $\varrho(A, A_{id}) \leq d$ can occur a tree node $A_j \neq A_i$. The size of the area defined in this manner will be $(\frac{2}{3}\pi - \frac{\sqrt{3}}{2}) \cdot d^2 \doteq 1.23d^2$. The probability that this area contains no node of the tree is (after inserting k = 0 into relation /1/):

$$p(0) = e^{-1.23\mu d^2} \qquad /2/$$

It may be said that it is practically equal to the probability that the minimum spanning tree edge exceeds d. The distribution function may be then written in the following form

$$F(d) = 1 - e^{-1.23\mu d^2} \qquad /3/$$

The distribution function of Rayleigh probability distribution has the following general form

$$F(d) = 1 - e^{-cd^2} \qquad /4/$$

The edges of the minimum spanning tree which has been created on a set of points whose coordinates have a uniform probability distribution, have then Rayleigh distribution with a parameter $c = 1.23$.

Let us have a cluster of points whose coordinates are independent random variables with normal probability distribution and standard deviations of σ_1 and σ_2. Let us further assume that the probability distribution of the minimum spanning tree edge lengths may be approximated by Rayleigh probability distribution (the minimum spanning tree is assumed to be created on this set of points). In other words, it means the substitution of this cluster by a set of points distributed uniformly throughout a certain region. For the equivalent density of the set, denoted μ_u, it may be written

$$\mu_u = \frac{N}{S_u}$$

N denoting the number of points in the cluster while $S_u = k_0 \sigma_1 \sigma_2$ denotes the equivalent area of the cluster, the area having a shape of an allipse whose semiaxes ratio is the same as that of standard deviation. Let us furthermore introduce a parameter $h = N/\sigma_1 \sigma_2$ which will be called the cluster density. By inserting and modifying relation /4/ it will be possible to derive from it for the distribution function

$$F(d) = 1 - e^{-1.23 \cdot \frac{1}{k_0} \cdot hd^2} = 1 - e^{-khd^2} \qquad /5/$$

In order to determine the value of the coefficient k, clusters of points with various densities h (various number of points, N, and various ratios of standard deviations σ_1/σ_2) have been generated on a digital computer. It has been proved that Rayleigh distribution is a good approximation of the probability distribution of edge lengths and, in addition, that its parameter c varies in linear proportion with the cluster density, h (see fig. 2).

2.3. PROBABILITY DISTRIBUTION OF THE MINIMUM SPANNING TREE EDGE LENGTHS FOR SEVERAL CLUSTERS

Let us consider a set of points having M subsets (clusters) with a normal distribution of the coordinates of points. As has already been shown in 2.2., the probability distribution of minimum spanning trees edge lengths of individual clusters may be approximated by Rayleigh distribution.

The distribution of the i-th cluster will be given as

$$w_i(d) = 2kh_i d \cdot \exp(-kh_i d^2) \qquad /6/$$

The minimum spanning tree edge lengths of the whole set may be then written in the following shape

$$w(d) = \frac{1}{N_c} \sum_{i=1}^{M} N_i w_i(d) \qquad /7/$$

where $N_c = \sum_{i=1}^{M} N_i$, N_i denoting the number of points of the i-th cluster.

Combining /7/ and /6/ yields

$$w(d) = \frac{2k}{N_c} \cdot d \cdot \sum_{i=1}^{M} N_i h_i \cdot \exp(-kh_i d^2) \qquad /8/$$

The distribution will be Rayleigh distribution only in two cases as follows:

1) one of the clusters is a dominant one, i.e. $N_i \gg N_j \quad (j = 1, 2, \ldots M;\ i \neq j)$
2) all the clusters have equal density, i.e. $h_i = \text{const.} \quad (i = j, 2, \ldots M)$.

If the set contains clusters with different densities wherefrom no cluster is dominant, distribution w(d) differs from Rayleigh distribution. It has been found experimentally that in such a case an adequate approximation provides the distribution of Weibull. Its form is as follows:

$$w(d) = cbd^{b-1} \cdot \exp(-cd^b) \qquad /9/$$

where parameter $c > 0$ and $b > 0$.

Rayleigh distribution is a particular case of Weibull distribution for $b = 2$.

2.4. DETERMINING THE VALUE OF THE CRITICAL EDGE, AND THE METHOD OF DETERMINING THE CLUSTERS

Let us have a distribution function $F(d)$ of the minimum spanning tree edge lengths for the whole set. After partition of the minimum spanning tree on the level of the critical edge d_k there will remain unbroken $F(d_k)N_c$ edges. This value corresponds to the number of points which will remain in the clusters. It is possible to write

$$F(d_k)N_c = \sum_{i=1}^{M} F_i(d_k)N_i \qquad /10/$$

Let us denote the cluster with the highest number of points by an index M and let us choose d_k so that

$$d_k = 2d_{mM} \qquad /11/$$

where d_{mM} denotes the coordinate of the mode of the minimum spanning tree edge lengths distribution of the cluster with index M. Then

$$F_M(d_k) = \int_0^{2d_{mM}} w_M(d)dd = 0.86$$

$w_M(d)$ denoting the probability density of the Rayleigh distribution of the edges of the M-th cluster.

Relation /11/ may be called the condition of conservation of the maximum cluster.

For the purpose of simplification let us furthermore assume that the cluster density depends only on the number of its points, i.e.

$\sigma_{i1}\sigma_{i2}$ = const. for all i.

Assuming in addition that

$$N_i = \alpha_i N_M \qquad /12/$$

($\alpha_i \in \langle 0,1\rangle$),
the distribution function of Rayleigh distribution of edges for the i-th cluster may be then written as follows

$$F_i(d) = 1 - e^{-KN_M\alpha_i d^2} \qquad /13/$$

K denoting a constant.

The coordinate of the mode of this distribution will be given by relation /14/:

$$d_{mi} = \left(\frac{1}{2KN_i}\right)^{1/2} \qquad /14/$$

Combination of relations /11/, /13/ and /14/ will result in

$$F_i(d_k) = 1 - e^{-2\alpha_i} \qquad /15/$$

Relations /10/, /12/ and /15/ enable to obtain an equation for choosing the critical edge in the following shape

$$F(d_k) = \frac{\sum_{i=1}^{M} \alpha_i e^{-2\alpha_i}}{\sum_{i=1} \alpha_i} \qquad /16/$$

Provided that the distribution function of the whole set minimum spanning tree edge lengths distribution has been obtained experimentally, then the value of d for which

$$F^*(d) = F(d_k)$$

is the length of the critical edge ($F^*(d)$ denotes the minimum spanning tree edge lengths distribution function for the whole set).

We do not know a priori the coefficients α_i in /16/. Nevertheless, it is possible to assume that they are terms of a certain series. $F(d_k)$ may be determined for an arithmetic and a geometric series. Let us further denote

$$\nu = \frac{\alpha_M}{\alpha_1}$$

M and 1 denoting the cluster with the highest and the lowest number of points, respectively. Using the relations for the n-th term and for the sum of the first n terms it is possible to derive that

$$\alpha_i = \frac{M - 1 + (i - 1)\cdot(\nu - 1)}{\nu\,(M - 1)} \qquad /17/$$

for the arithmetic series, and

$$\alpha_i = \nu^{\left(\frac{i - M}{M - 1}\right)} \qquad /18/$$

for the geometric one.

Table 1 presents calculated values of $F(d_k)$ for various ν and M. Added are values of $F(d_k)$ for limiting cases mentioned in 2.3., i.e. for equal density of clusters or for one dominant cluster.

The procedure of clustering may be characterized as follows:

1. We shall create the minimum spanning tree on the set of points, determine the values of $F^*(d)$ and fit them by the distribution function of Weibull distribution. The values of parameters of this distribution contain information about the structure of the set under investigation. Nevertheless, these relations are not sufficiently known for the time being.

Value of the parameter b approaching 2 suggests that there is involved one of the limiting cases stated in 2.3., if $b<2$, it may be concluded that the set consists of clusters of different densities.

2. On the basis of the available a priori information about the set and the obtained parameter of Weibull distribution it is possible, taking advantage of table 1, to obtain $F(d_k)$. By its comparison with $F^*(d)$ we shall determine the length of the critical edge, d_k.

3. The composition of clusters will be obtained by partition of the minimum spanning tree on the level of the critical edge, d_k.

4. We shall now omit all points of the set which correspond to clusters with the maximum number of points. The composition of these clusters will be now considered for definitive. The numbers of points in the clusters resulting from the partition of the tree on a level d_k, are obtainable from relation /19/

$$N_{si} = F_i(d_k) \qquad /19/$$

In order for the greatest clusters to be omitted it is necessary to determine the limiting value of probability, $F_{im}(d_k)$. Using relations /12/, /15/ and /19/ it is possible to determine the limiting number of the points in the cluster which may be omitted. This number of points is given as

$$N_{si_m} = -0.5\frac{F_{im}(d_k)}{F_M(d_k)}\cdot N_{sM}\cdot \lg\left[1-F_{im}(d_k)\right] \qquad /20/$$

5. The procedure of clustering is now repeated for the rest of the set. The repetition of the procedure is an important step because it enables to obtain the composition of clusters even in case of larger differences in their densities. In using the method for the localization of emission sources it is of importance e.g. in such a case when the region under investigation contains an emission source whose activity markedly exceeds the activity of the other sources (it is usually a

disturbing source). The procedure may be repeated until the numbers of points of the clusters found in the sets decrease under the preset limit.

6. The final phase consits in determining the boundaries of regions of found clusters by means of the probability ellipses.

3. PRACTICAL EXAMPLE OF USING THE METHOD

The method can be exemplified using the data obtained in the course of pressurizing the vessel serving for testing the covers of reactor pressure vessels.

Fig. 3 depicts emission events in the form of points on developed surface of the vessel.

The procedure of clustering has been repeated three times. For determining the length of the critical edge we have chosen $F(d_k) = 0.64$ which, according to tab. 1, corresponds to $M = 15$, $\nu = 10$, and coefficients α_i forming a geometrical series. Of course, this assumption is only approximative because it is based only on visual evaluation of the structure of the set as shown on fig. 3, and on the value of Weibull distribution exponent obtained by fitting the experimentally found distribution function of edges of the minimum spanning tree of the set.

The results of individual steps of the clustering procedure are summarized in table 2. The probability for omitting the clusters, $F_{im}(d_k)$, is approximatively equal to 0.7. The table presents the numbers of points of found clusters which contain more than 7 points. Final composition of clusters should be considered to consist of those clusters which have been omitted from processing in the first and second step of the procedure, as well as all clusters with more than 7 points found in the third step.

Fig. 4 presents the probability ellipses for all clusters with final composition (the semiaxes of the probability ellipses equal $2\sigma_1$, $2\sigma_2$). If the probability ellipses overlap as seen in fig. 4, the points inside both ellipses belong to one cluster.

4. CONCLUSIONS

Presented method of localization of the acoustic emission sources has been tested in practice and proves to be useful. Its main advantages consist in sufficient local sensitivity and adaptivity from the viewpoint of different activity of emission sources which results in different density of the clusters.

The method of finding the clusters which is based on partition of the minimum spanning tree, enables the whole clustering procedure to be automated. This requires to formulate the criteria and techniques of evaluating the minimum spanning tree which have a more general validity and may be applied to solving various situations occurring in the practice. The first step in this direction has been made. Refining the criteria will be the aim of further investigation.

REFERENCES

/1/ Zahn C.T.: Graph-Theoretical methods for detecting and describing gestalt clusters. IEEE Transactions on computers, Vol. C-20, No. 1, 1971.

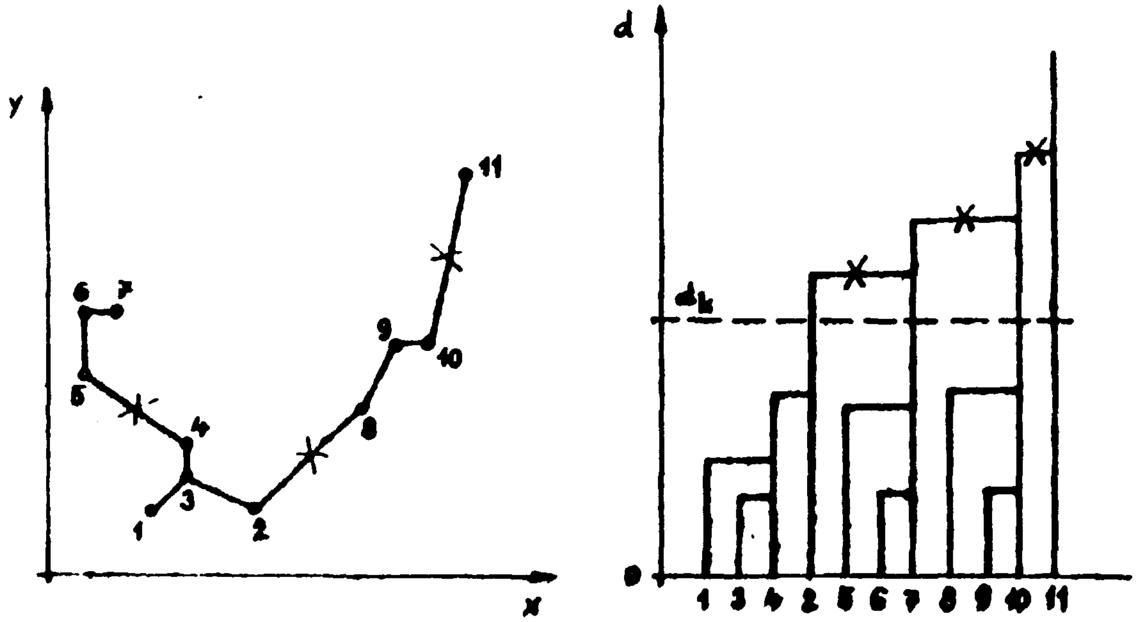

Fig. 1. Minimum spanning tree Dendrogram

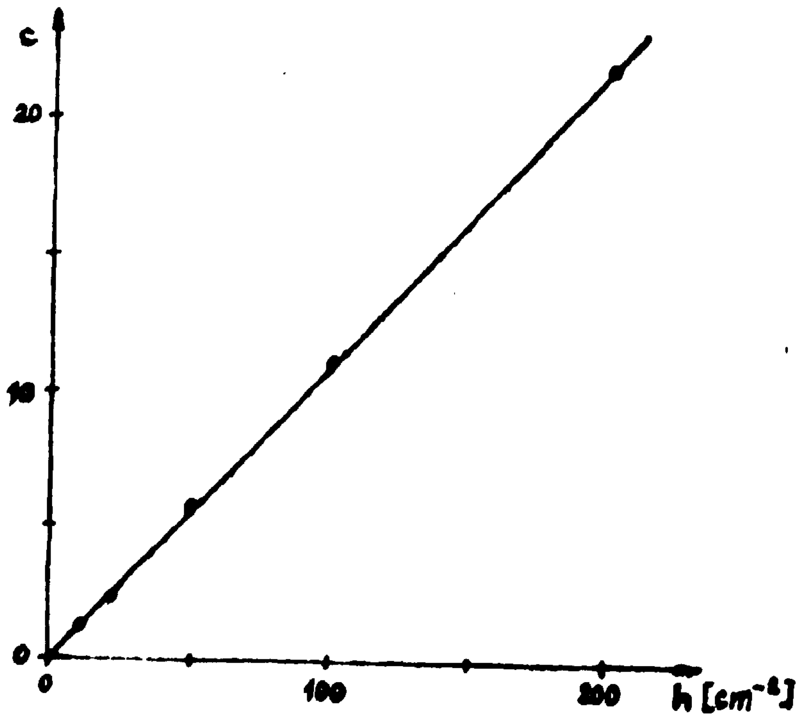

Fig. 2. Dependence of parameter c in Rayleigh distribution on cluster density h

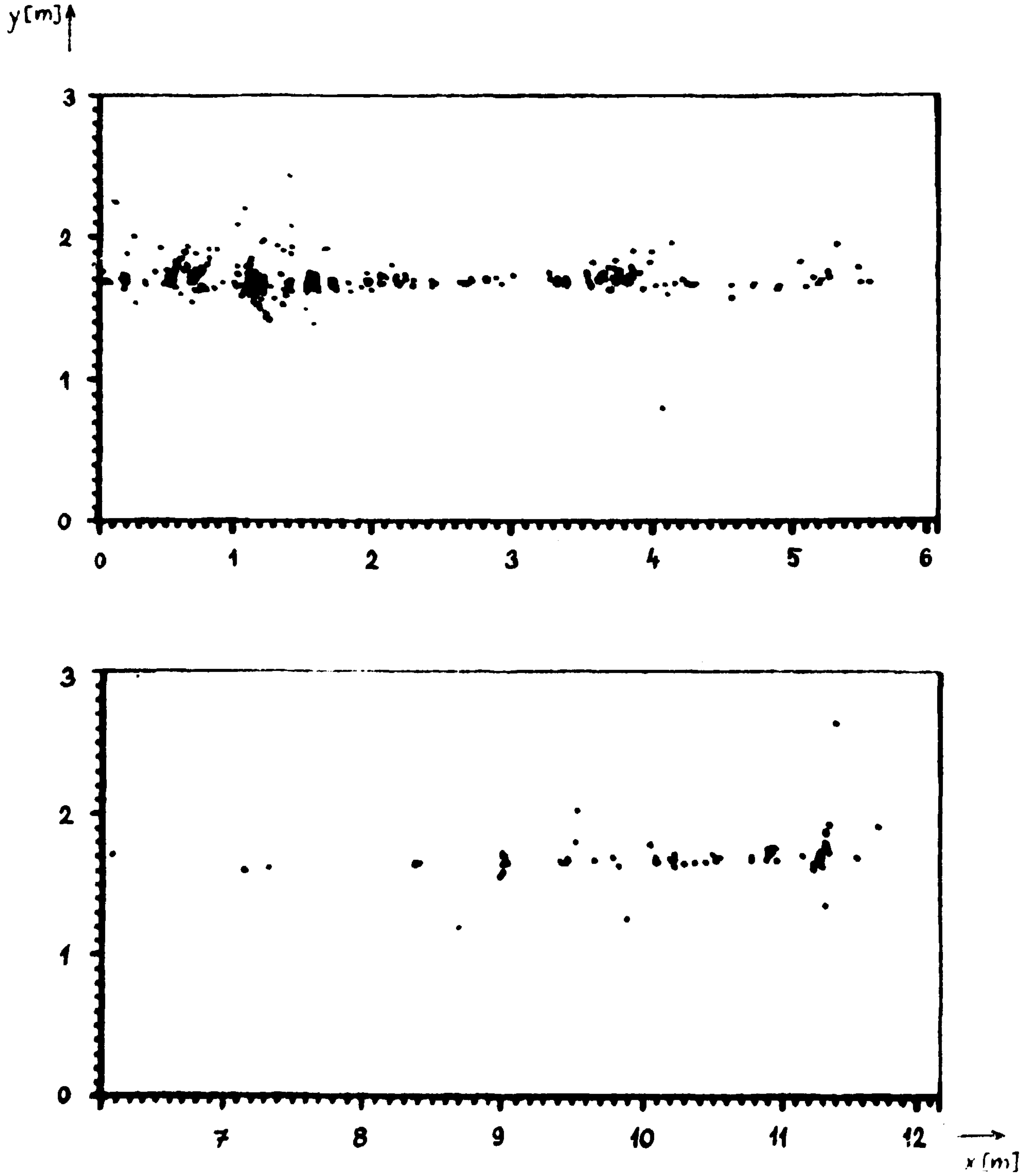

Fig. 3. Pictorial representation of emission events in the form of points on developed surface of the vessel

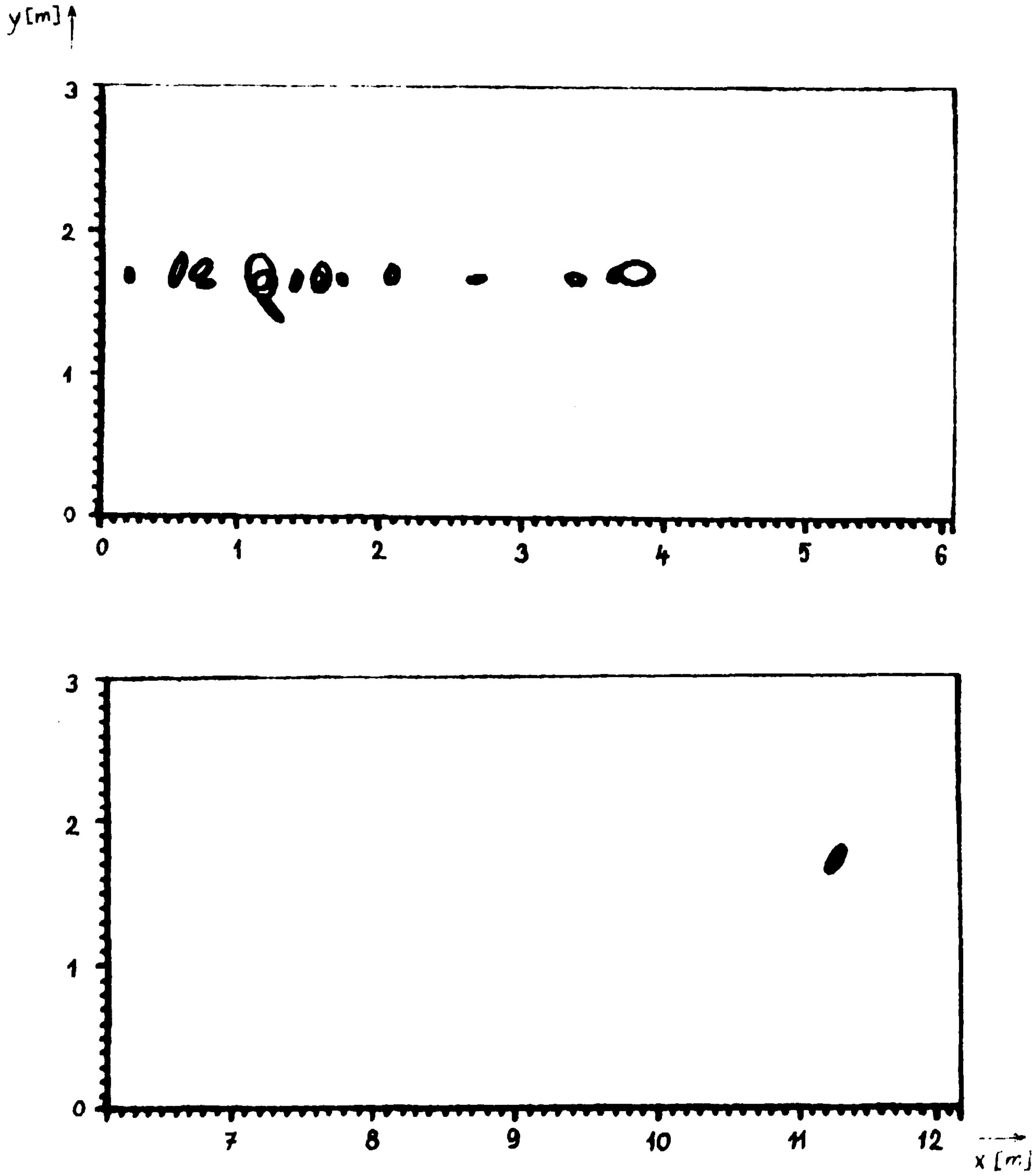

Fig. 4. The probability ellipses of found clusters

Table 1.

		series					
		arithmetic			geometric		
		ν			ν		
		5	10	15	5	10	15
M	5	0.737	0.733	0.735	0.697	0.680	0.679
	10	0.729	0.715	0.719	0.680	0.644	0.639
	15	0.724	0.714	0.714	0.672	0.640	0.628
M = 1		0.864	dominant cluster				
ν = 1, M arbitrary		0.864	clusters with equal density				

Table 2.

minimum spanning tree No	1	2	3	final number of points in cluster
parameters of Weibull distribution b c	 0.51 $240\cdot10^{-3}$	 0.73 $71\cdot10^{-3}$	 0.71 $72\cdot10^{-3}$	
the length of critical edge [mm]	17	38	45	
the number of points in clusters (asterisk denotes the number of points in that cluster which is in further processing disregarded)	12	24	26	26
		20	22	22
	22	23	23	23
	20	22	22	22
	8	10	10	10
		8	8	8
		9	9	9
	14 10	30	32	32
		9	9	9
		8	8	8
		10	12	12
			8	8
	34	44*		44
	261*	53*		261+53
	95*	22	22	95+22