

## STATISTICAL MODEL FOR LIDAR SIGNAL VALIDATION

V. Babin, M. Ciobanu, C. Talianu, D. N. Nicolae

National Institute of R&D for Optoelectronics INOE 2000, Bucharest-Magurele,  
P.O. Box MG 5, Romania

In this paper we build a statistical model for the LIDAR signal validation (estimation, selection, statistical validation of the Mie backscattered optical field from pollutants at medium and long range distances). The LIDAR backscattered optical signal is realized via the operative characteristic. Defining  $Q_o$  as the false alarm probability and  $Q_d$  the detection probability, the operative characteristic is a curve in the plane defined by the indecision zone of the LIDAR signal validation. The aim of this paper is to propose some optimization methods of the operative characteristic associated with the LIDAR signal validation.

(Received October 9, 2002; accepted October 31, 2002)

*Keywords:* LIDAR signal validation, Mie backscattered optical field, Pollutants

### 1. Theoretical considerations

We present a quantum detection model for the non-coherent optical radiation at the noise limit. Using the coherent states representation of the density matrix operator  $(\bar{\rho})$  and of the detection operators  $(\bar{I})$ , we estimate the detection probability ( $Q_d$ ) and the false alarm probability ( $Q_o$ ), the signal - to - noise ratio (S/Z), as function of the number of degrees of freedom of both temporal oscillation modes ( $M_t$ ) and spatial oscillation modes ( $M_s$ ). Using the Neyman-Pearson criterion [1] we build an algorithm for the validation of the statistical hypothesis in analyzing the non-coherent optical field.

We consider the optical field to be composed of a large number of oscillators, in thermal equilibrium with the substance [2]. In this case, the generation and annihilation operators,  $(\hat{C}^+)$  and  $(\hat{C}^-)$  will be  $\hat{C}_{pm}^+; \hat{C}_{pm}^-$ , where:

$p \Rightarrow$  the spatial oscillation mode index;  
 $m \Rightarrow$  the temporal oscillation mode index.

If a filter selects a narrow band and a polarizer selects the linear polarized radiation, the field operator reads [2]:

$$\hat{\phi}^-(\vec{r}, t) = \sqrt{\frac{1}{2} \hbar \Omega Z_0} \sum \hat{C}_{pm}^- \cdot \eta_p(\vec{r}) \cdot g_m(t) e^{-i\Omega t} \quad (1)$$

where:

$\Omega$  is the central frequency of the filter;

$Z_o = \sqrt{\frac{\mu}{\epsilon}}$  is the impedance of the medium;

$g_m(t), \eta_p(\vec{r})$  are the wave functions associated with the temporal and spatial modes.

If  $(\hat{\rho})$  is the density matrix operator characterizing the optical field, the quantum correlation function reads:

$$\Gamma(\vec{r}_1, \vec{r}_2; \vec{t}_1, \vec{t}_2) = \frac{2}{Z_0} \text{Tr} \left[ \hat{\phi}^-(\vec{r}_2, \vec{t}_2) \hat{\rho} \hat{\phi}^+(\vec{r}_1, \vec{t}_1) \right] \quad (2)$$

Integrating on the aperture surface (A) and on the illumination time ( $0, T'$ ) it results:

$$\hat{C}_{pm}^- = \frac{1}{\sqrt{\frac{1}{2} \hbar \Omega Z_0}} \int_0^{T'} \int_A g_m^*(t) e^{-i\Omega t} \cdot \hat{\phi}^-(\vec{r}, t) \cdot d^2\vec{r} \cdot dt \quad (3)$$

where:  $d^2\vec{r} = d\mathbf{x} \cdot d\mathbf{y}$ .

From the thermal equilibrium condition we find the optical field ( $T_r$  being the trace of the operator)

$$\text{Tr} \left[ \hat{\rho} \hat{\phi}^-(\vec{r}, t) \right] = 0 \quad (4)$$

it results:

$$\text{Tr} \left[ \hat{\rho} \hat{C}_{pm}^- \right] = 0 \quad (5)$$

Following Helstrom [3] and Glauber [4] it results:

$$\Gamma(\vec{r}_1, \vec{r}_2; t_1, t_2) = \Gamma_1(\vec{r}_1, \vec{r}_2; t_1, t_2) + \Gamma_0(\vec{r}_1, \vec{r}_2; t_1, t_2) \quad (6)$$

where:  $(\Gamma_1) \Rightarrow$  the correlation function of the detected optical field;

$(\Gamma_0) \Rightarrow$  the background radiation correlation function.

We define the correlation matrix elements, taking into account (6):

$$\Phi_{pm, qn} = \text{Tr} \left[ \hat{C}_{pm}^- \hat{\rho} \hat{C}_{qn}^+ \right] = \Phi_{pm, qn}^{(1)} + \Phi_{pm, qn}^{(0)} \quad (7)$$

In the case of monochromatic radiation detection, it results:

$$\Gamma(\vec{r}_1, \vec{r}_2; t_1, t_2) = \varphi_1(\vec{r}_1, \vec{r}_2) \chi(t_1 - t_2) e^{i\Omega(t_1 - t_2)} \quad (8)$$

where  $\chi(t_1 - t_2)$  is the eigen function (integral Kernel for the evolution equation).

The correlation matrix associated to the signal will be  $(\hat{\rho}_1)$ :

$$\Phi_{pm, qn}^{(1)} = \left\{ \frac{1}{\hbar \Omega} \int_0^T \int_0^T g_m^*(t_2) \chi(t_1 - t_2) g_n(t_1) dt_1 \cdot dt_2 \right\} \cdot \left\{ \int_A \int_A \eta_p^*(\vec{r}_2) \varphi_1(\vec{r}_1, \vec{r}_2) n_q(\vec{r}_1) d^2\vec{r}_1 \cdot d^2\vec{r}_2 \right\} \quad (9)$$

The background radiation correlation matrix will have the form  $(\hat{\rho}_0)$ :

$$\Phi_{pm, qn}^{(0)} = N \cdot \delta_{pq} \cdot \delta_{mn} \quad (10)$$

Because the thermal radiation is accomplished at low levels, we search statistical strategies based on the Neyman-Pearson criterion order to identify two statistical hypotheses:  $H_0$ ,  $H_1$ , defined as follows [5]:

$$H_0 : \Phi_{pm, qn} = \Phi_{pm, qn}^{(0)} \quad (11)$$

(only the background is detected)

$$H_1 : \Phi_{pm,qn} = \Phi_{pm,qn}^{(0)} + \Phi_{pm,qn}^{(1)} \quad (12)$$

(both the background and the emitted radiation are detected).

We associate to the two hypotheses  $H_0$ ,  $H_1$ , the density matrix operators ( $\hat{\rho}_0, \hat{\rho}_1$ ) and the detection operator ( $\hat{\Pi}$ ). We define the detection probability [6]:

$$Q_d \stackrel{\text{Def.}}{=} P_r [H_1 | H_1] = \text{Tr} [\hat{\rho}_1 \hat{\Pi}] \quad (13)$$

and the false alarm probability [6]:

$$Q_0 \stackrel{\text{Def.}}{=} P_r [H_1 | H_0] = \text{Tr} [\hat{\rho}_0 \hat{\Pi}] \quad (14)$$

From (10) we see the average number of emitted photons by the background radiation is  $\langle N \rangle$ . The average number of photons emitted by the source is [6]:

$$N_s = \frac{1}{\hbar \Omega} \int_A \int_0^T I(\vec{r}, t) \cdot d^2\vec{r} \cdot dt \quad (15)$$

where:  $I(\vec{r}, t) = \Gamma(\vec{r}, \vec{r}; t, t)$

$$A_0 = \frac{B_\tau^2}{\int_{(0)} [B(\vec{U})] d^2\vec{U}} \rightarrow \text{effective area of the optical field source.}$$

## 2. Contributions

### 2.1. The detection of a spatial coherent signal in the presence of the background radiation

Conformal to the principles described in chapter 1, the empirical average of the measurements number ( $n$ ) is:

$$\bar{V}_n = \frac{1}{n} \sum_{K=1}^n V_K \quad (16)$$

The statistical hypothesis in the limit of which the test is performed has the form:

$$\hat{H}_0 : V_{K_0} = V_0 \quad (\text{lower limit of the signal}) \quad (17)$$

$$\hat{H}_1 : V_{K_I} = V_I = \frac{1}{N_{max}} \sum_{n=1}^{N_{max}} V_n \quad (\text{upper limit of the signal}) \quad (18)$$

$$\bar{V} = \frac{1}{N_{max}} \sum_{n=1}^{N_{max}} \bar{V}_n \quad (\text{average limit of the signal}) \quad (19)$$

$$\lambda_K = \sigma^2 = \frac{1}{N_{max}} \sum_{n=1}^{N_{max}} (\bar{V}_n^2 - \bar{V}^2) \quad (\text{data dispersion}) \quad (20)$$

In the following, we choose the repartition functions- in our case, the Gauss functions [7]:

$$\begin{aligned}
 P_0 &= P_0(V, \frac{\sigma}{\sqrt{N_{\max}}}, V_0) = \frac{1}{\sqrt{\pi} \cdot \sigma / \sqrt{N_{\max}}} \cdot \exp\left[-\frac{(V - V_0)^2}{(\sigma / \sqrt{N_{\max}})^2}\right] \\
 P_1 &= P_1(V, \frac{\sigma}{\sqrt{N_{\max}}}, V_1) = \frac{1}{\sqrt{\pi} \cdot \sigma / \sqrt{N_{\max}}} \cdot \exp\left[-\frac{(V - V_1)^2}{(\sigma / \sqrt{N_{\max}})^2}\right]
 \end{aligned}
 \tag{21}$$

where:  $N_{\max}$  represents the maximum number of data submitted to statistical validation;

The next step consists in computing the probability to detect a false signal. This probability is given by:

$$P_r \left\{ \hat{H}_1 \mid \hat{H}_0 \right\} = \int_{V_C}^{+\infty} P_0(V, V_0, \frac{\sigma}{\sqrt{N_{\max}}}) dV = \alpha
 \tag{22}$$

In the case of LIDAR detection, where the signal is low when compared with the noise, we choose:  $\alpha \approx 0.1$ . From (21) we determine  $V_C$ , which indicates the best critical region.

The power function of the test is the probability to detect a signal in the presence of the target. This probability has the form:

$$P_r \left\{ \hat{H}_1 \mid \hat{H}_1 \right\} = \int_{V_C}^{+\infty} P_1(V, \bar{V}, \frac{\sigma}{\sqrt{N_{\max}}}) dV = 1 - \beta
 \tag{23}$$

where  $V_C$  is the best critical zone that satisfies:

$$W^* : V_K > V_{KC}
 \tag{24}$$

With:

$$u = \frac{V - V_0}{\sigma / \sqrt{N_{\max}}}
 \tag{25}$$

From (21) we can explicitate  $V_C$ :

$$V_C = V_0 + \frac{\sigma}{\sqrt{N_{\max}}} u_{1-\alpha}
 \tag{26}$$

where  $u$  represents the inverse error function:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt
 \tag{27}$$

We next compute the signal - to - noise ratio  $S/Z$ :

$$(S / Z) = u_{1-\alpha} + u_{1-\beta}
 \tag{28}$$

In order the algorithm to be operative, we choose the lower limit ( $V_0$ ) and the upper limit ( $V_1$ ) as:

$$V_1 - V_0 > \frac{\sigma}{\sqrt{N_{max}}} \cdot (S/Z) \quad (29)$$

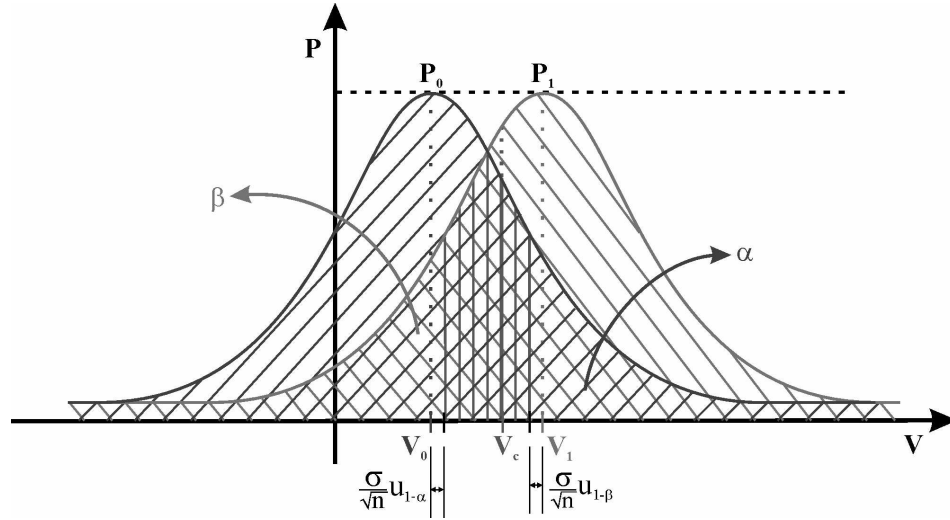


Fig. 1. Integration limits that defines  $\alpha$  and  $\beta$  probabilities.

In the case of LIDAR-type data, the dispersion is large, which implies large limits for the validation interval.

In the following, we explicate the statistical signification of  $\alpha$  and  $\beta$ :

$$\alpha = P_r \left\{ \overline{V}_n > V_C \mid \hat{H}_0 \right\} = Q_0 \quad (30)$$

so  $\alpha$  is the probability of validating a target when it does not exist, and  $\alpha$  must be much lessar then unity. Therefore,  $\alpha$  is called the false alarm probability:  $\alpha \ll 1$ .

$\beta$  is defined as follows:

$$\beta = P_r \left\{ \overline{V}_n < V_C \mid \hat{H}_1 \right\} \quad (31)$$

so  $\beta$  is the probability to invalidate a target when it exists, and, as well, and  $\beta$  must be much lessar than unity:  $\beta \ll 1$ . With the aid of this, we define the detection probability ( $Q_d$ ) as:

$$(Q_d) = 1 - \beta \quad (32)$$

therefore  $Q_d \leq 1$

The ideal case is when ( $Q_d$ ) equals unity, but typical values are ( $Q_d$ ) = 0.9.

## 2.2. Software development for experimental data

The code for the data is intended to optimize the number of measurements that, in the limits of a risk, validate or not the existence of a target - in our case a pollutant from the lower atmosphere, - and to return, before validation, only those measurements that, with some probability, contain real information.

The LIDAR data are not homogeneous, cover few orders of magnitude (from  $10^2$  for signals received from tens of meters to  $10^{-5}$  for signals received from kilometers), and decrease cvasi-exponentially with distance. Therefore, the algorithm cannot be applied to the whole series of data, but only to some 'regions' of interest, which present relatively low dispersion and which are suspected to represent a target to be validated (hypothesis  $H_1$ ). For that, we choose the limits  $V_0$  and  $V_1$  as to satisfy (29) for a mean dispersion of 0.01 and an average number of input data of 50 (representing data from 90 m distance consecutive).

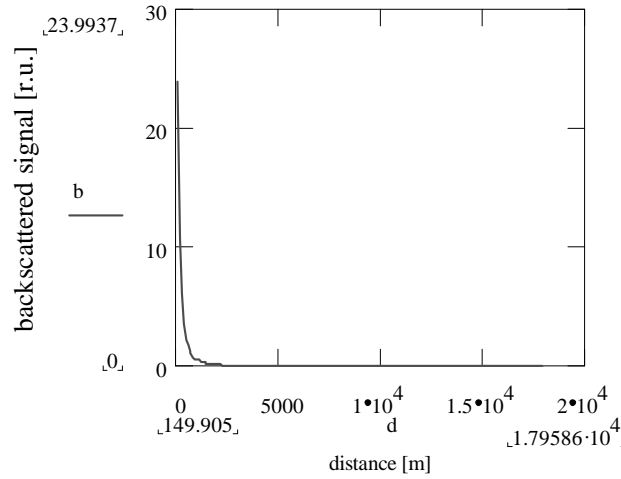


Fig. 2. Backscattered signal (r.u.) function of distance – LIDAR data.

In the following, we explicate the algorithm, the organigram of which is presented in Fig. 3. Each experimental result  $V_n$ , is submitted to statistical validation using two loops, having the following signification:

$$1. \text{ if: } \quad \bar{V}_n > V_0 + \left( \frac{\sigma}{\sqrt{n}} \right) \cdot u_{1-\alpha} \quad (33)$$

$$\text{and: } \quad \bar{V}_n < V_1 - \left( \frac{\sigma}{\sqrt{n}} \right) \cdot u_{1-\beta} \quad (34)$$

then the  $\hat{H}_1$  hypothesis is validated, meaning we have validated a target when it exists.

$$2. \text{ if: } \quad \bar{V}_n \leq V_0 + \left( \frac{\sigma}{\sqrt{n}} \right) \cdot u_{1-\alpha} \quad (35)$$

$$\text{or: } \quad \bar{V}_n \geq V_1 - \left( \frac{\sigma}{\sqrt{n}} \right) \cdot u_{1-\beta} \quad (36)$$

then the  $\hat{H}_0$  hypothesis is validated, namely we 'see' the target when it does not exist.

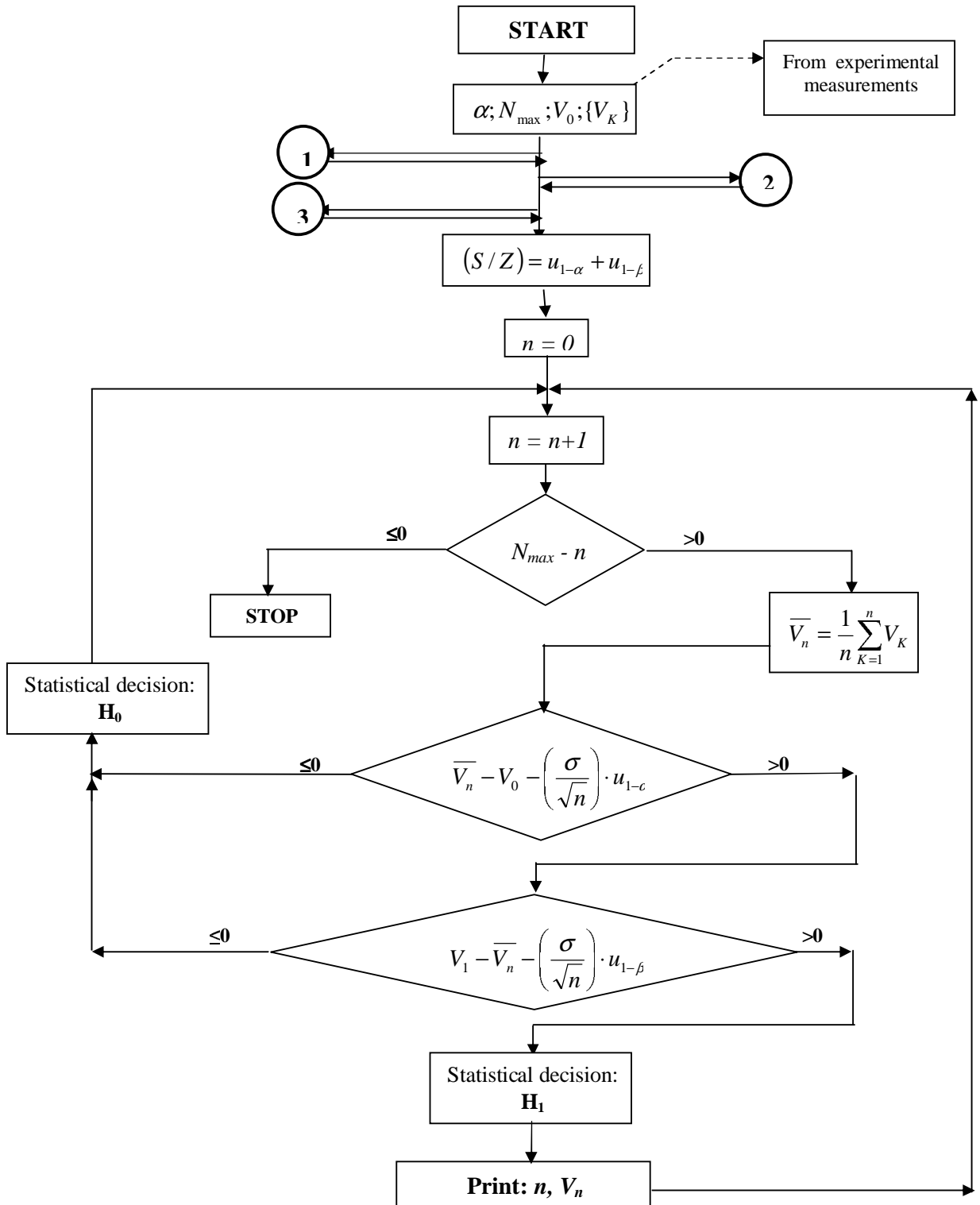


Fig. 3. The data validation organigram.  
 where: 1 – is the subroutine for computing  $\sigma$  using eq. (19)  
 2 – is the subroutine for computing  $\beta$  using eq. (21) and (22)  
 3 – is the subroutine for computing the inverse functions  $u_{1-\alpha}$  and  $u_{1-\beta}$ .

### 3. Results

Four distinct set of LIDAR data, were submitted to statistical validation, by extracting partial subsets. The results presented below refers to three subsets extracted from a single set and represents three interest zones: the near zone, characterized by a large signal to noise ratio and a large dispersion of data (Fig. 4), the far zone having a small but supra-unitary signal to noise ratio (Fig. 5) and the incertitude zone, having a signal to noise ratio less than unity (Fig. 6).

Distance	Backsc. sign.	Distance	Backsc. sign.	Distance	Backsc. sign.
4,287.270	1.71289E-02	6,086.120	6.10451E-03	7,884.980	2.98123E-03
4,377.210	1.64665E-02	6,176.060	6.29627E-03	7,974.920	2.89236E-03
4,467.150	1.54131E-02	6,266.010	6.04010E-03	8,064.860	2.93502E-03
4,557.100	1.43143E-02	6,355.950	5.76585E-03	8,154.810	2.59470E-03
4,647.040	1.36350E-02	6,445.890	5.41251E-03	8,244.750	2.52049E-03
4,736.980	1.31374E-02	6,535.840	5.33710E-03	8,334.690	2.66854E-03
4,826.920	1.28244E-02	6,625.780	4.81992E-03	8,424.630	2.34885E-03
4,916.870	1.17361E-02	6,715.720	4.54518E-03	8,514.580	2.21332E-03
5,006.810	1.13499E-02	6,805.660	4.82870E-03	8,604.520	2.27948E-03
5,096.750	1.10528E-02	6,895.610	4.60634E-03	8,694.460	2.13801E-03
5,186.700	1.03845E-02	6,985.550	4.39572E-03	8,784.400	2.07245E-03
5,276.640	9.66683E-03	7,075.490	4.07690E-03	8,874.350	2.00675E-03
5,366.580	9.16400E-03	7,165.440	3.88962E-03	8,964.290	1.95560E-03
5,456.520	9.06568E-03	7,255.380	3.66425E-03	9,054.230	1.90143E-03
5,546.470	8.42323E-03	7,345.320	3.61423E-03	9,144.180	1.83255E-03
5,636.410	7.78226E-03	7,435.260	3.44130E-03	9,234.120	1.81620E-03
5,726.350	7.81105E-03	7,525.210	3.25654E-03	9,324.060	2.02492E-03
5,816.290	7.24820E-03	7,615.150	3.35238E-03	9,414.000	1.92656E-03
5,906.240	7.15578E-03	7,705.090	3.01539E-03	9,503.950	1.98310E-03
5,996.180	6.80406E-03	7,795.030	2.98813E-03	9,593.890	2.25011E-03

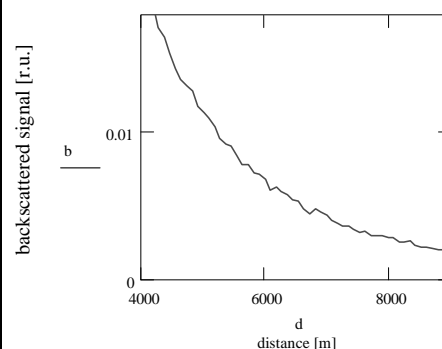


Fig. 4. Results of validation program for the first subset data ( $H_0$  decision – white zone,  $H_1$  decision – grey zone).

Distance	Backsc. sign.	Distance	Backsc. sign.
8,874.350	2.00675E-03	10,493.300	4.21244E-03
8,964.290	1.95560E-03	10,583.300	7.95000E-03
9,054.230	1.90143E-03	10,673.200	1.71054E-02
9,144.180	1.83255E-03	10,763.100	2.85091E-02
9,234.120	1.81620E-03	10,853.100	3.08959E-02
9,324.060	2.02492E-03	10,943.000	2.97272E-02
9,414.000	1.92656E-03	11,033.000	2.93977E-02
9,503.950	1.98310E-03	11,122.900	2.77904E-02
9,593.890	2.25011E-03	11,212.900	2.57241E-02
9,683.830	3.06676E-03	11,302.800	2.64090E-02
9,773.770	9.92985E-03	11,392.700	2.76933E-02
9,863.720	1.57287E-02	11,482.700	2.71736E-02
9,953.660	1.94575E-02	11,572.600	2.35021E-02
10,043.600	1.89263E-02	11,662.600	1.87486E-02
10,133.500	1.32579E-02	11,752.500	1.29718E-02
10,223.500	7.91412E-03	11,842.500	8.57789E-03
10,313.400	5.55261E-03	11,932.400	4.96461E-03
10,403.400	3.57997E-03	12,022.300	3.03832E-03

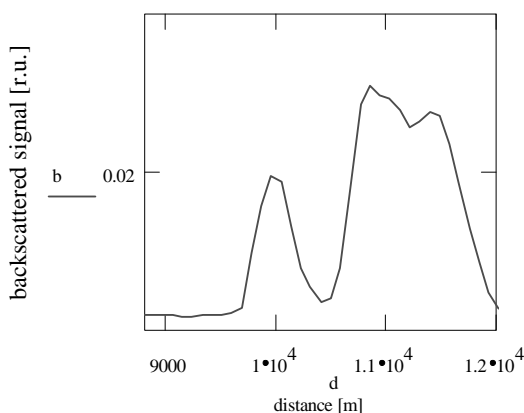


Fig. 5. Results of validation program for the second subset data ( $H_0$  decision – white zones,  $H_1$  decision – grey zones).

Distance	Backsc. sign.	Distance	Backsc. sign.	Distance	Backsc. sign.
12,382.100	2.62432E-04	14,270.900	3.91212E-05	16,159.700	3.76465E-05
12,472.100	9.32470E-05	14,360.900	1.31563E-05	16,249.600	2.63138E-05
12,562.000	7.90184E-05	14,450.800	7.67621E-06	16,339.600	2.66491E-05
12,651.900	7.64818E-05	14,540.700	5.14443E-06	16,429.500	8.25131E-05
12,741.900	1.83941E-05	14,630.700	0.00000E+00	16,519.500	3.60204E-05
12,831.800	4.80250E-05	14,720.600	6.00963E-06	16,609.400	0.00000E+00
12,921.800	3.08797E-05	14,810.600	1.81827E-05	16,699.400	1.02249E-05
13,011.700	5.17500E-05	14,900.500	1.86897E-05	16,789.300	1.92280E-05
13,101.700	5.21585E-05	14,990.500	3.96854E-05	16,879.200	0.00000E+00
13,191.600	2.04085E-05	15,080.400	4.90035E-05	16,969.200	0.00000E+00
13,281.500	4.12841E-05	15,170.300	7.88042E-05	17,059.100	3.44315E-05
13,371.500	1.53675E-05	15,260.300	1.50541E-05	17,149.100	3.75209E-05
13,461.400	8.88407E-05	15,350.200	5.07148E-05	17,239.000	0.00000E+00
13,551.400	0.00000E+00	15,440.200	5.12895E-05	17,329.000	5.53651E-05
13,641.300	6.02922E-05	15,530.100	7.81705E-05	17,418.900	2.91879E-05
13,731.300	3.13705E-05	15,620.000	0.00000E+00	17,508.800	4.39351E-05
13,821.200	2.58303E-05	15,710.000	4.70883E-05	17,598.800	1.48188E-05
13,911.100	6.99938E-05	15,799.900	9.58410E-06	17,688.700	0.00000E+00
14,001.100	2.64465E-05	15,889.900	6.56169E-05	17,778.700	1.50585E-05
14,091.000	5.60019E-05	15,979.800	6.89908E-05	17,868.600	0.00000E+00
14,181.000	5.34033E-05	16,069.800	2.26338E-05	17,958.600	4.16652E-05

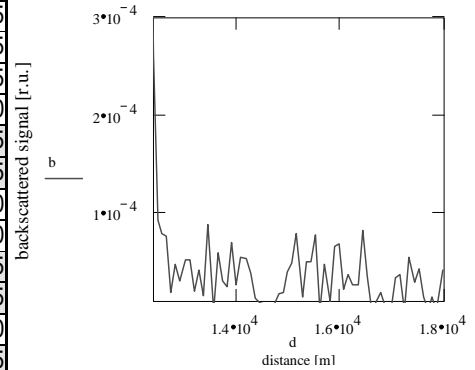


Fig. 6. Results of validation program for the last subset data ( $H_0$  decision – white zones,  $H_1$  decision – grey zones).

#### 4. Conclusions

Analyzing the results obtained from experimental data from different sites in Bucharest, we can draw the following conclusions:

1. The representation of data at a large scale do not allow the direct identification of the targets, which imposes the use of the statistical validation method as a supplementary filter in order to extract the useful information.
2. The method returns only the data corresponding to a possible target, as we can see from Fig. 4 (which shows the lack of the target) and from Fig. 5 (which evidences a double target).
3. We observe a good agreement between the processed data and the measured data, in the range where these observations are possible (sufficiently large signal-to-noise ratio)-Figs. 4,5
4. Validation of measurements was realized even when the signal was hidden by the noise.
5. In the conditions of system calibration, the validation algorithm allows for pollutant density measurements at large and very large distances.
6. For the detection of low amplitude signals a large number of measurements are necessary (Fig. 6), whereas for large amplitude signals a lower number of measurements is necessary (Fig. 5).
7. The sensitivity of the method depends, in the case of these data with large dispersion, on an appropriate choice of the subsets, which will be the object of further studies.

#### References

[1] Neyman J. Pearson, On the problem of the most efficient tests of statistical hypotheses, Joint statistical Papers of J. Neyman, E. S. Pearson, Cambridge University Press, 1967.  
 [2] M. Born, E. Wolf, Principles of Optics. Electromagnetic theory of propagation, interference and diffraction of light, Pergamon Press, New York, 1973.  
 [3] C. W. Helstrom, Quantum Detection and Estimation Theory, Academic Press, New York, Math in Science and Engineering, 1976.

- [4] R. J. Glauber, The quantum theory of optical coherence, *Phys. Rev.* **130**, June 15, 2529 (1963).
- [5] S. Kerler, S. Haykin, Conventional MLM and MEN Mapping of Inhomogeneous Spatial Field, First IEEE Workshop on Spectral Estimation, Hamilton, Ont., pp 7.3.1-7, 1981.
- [6] V. Babin, A model to verify the validity of statistical hypotheses in non-coherent (thermal) radiation detection, *Proceedings SPIE*, nr. 54, pag. 11-13, Los Angeles, SUA, 1993.
- [7] V. I. Vlad, V. Babin, A. Mocofanescu, Analytical treatment of the three-dimensional model of stimulated Brillouin scattering with axial symmetric pump wave, *J. Optoelectron. Adv. Mater.* **4**(3), 581 (2002).