A Second Order Autoregressive Based Technique for Pipeline Leak Detection

MARLLENE DANETI

Faculty of Electronics and Communication, Applied Electronics Department
"Politehnica" University of Timisoara
2 Bd. Vasile Parvan, 300233, Timisoara
ROMANIA
marllene.daneti@etc.upt.ro   http://www.etc.upt.ro

Abstract: - Efficient leak detection techniques need to be characterized both by rapidity and robustness. This paper studies a simple detection method based on the second order autoregressive (AR) parameters of the pipeline signals- a trade-off between the two required characteristics. The theoretical geometrical positions defined by the AR coefficients are developed for some particular cases of interest. The resulted model proves to be in concordance with the experimental data. The algorithm is next tested under non-stationary burst-type conditions induced both soft and hard, indicating a good stability in comparison with the basic most inexpensive computational detection method. The area under the Receiver Operating Characteristics (ROC) curves indicates also a better performance of the proposed detector.

Key-Words: - Leak detection, pipeline monitoring, AR models, ROC curves.

1   Introduction

Fluid transportation through pipeline systems has been since ancient times a preoccupation of great practical importance. However, the major problem concerning these systems is that a significant amount of the transported fluid is lost if leakage occurs in the distribution pipes. On the other hand, besides the fact that a great quantity of material is wasted, another problem that rises is the environment erosion and pollution. Therefore, a great attention must be paid to a careful pipeline monitoring, in order to be able to diagnose the pipeline's state at every time moment, if possible. Consequently, two major problems must be solved here: firstly, the detection as soon as possible of a leak occurrence and secondly, in case of a leak has been detected, the accurate position of the leak on the pipe. This paper addresses to the first problem of leak detection.

In practice, many methods for solving this problem have been developed. Among them we recognize: methods that use the measurements of pressure and flow investigating the material balance between input and output; detection by patrolling and inspecting along the pipeline; remote acoustic methods which detect leaks through processing the information contained in the random signals captured from the sensing devices placed on the pipeline [2],[3]. Other methods involve different techniques such as tracer gas, infrared imaging or ground penetrating radar but their use is limited.

Until now, no universal effective method for detecting and locating leaks was found. Lately, combined techniques of flow measurement and acoustic methods are used in practice. An important step in improving the effectiveness of the leak detection systems would be to find algorithms that speed up the detection process and permit a remote detection from the pipeline inlet.

A quick detection method frequently used is based on monitoring the increase of the acquired signals' power from the background noise [3]. However, in practice, the pipeline signals prove to have a non-stationary behavior with respect to both their mean and variance, due to various internal and external conditions such as road traffic, sudden pressure and flow velocity variations (turbulences), etc. The non-stationary behavior can interfere with the de decision criterion for leak detection.

Another approach is based on performing the spectral analysis of the acquired signals and deciding if any major changes occurred in their spectral components. On this principle, for metal water pipelines, leak detection professionals can discern if there is a leak on the pipe or not, by using a simple listening device (e.g. ground microphone, etc.) [2]. Fig.1 shows two typical power spectra corresponding respectively to both situations when a leak isn't and is present, at the same locations on an experimental pipeline installation. It can be observed that in the presence of a leak the corresponding signal has higher spectral components.
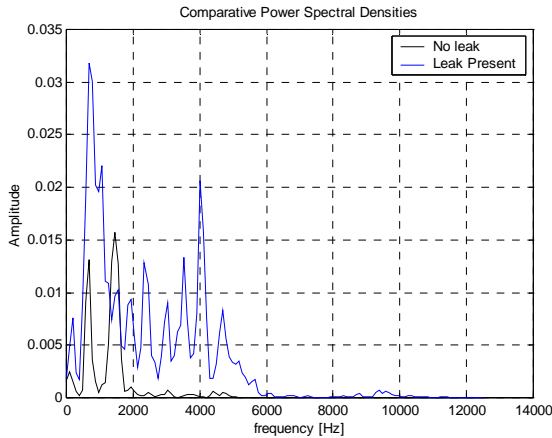
Fig.1 Typical power spectral densities for pipeline signals

Observing also that both signals can be modeled as a couple of low-pass harmonics plus noise, a simple method for detection is developed based on modeling the received signals using only the second order AR parameters, a_1 and a_2 [5],[8]. The distance measure between the obtained points in the plane defined by the a_1, and a_2 coordinates is used as a test statistic for detection. This is a robust technique, especially for non-stationary situations. Also, this method is simple, quick, need only one non-intrusive sensing device and is passive (i.e. no external test signals are used).

Fig.2 illustrates the algorithm's principle. Briefly, the received signals are divided in short segments, which can be viewed as piecewise stationary; for each segment the second order AR coefficients are computed [8]; the "gravity" center of each set of data is estimated, using their median values; finally, the distance measure between the data "gravity" centers is evaluated: if the value of this random variable exceeds a certain threshold, then the algorithm decides that there is a leak in the pipeline system; otherwise, not.
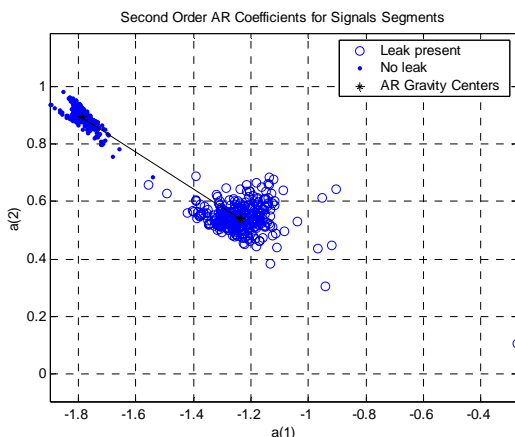

Fig.2 The estimated AR parameter sets for two typical pipeline signals.

2 Problem Formulation

As previously stated, the major question here is included in the binary detection problem [4],[5], where the receiver processes the acquired random signals in order to decide if any leak is present according to the general detection model described by [4]:

$$r(t) = s_0(t) + n(t), \quad t \in [T_i, T_f] \ : H_0,$$
$$r(t) = s_1(t) + n(t), \quad t \in [T_i, T_f] \ : H_1, \tag{1}$$

where $r(t)$ is a sample function of the received random signal; $s_0(t)$ denotes the signal produced by the main stream flowing through the pipeline; $s_1(t)$ denotes the signal produced by the leak and the main stream; $n(t)$ is the background disturbing noise; t denotes the time variable which takes values between the initial and final moment, T_i and T_f, respectively; finally H_0 and H_1 denote the null and the alternative hypothesis, respectively. The null hypothesis assumes that the source leak signal is not present, while the alternative hypothesis assumes the opposite. In this problem, the received signals may be characterized mainly by a non-stationary behavior.

In attempting to solve this problem, a case study is developed in order to find a possible connection between the received signal's spectral components and their second order AR parameters, a_1 and a_2. Therefore, the following ideal cases for the received signals are presented in Table 1, while the corresponding autocorrelations functions are described by equations (2) to (6).

Table 1

No.	Considered cases	Received signal	Auto-correlation function
1.	one sinusoid	$A \sin 2\pi f t$	Eq.(2)
2.	one sinusoid in white noise	$A \sin 2\pi f t + w(t)$	Eq.(3)
3.	two sinusoids	$A_1 \sin 2\pi f_1 t + A_2 \sin 2\pi f_2 t$	Eq.(4)
4.	two sinusoids in white noise	$A_1 \sin 2\pi f_1 t + A_2 \sin 2\pi f_2 t + w(t)$	Eq.(5)
5.	two sinusoids in colored noise	$A_1 \sin 2\pi f_1 t + A_2 \sin 2\pi f_2 t + w(t)$	Eq.(6)

$$R(\tau) = \frac{A^2}{2}\cos 2\pi f \tau \qquad (2)$$

$$R(\tau) = \frac{A^2}{2}\cos 2\pi f \tau + \sigma_w^2 \delta(\tau) \qquad (3)$$

$$R(\tau) = \frac{A_1^2}{2}\cos 2\pi f_1 \tau + \frac{A_2^2}{2}\cos 2\pi f_2 \tau \qquad (4)$$

$$R(\tau) = \frac{A_1^2}{2}\cos 2\pi f_1 \tau + \frac{A_2^2}{2}\cos 2\pi f_2 \tau + \sigma_w^2 \delta(\tau) \qquad (5)$$

$$R(\tau) = \frac{A_1^2}{2}\cos 2\pi f_1 \tau + \frac{A_2^2}{2}\cos 2\pi f_2 \tau + \sigma_w^2 e^{-\alpha\tau} \qquad (6)$$

In the above relationships A denote amplitudes, f denote frequencies, τ stands for the time lag argument; δ is the unit impulse; σ_w^2 is the white noise's power. Finally, α is a constant factor defining the spectrum of the colored noise according to:

$$S_w(f) = \frac{4\alpha\sigma_w^2}{\alpha^2 + 4\pi^2 f^2}. \qquad (7)$$

Other answers that need to be found regard the degree by which the experimental data match the theoretical results; finding a variable that can be used as a test statistic and it's probability density functions for both situations; selecting a hypothesis test procedure; computing the ROC curves for describing the detector's performance.

3 Problem Solution

The second order AR parameters a_1 and a_2 are obtained by solving the normal equations [5], [8] described, for this particular case, as:

$$\begin{cases} R[1] + a_1 R[0] + a_2 R[-1] = 0; \\ R[2] + a_1 R[1] + a_2 R[0] = 0. \end{cases} \qquad (8)$$

Here, R denotes the received signal's autocorrelation function computed at the time lags corresponding to the number of samples indicated in the square brackets. Hence, for each case described in Table 1 the autocorrelation samples $R[0]$, $R[1]=R[-1]$ and $R[2]$ are found from equations (2)-(6), prior to calculating the AR coefficients according to:

$$\begin{cases} a_1 = \dfrac{R[1] \cdot (R[2] - R[0])}{R^2[0] - R^2[1]}, \\ a_2 = \dfrac{R^2[1] - R[0]R[2]}{R^2[0] - R^2[1]}. \end{cases} \qquad (9)$$

3.1 The theoretical geometrical positions

For each case presented in Table 1, in the plane specified by the a_1 and a_2 coordinates, the theoretical points obtained define specific geometrical positions with varying frequencies and amplitudes. Next, these geometrical positions are derived with the purpose of comparing them with the experimental data.

3.1.1 One sinusoid

In this case, the relationships (2) and (9) produce the following simple solution:

$$a_1 = -2\cos\frac{2\pi f}{F_s}; \qquad a2 = 1, \qquad (10)$$

where Fs is the sampling frequency used in the signal acquisition. The geometrical positions of these points vary only with the signal's frequency as in Fig.3.
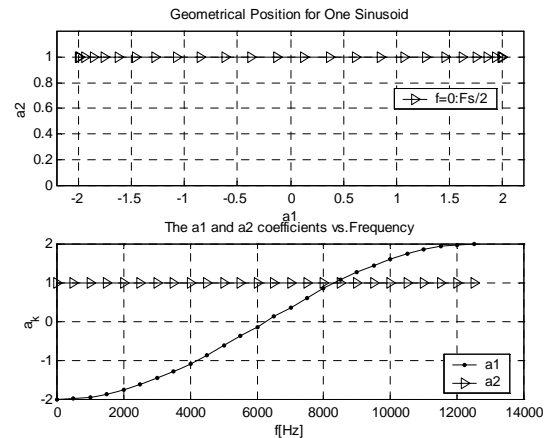

Fig.3 The theoretical AR coefficients for case 1

3.1.2 One sinusoid in white noise

Following the steps mentioned above, the AR coefficients become:

$$a_1 = \frac{2\cos^3 y - (2 + \eta^2)\cos y}{(1+\eta^2)^2 - \cos^2 y};$$

$$a2 = \frac{1 + \eta^2 - (1 + 2\eta^2)\cos^2 y}{(1+\eta^2)^2 - \cos^2 y}; \qquad (11)$$

$$y = \frac{2\pi f}{F_s}; \qquad \eta^2 = 2\frac{\sigma_w^2}{A^2}.$$

By eliminating the cos y factor between a1 and a2 yields:

$$a_1^2 = \frac{(ma_2 - 1)(a_2 + 1)^2}{a_2 - 2m + 1} \cdot \frac{1}{m};$$

$$m = 1 + \eta^2, \qquad m \in (1, +\infty). \tag{12}$$

The AR coefficients for different m values and frequencies varying from 0 to $Fs/2$ are depicted in Fig.4 and Fig.5.
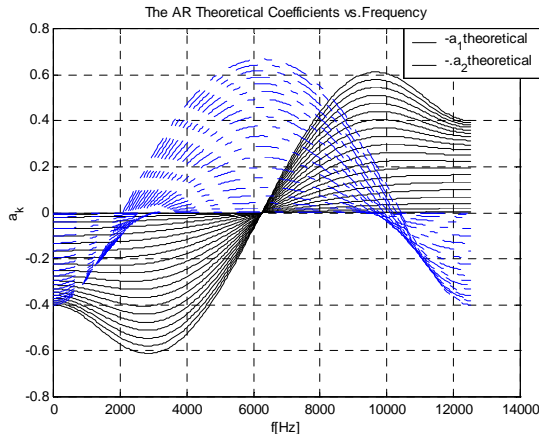

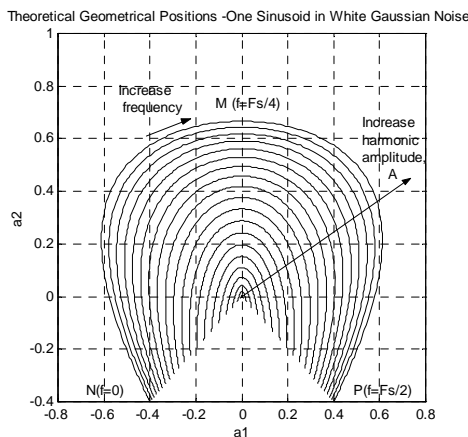Fig.4 Variation of the a_1 and a_2 parameters with frequency for different harmonic-to-noise ratios.

Fig.5 Corresponding geometrical positions for the second case.

In can be seen that the specified M, N and P points in Fig.5, corresponding respectively to zero, a quarter and a half of the sampling rate, have the following particular coordinates in the $(a_1 a_2)$ plane: $M\left(0, \frac{1}{m}\right)$; $N\left(-\frac{1}{m+1}, -\frac{1}{m+1}\right)$ and P, symmetric of N in rapport to $a_1=0$. When the m parameter approaches unity (i.e. the harmonic component is much stronger than the white noise component), this case tends be reduced to case one. Conversely, when m tends to infinity (i.e. the white noise component predominates), the geometrical position is concentrated towards the axes origin.

3.1.3 Two sinusoids

The second order AR coefficients in this case can be written as:

$$a_1 = \frac{(\rho^2 \cos y + \cos vy)(\rho^2 \cos 2y + \cos 2vy - \rho^2 - 1)}{(\rho^2 + 1)^2 - (\rho^2 \cos y + \cos vy)^2};$$

$$a_2 = \frac{(\rho^2 \cos y + \cos vy)^2 - (\rho^2 + 1)(\rho^2 \cos 2y + \cos 2vy)}{(\rho^2 + 1)^2 - (\rho^2 \cos y + \cos vy)^2} \tag{13}$$

$$\rho = \frac{A_1}{A_2}; \qquad v = \frac{f_2}{f_1}; \qquad y = 2\pi \frac{f_1}{F_s}$$
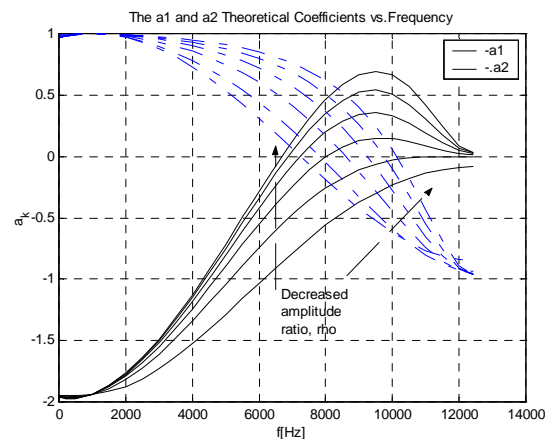
The corresponding curves are shown in Fig.6 to Fig.9.

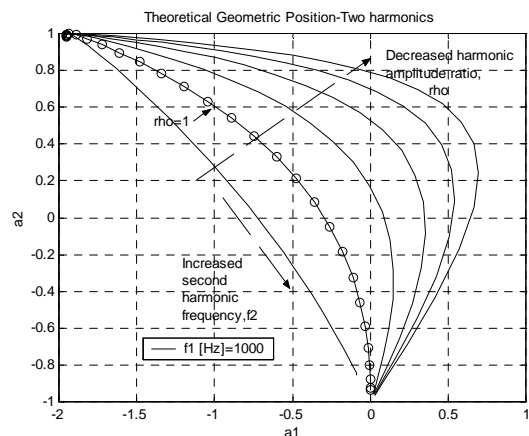Fig.6 Family curves for AR coefficients versus f_2 when f_1 is fixed ρ varies.

Fig.7 Geometrical positions for the third case (f_1-fixed, f_2, ρ varies)

Also it can be observed that this case reduces to the first case if one of the harmonics is significantly grater than the other one.

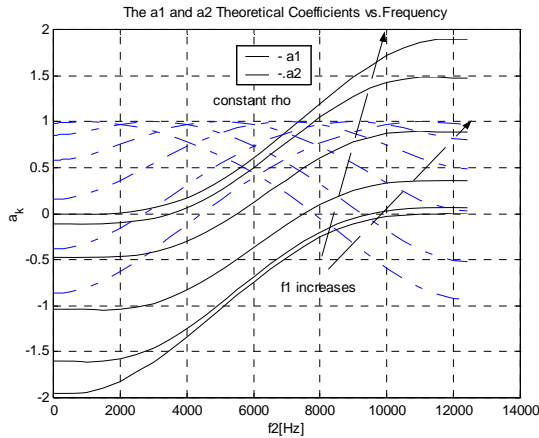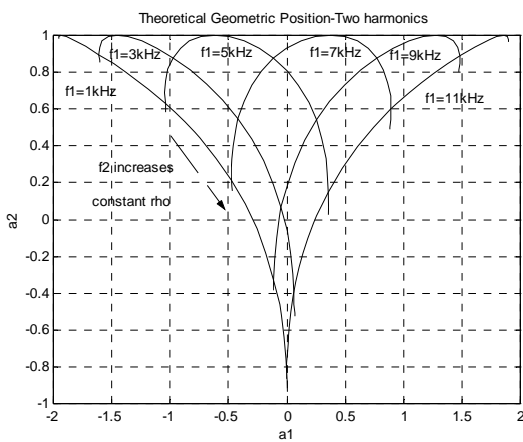Fig.8 Family curves for AR coefficients versus f_2 when ρ is fixed while f_1 varies.

Fig.9 Geometrical curves for the third case (ρ is fixed, f_1, f_2 varies)

3.1.4 Two sinusoids in white noise

In a similar fashion, in this case, the corresponding relations for the a_1 and a_2 parameters are described by the next equations while their variations are illustrated in Fig.10 to Fig.13.

$$a_1 = \frac{\left(\rho^2 \cos y + \cos \nu y\right)}{\left(\rho^2 + 1 + \eta^2\right)^2 - \left(\rho^2 \cos y + \cos \nu y\right)^2}$$
$$\cdot \left(\rho^2 \cos 2y + \cos 2\nu y - \rho^2 - 1 - \eta^2\right);$$

$$a_2 = \frac{\left(\rho^2 \cos y + \cos \nu y\right)^2}{\left(\rho^2 + 1 + \eta^2\right)^2 - \left(\rho^2 \cos y + \cos \nu y\right)^2} - \quad (14)$$
$$- \frac{\left(\rho^2 + 1 + \eta^2\right)\left(\rho^2 \cos 2y + \cos 2\nu y\right)}{\left(\rho^2 + 1 + \eta^2\right)^2 - \left(\rho^2 \cos y + \cos \nu y\right)^2};$$

where the variables y, ρ and ν have the same definitions as in the previous section, and η^2 is defined as:

$$\eta^2 = 2\frac{\sigma_w^2}{A_2^2} \quad (15)$$

This time also, the previous case can be viewed as a particular case of the current one, by substituting $\eta^2 = 0$ in equations (14).
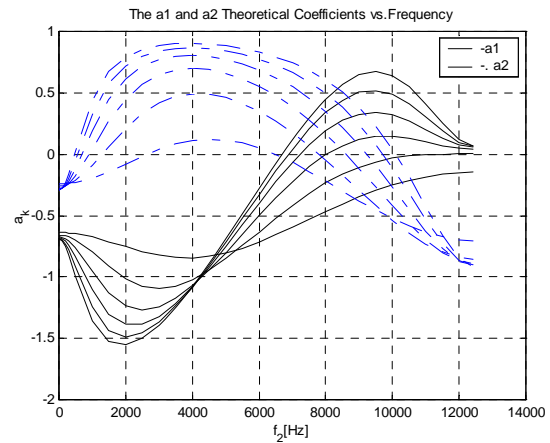

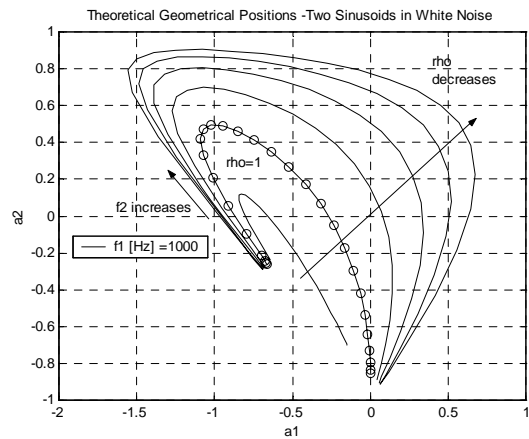Fig.10 AR coefficients versus f_2 when f_1 is fixed
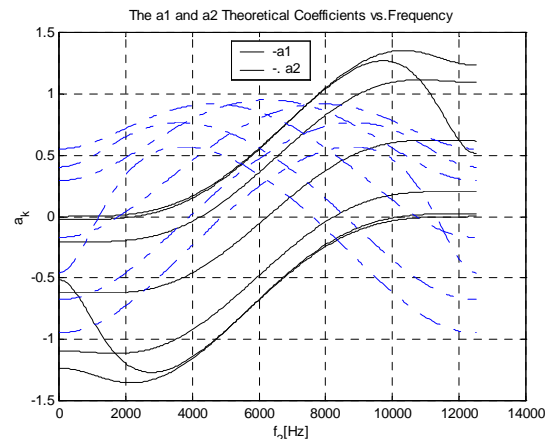

Fig.11 Geometrical positions for f_1 fixed
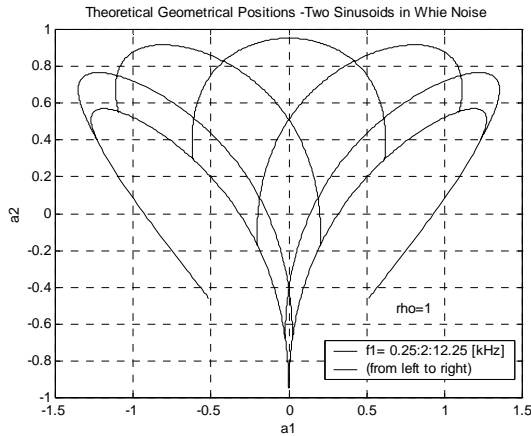

Fig.12 AR coefficients for ρ fixed
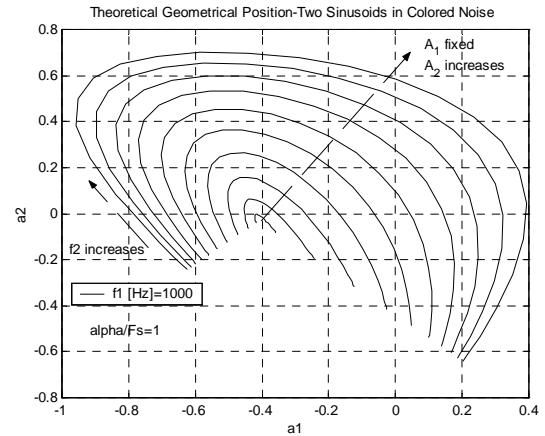
Fig.13 Geometrical positions for ρ fixed

3.1.5 Two sinusoids in colored noise

Finally, the two AR parameters for the fifth case become:

$$
\begin{cases}
a_1 = \dfrac{q \cdot \left(r - \rho^2 - 1 - \eta^2 \right)}{\left(\rho^2 + 1 + \eta^2 \right)^2 - q^2}; \\[2ex]
a_2 = \dfrac{q^2 - r \cdot \left(\rho^2 + 1 + \eta^2 \right)}{\left(\rho^2 + 1 + \eta^2 \right)^2 - q^2}; \\[2ex]
q = \rho^2 \cos y + \cos vy + \eta^2 e^{-\frac{\alpha}{F_s}}; \\[2ex]
r = \rho^2 \cos 2y + \cos 2vy + \eta^2 e^{-2\frac{\alpha}{F_s}}.
\end{cases}
\tag{16}
$$

In analogy with the previous sub-sections, Fig.14 to Fig.17 illustrate the a_1 and a_2 theoretical variations corresponding to this case. The obtained curves look somehow similar with the ones from the previous cases depending on the amount of the harmonics or noise allowed to be contained in the signal.
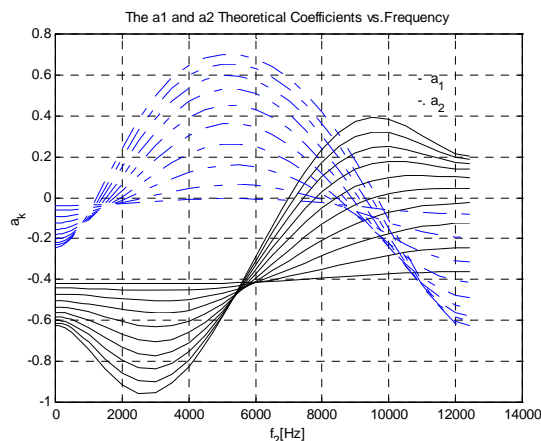

Fig.14 Case five: AR coefficients variations with 2nd harmonic frequency when f_1 is fixed.
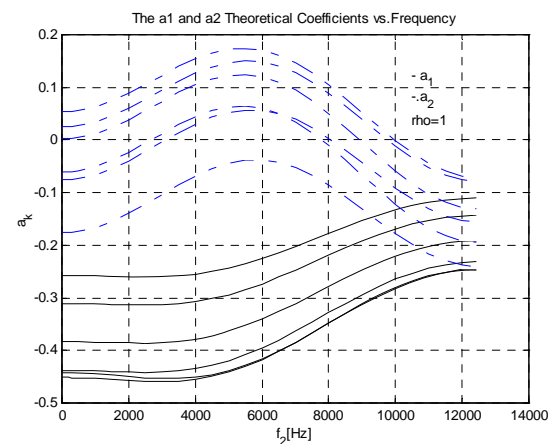

Fig.15 Geometrical curves for f_1, A_1, α and σ_w^2 fixed

Fig.16 AR coefficients families of curves when ρ is fixed, while f_1 is taken as parameter
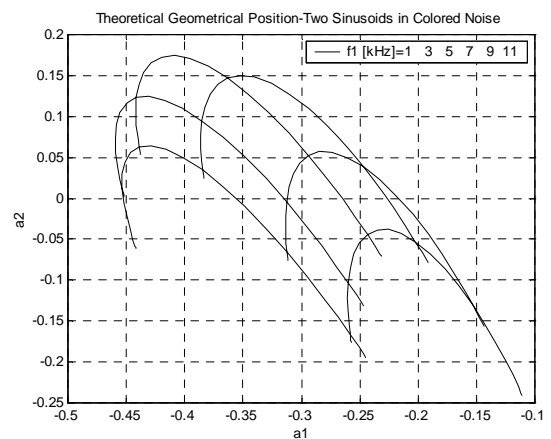

Fig.17 Geometrical curves (from left to right) when f_1 is taken as a parameter and ρ is fixed.

The parameter α from (16), that reflects the noise "coloring" degree, produces the asymmetrical aspect of the curves in Fig.17 (i.e. as α decreases, the inclination of the graphs increases to the left).

3.2 Algorithm behavior in burst-type environments. Computer simulation results.

One of the most common problems encountered in leak signals is the non-stationary burst-type component that overlaps over the basic signal. This paragraph studies the proposed algorithm's behavior using computer generated burst perturbations of exponential type with a random Poisson occurrence [6],[9]. Both amplitude and concentration of the induced bursts were varied in order to establish the algorithm's robustness. Fig.18 presents a typical non-stationary signal obtained by computer simulation, similar to signals encountered in practice.
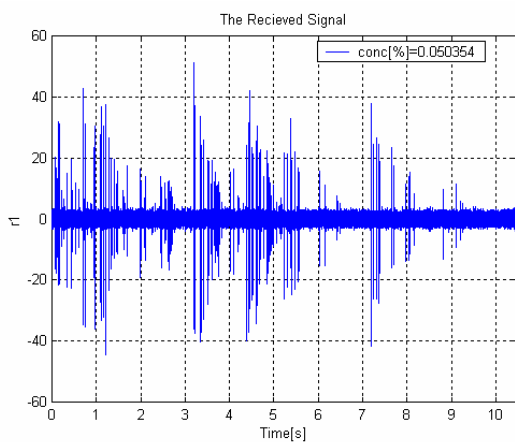

Fig.18 A typical test signal with burst component

In order to be able to evaluate the effect of the burst-type component on the algorithm's performance the burst signal-to-noise ratio factor (*BSNR*) is defined as in:

$$BSNR[dB] = 10 \cdot \lg\left(\frac{P_s}{P_b}\right) \qquad (17)$$

where P_s and P_b denote the power of the primary source signal and the burst component, respectively. Fig.19 displays the positions of the second order AR coefficients for a computer generated stationary source signal (blue marks) and for the signal in Fig.18, (green marks). It can be observed that the distance measure *L* between the data gravity centers (red marks) is practically not affected by the existence of the burst component, due to the segmentation procedure performed by the algorithm. Although the burst component produces a scattering effect on the AR coefficients' locations, their median values remain stable to the environment's abrupt variations. On the other hand, the total power, *P* of the observed signal increases if burst events occur. This can affect the decision results in

detection systems based on monitoring the power increase above a detection threshold [2],[3].
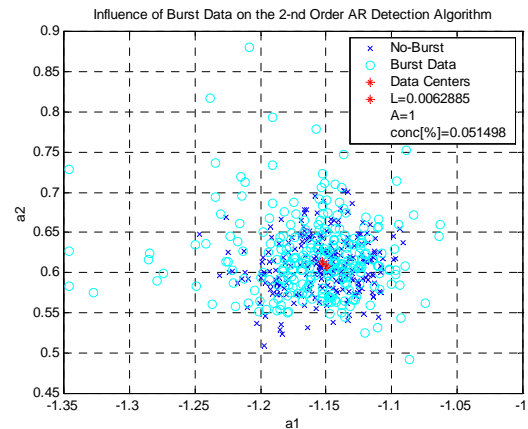

Fig.19 AR coefficients' locations in computer burst simulated environments

Fig. 20 and fig. 21 present comparative simulation results corresponding to the proposed detection algorithm and to the basic algorithm of monitoring the power increase of the received signal, respectively. In the first case, presented by fig. 20, the burst concentration was kept constant and the burst amplitude varied, while in the second case (fig.21) the burst concentration was varied and the burst amplitude was kept constant. The random variables *L* and *P* were obtained by choosing the maximum values among a large number of trials and are represented versus the burst signal-to-noise ratio factor. The maximum limits of these variables in the no-burst case are also shown (blue dotted lines). It can be observed that, for the considered domains of burst amplitudes and concentrations, the *L* variable does not exceed the maximum limit value corresponding to the no-burst case, while the random variable *P* always exceeds its maximum limit.
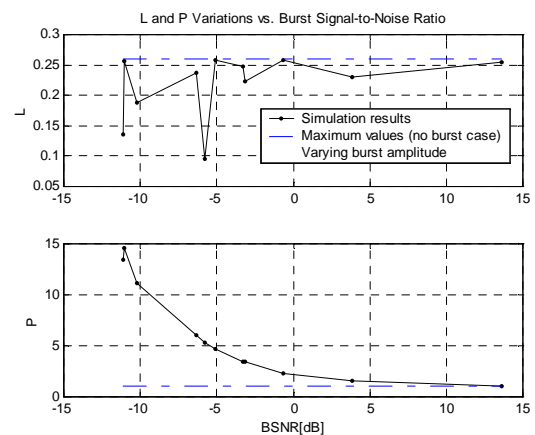

Fig.20 Comparative simulation results for variables L and P. Varying burst amplitude case.