Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 52(66), Fascicola 2, 2007

Parametric Analysis and Spectral Whitening of Signals Generated by Leaks in Water Pipes

Raul Ionel[1] Alimpie Ignea[2]

Abstract – This paper presents a way to determine the best modeling algorithms for working with signals generated by water pipe leaks. Three methods of parametric modeling are presented in this paper: auto-regressive modeling AR, moving average MA modeling and auto-regressive – moving average ARMA modeling. From these methods, the auto-regressinve modeling is the best one for analysing signal sequences from water pipe leaks. A special MATLAB Toolbox was used in order to work with the signals and the parametric models. The name of the Toolbox is ARMASA. Several programs were written in order to work with ARMASA functions and with the leak signals. The influence of signal length and number of estimation coefficients, are studied in order to show which parametric modeling method works best with signals generated by pipe leaks. The conclusion is that for these type of signals, the AR autoregressive model is the optimal solution. With the help of the obtained spectral distribution values, we can further analyze the signals in order to find with precision the position of the leaks. After the exact choice of a parametric modeling algorithm, in ths case the AR model, we are able to see the benefits of this choice when dealing with spectral analisys. Signal whitening can be used in order to improve the quality of the Cross Correlation Function (CCF).
Keywords - parametric modeling, water pipes, leak detection, leak location, MATLAB, ARMASA Toolbox, Cross Correlation Function, signal whitening.

I. INTRODUCTION

The flow of water trough a pipe generates specific auditive (noise) signals. If the pipe has leakage points or other faults, then we face problems of liquid loss. These problems must be solved, by locating with precision, the position of the leaks. The possition of the leak, must be found with the highest accuracy. When dealing with pipes that are very long (measuring kilometers), the leaks must be located with an error of a few meters (less than 5 meters).

The analysis of data sequences (noise signals from pipe leaks) by means of parametric modeling is a modern alternative which can be used in this domain. With the help of improved software applications and hardware possibilities, we are allowed to use applications based on parametric modeling (which involve lots of calculations) and to continuously monitor time varying processes.

As mentioned in literature, the use of non-parametric methods of signal processing is more suitable for periodical signals. When dealing with random signals (signals from water pipe leaks), the use of non-parametric methods is considered "quick and dirty" [3].

A more accurate approach would be the use of parametric methods. We are interested in determining the spectral distribution of the signals and also the ease and the calculus volume which is involved. Furthermore, we are interested in having a program that works with the signals in an automatic way. One should be able to determine the spectral distribution of the signals, with the help of automated parametric modeling methods (automated application), without much knowledge about signal processing techniques.

ARMASA is a collection of MATLAB programs that helps the user perform signal processing algorithms in an automated manner. Some of the offered features involve automatic spectral analysis.

The use of parametric modeling methods (AR, MA, ARMA) can turn the application for signal analysis into an automated program.

After determining which method is best for analyzing the signals generated by water leaks, we can proceed with the calculation of the CCF.

II. THE INSTALLATION

An experimental pipes installation is presented in the following image. With the help of this installation, leak signals were acquired and analyzed.

Piezoelectric sensors are placed on both sides of the simulated Leak A. The purpose of the sensors is to simultaneously acquire pairs of signals generated when water comes out of the pipe trough the simulated leak. When the faucet is opened, water came out of the pipe. Noise signals are sent from the simulated leak to the sensors trough the pipe material (metal or PVC) and trough the liquid that flows inside the pipe.

The sensors were placed, at about the same distances, on a straight part of the pipe in order to avoid possible perturbations which appear at pipe elbows.

[1] Faculty of Electronics and Telecommunications, Measurement and Optical Electronics Dept.
Bd. V. Parvan 2, 300223 Timisoara, Romania, e-mail raul.ionel@etc.upt.ro
[2] Faculty of Electronics and Telecommunications, Measurement and Optical Electronics Dept.
Bd. V. Parvan 2, 300223 Timisoara, Romania, e-mail alimpie.ignea@etc.upt.ro

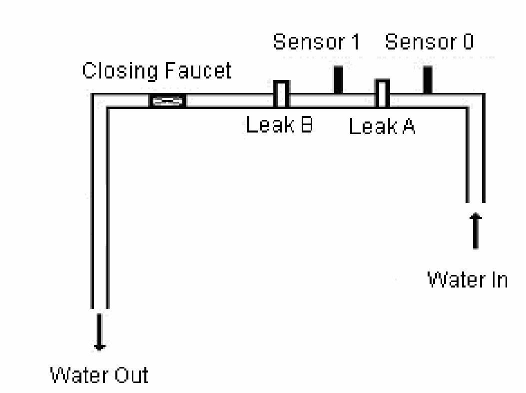

Fig.1. Experimental Pipe Installation.

III. SIGNALS CHARACTERISTICS

Several pairs of signals were acquired for different water debits trough the leak. It is expected that when dealing with small leaks (small leak debits), the spectral power distribution of the signals should cover higher frequencies. For larger leaks and larger debits, the spectral distribution should cover lower frequencies.

The signals were amplified and then transmitted to a data acquisition board for analog to digital conversion. The sampling frequency at which the signals were acquired was *Fs =25 kHz*. Each signal file contains 16384 samples. The sampling period is $Ts = 40\ \mu s$. Each signal sequence lasts *0.665* seconds.

An important aspect deals with the turbulent or laminar (non-turbulent) flow of the water trough the pipe. For the same experimental conditions, without any changes to the installation, different sounds could be heard. The conclusion was that in some cases, the laminar or normal flow of water is replaced with a turbulent flow.

The following tables, show the recorded laminar and turbulent signals, for different water debits.

Signal Pair	Leak Debit
Fa2	0,4 l/min
Fa3	0,7 l/min
Fa4	1,1 l/min
Fa5	1,2 l/min
Fa6	1,46 l/min
Fa8	3 l/min
Fa11	4,36 l/min
Fa12	7,74 l/min
Fa13	9,6 l/min
Fa14	10,4 l/min
Fa15	13,3 l/min
Fa16	16 l/min

Table 1. "Leak" Signal Pairs – laminar flow.

Signal Pair	Leak Debit
Fa8tur	3 l/min
Fa10t	3,88 l/min
Fa10tur	3,88 l/min

Table 2. "Leak" Signal Pairs – turbulent flow.

It is also important to look at the power of these signals. We should be able to see that for turbulent flows, the power of the acquired signals is higher than the power for laminar flow, at the same leak debits. The following tables show the power of the signals, in μW, for different leak debits.

Leak Debit	Power - Signal 0	Power - Signal 1
0,4 l/min	4.9715e+003	5.0098e+003
0,7 l/min	6.2775e+003	6.7602e+003
1,1 l/min	3.2664e+004	3.1136e+004
1,2 l/min	4.4374e+004	3.8689e+004
1,46 l/min	2.2698e+004	2.2709e+004
3 l/min	2.6108e+004	2.2136e+004
4,36 l/min	8.3687e+004	6.5207e+004
7,74 l/min	2.3702e+005	1.6604e+005
9,6 l/min	2.7193e+005	2.3001e+005
10,4 l/min	1.5093e+005	1.9958e+005
13,3 l/min	6.4635e+004	1.0561e+005
16 l/min	1.0812e+004	2.0568e+004

Table 3. "Leak" Signals Powers – laminar flow.

Leak Debit	Power - Signal 0	Power - Signal 1
3 l/min	4.0533e+005	2.5927e+005
3,88 l/min	2.3010e+005	1.6098e+005
3,88 l/min	3.9009e+005	2.8840e+005

Table 4. "Leak" Signals Powers – Turbulent flow

The following image shows a pair of such signals for 0.7 l/min debit, laminar flow –(Fa3).

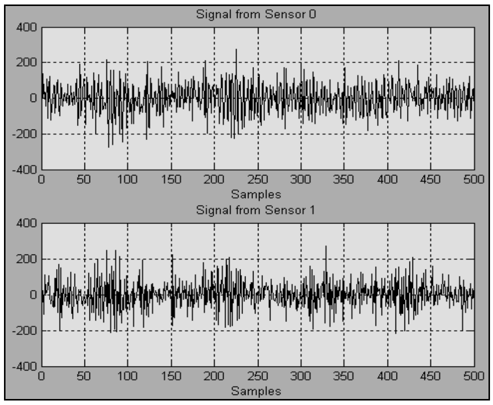

Fig.2. Leak Signals – 0.7 l/min. debit.

From the following image, one can see that the powers of the signals acquired for laminar flow, seem to follow a parabolic distribution depending on the debit. For turbulent flow, the powers of the signals are above this parabolic distribution.

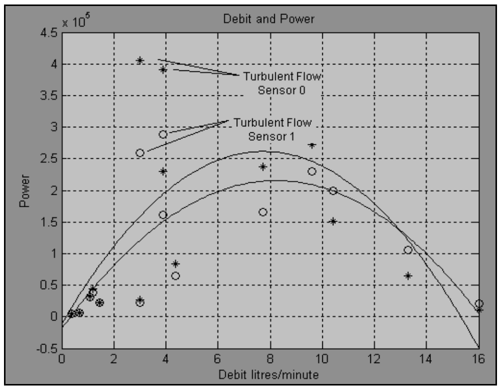

Fig.3. Powers and Debits dependency.

IV. PARAMETRIC MODELING ALGORITHMS

The parametric modeling process is based on the following idea. If we have a set of data (a set of signal values), we can characterize that set with the help of a linear filter which has a white noise as input. In order to do that, we have to determine the power of the white noise signal and the parameters that characterize the filter. The use of such models is highly recommended for dealing with random signals. For periodic sequences, the use of periodograms is prefered.

From this analisys, the appropriate model that characterizes the data set can be selected. Three such models will be presented and for each of them, several tests will be made in order to determine which model is best for characterizing the acquired leak signals.

The transfer function, *H(z),* of the linear filter is presented in relation (1).

$$X(z) = H(z)U(z) = \frac{b(1)+b(2)z^{-1}+....+b(n+1)z^{-n}}{1+a(2)z^{-1}+....+a(m+1)z^{-m}}U(z) \qquad (1)$$

The output of the linear filter is represented by $X(z)$. This is the resulting signal characterization. The white noise input is represented by $U(z)$.

The first model is the auto-regressive or recursive model (AR). The value of the present sample, is determined from the past samples values of the sequence. The transfer function of the model is presented in relation (2).

$$H^{MA} = \frac{X(z)}{U(z)} = \frac{1}{1+\sum_{i=1}^{p} a_i z^{-i}} \qquad (2)$$

The order of the model is p. As one can see, the numerator is 1, so n is 0. As an example, we can determine the AR model for $p = 1$. The relation (1), becomes:

$$U(z) = X(z) + a_1 z^{-1} X(z) \qquad (3)$$

The present sample value, depends on the white noise input and on the value of a prior sample.

$$x[k] = u[k] - a_1 x[k-1] \qquad (4)$$

In order to determine the value of a sample, which depends on the values of p other prior samples and the white noise, one can write the following relation.

$$x[k] = u[k] - a_1 x[k-1] - a_2 x[k-2] - .... - a_p x[k-p] \qquad (5)$$

As it will turn out, this kind of model is most appropriate for dealing with signals coming from leaks in pipe systems.

The next model is the moving average model (MA). The denominator from the filter transfer function is 1. This transfer function is presented in relation (6) and the relation for determining one sample in relation (7).

$$H^{MA} = \frac{X(z)}{U(z)} = 1 + \sum_{j=1}^{q} b_j z^{-j} \qquad (6)$$

$$x[k] = u[k] + b_1 u[k-1] + b_2 u[k-2] + ... + b_q x[k-q] \qquad (7)$$

The last model, which in a combination of the ones presented above, is the auto regressive – moving average model (ARMA). The transfer function for this case is presented in relation (8) and the equation for a specific sample is presented in relation (9).

$$H^{AR} = \frac{X(z)}{U(z)} = \frac{1+\sum_{i=1}^{p} b_i z^{-i}}{1+\sum_{j=1}^{q} a_j z^{-j}} \qquad (8)$$

$$x[k] = u[k] + b_1 u[k-1] + ... + b_q u[k-q] - a_1 x[k-1] - ... - a_p x[k-p] \qquad (9)$$

The main concern is to determine the spectral density of the signals, if we know the power of the white noise and the transfer function model for the filter. We can calculate the spectral density of the signal with the following relation. The power of the white noise is represented by σ_U^2 and the transfer function of the filter is represented by $H^{model}(e^{j\Omega})$.

$$S_x(e^{j\Omega}) = \sigma_U^2 \, |H^{model}(e^{j\Omega})|^2 \qquad (10)$$

As an example, for an AR model, where $p = 1$, we can write the following transfer function.

$$H^{model}(e^{j\Omega}) = H^{AR}(e^{j\Omega}) = \frac{1}{1+a_1 z^{-1}} \qquad (11)$$

In this function, $z = e^{j\Omega}$. The spectral density relation can be written as follows.

$$S_x(e^{j\Omega}) = \frac{\sigma_U^2}{|1+a_1 e^{-j\Omega}|^2} = \frac{\sigma_U^2}{1+2a_1\cos\Omega + a_1^2} \qquad (12)$$

V. AUTOMATED SPECTRAL ANALISYS

The ARMASA Toolbox, is a collection of programs that can be used with MATLAB, in order to implement applications that deal with parametric modeling and spectral density estimation.

The quality of spectral estimation by means of AR models mainly depends on the length of the input sequence and on the algorithm used to make the estimation (Burg or Yule-Walker).

ARMASA uses the Burg method in order to calculate the estimation coefficients.

The Burg method decreseases the prediction error for the estimation. For a system of order *r*, the error at iteration *k* can be expressed by the following relation.

$$e[k] = e_r[k] = x[k] + \sum_{i=1}^{r} a_{r,i} \cdot x[k-i] \quad (13)$$

This means that to the determined value for one sample, we will add a linear combination of r past values of the sequence. Despite the fact that past values of the sequence are included in the calculus, the name of the error is ”the forward prediction error”.

Another type of error which can be defined for this *r* order system, involves the adding between a past sample x[k-r] and a linear combination of *r* future values of the sequence. This can be defined as the ”backward prediction error”.

$$b[k] = b_r[k] = x[k-r] + \sum_{i=1}^{r} a^*_{r,i} \cdot x[k-r+i] \quad (14)$$

The idea of the Burg method is to minimize the sum of these errors which appear during the estimation process.

$$E\left\{\left|\{e_r[k]\}\right|^2\right\} + E\left\{\left|\{b_r[k]\}\right|^2\right\} = \text{minimal} \quad (15)$$

The Burg method is recommended for sets of reduced length. Signals with low frequencies are well aproximated by this method.

VI. EXPERIMENTAL RESULTS

The execution of the ARMASA functions involves a number of steps and a high volume of operations. For a given signal, the program calculates all models (AR, MA, ARMA) and then makes other operations in order to determine which of them is most suitable. If one knows that for example, the AR model is the most appropriate for certain types of signals, then some parts of the calculus (dealing with MA and ARMA models)
can be skipped.

In the following table, there is presented the execution time of the ARMASEL program, as a function of sequence length and model order. The purpose of the application was to calculate the power spectrum of the input sequence. The values of the execution time are to be considered orientative. The calculus depends on the configuration of the PC which runs the program. The length of the input sequence is between [2000 - 12000] samples and the order of the model is set between [10 - 70] coefficients.

N / p	2000	4000	6000	8000	10000	12000
70	5.906	7.969	9.078	11.125	12.609	13.812
60	4.703	6.437	7.547	9.500	10.687	11.938
50	3.750	5.203	6.031	7.829	9.063	10.157
40	2.891	4.109	4.922	6.422	7.500	8.516
30	2.172	3.109	3.922	5.140	6.156	7.125
20	1.500	2.157	2.719	3.906	4.813	5.672
10	*0.891*	1.360	1.781	2.797	3.532	4.281

Table 5. ARMASA execution time as a function of sequence length and model order.

If we increase the input signal length, the influence on the execution time is not so accentuated. However, an increase in the model order can be quite costly.

As we can notice, for only 70 coefficients, the execution time of the program increases. If we want to improve the execution time of the calculus, we need to reduce the number of operations by working with only one preferred model, in our case, the AR model.

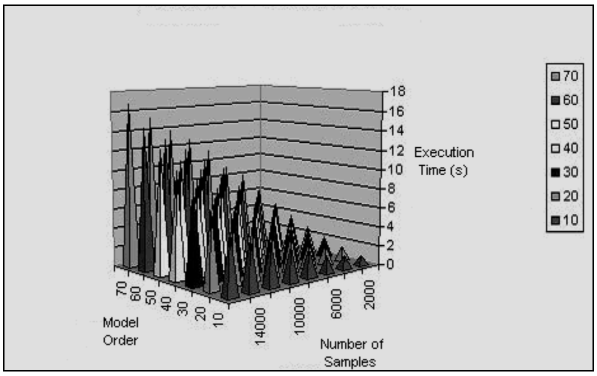

Fig.4. Execution Time as function of Signal Length and Model Order.

In order to determine which model (AR, MA or ARMA) is most suitable for working with signals coming from leaks, we can use automatic estimation on different intervals (same length) of the same signal and on intervals of different length of the signal.

As example, for one pair of signals (Fa2, 0.4 l/min/debit, laminar flow), we can start from the beginning of the sequences and increase the signal length which we analyze. In the following table, there are presented values of estimation for the signal acquired by Sensor 0. The same results are valid for the signal acquired by Sensor 1. As we can see, the AR model is most suitable.

Length	Model	AR Err.	MA Err.	ARMA Err.
2000	**AR**	885.0388	1.0768e+003	904.5822
4000	**AR**	861.1347	1.0799e+003	879.8599
6000	**AR**	805.8118	1.3258e+003	828.0465
8000	**AR**	748.3188	1.2333e+003	776.3164
10000	**AR**	728.1539	1.1962e+003	757.8679
12000	**AR**	699.6100	1.1444e+003	727.7679
14000	**AR**	692.5061	1.0967e+003	719.7552
16000	**AR**	687.3806	1.0700e+003	715.7700

Table 6. Best model for different Signal Lengths, signal from Sensor 0, laminar flow, 0.4 l/min debit.

Model	AR Err.	MA Err.	ARMA Err.
MA	1.9424e+005	1.9364e+005	1.9367e+005
AR	7.5067e+004	7.7144e+004	7.7179e+004
MA	2.0527e+005	2.0395e+005	2.0530e+005

Table 7. Best model for different Signals from Sensor 0, turbulent flow.

For signals acquired during turbulent flow, very small errors appear between the different estimation models. As a convention, we can also choose the AR model for their estimation.

The next images will show the error of the estimated model, for different signals. The horizontal axis, shows the order of the model, which ranges from [0 - 900].

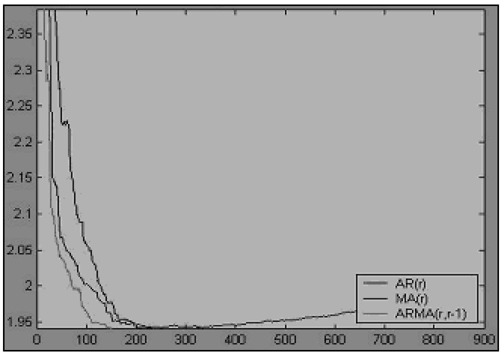

Fig.5. Error of Estimated Models,
Cfa8tur, Sensor 0, turbulent flow.

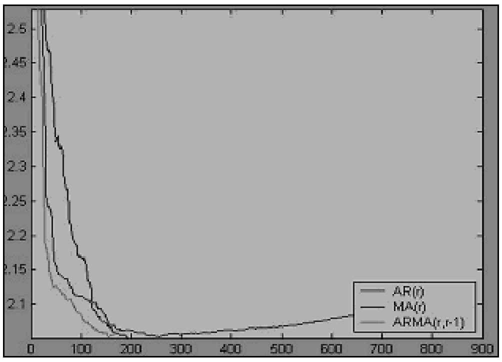

Fig.6. Error of Estimated Models,
Cfa10tur, Sensor 0, turbulent flow.

Signal	Model	AR Err.	MA Err.	ARMA Err.
Cfa2	AR	687.3806	1.0700e+003	715.7700
Cfa3	AR	1.5481e+003	2.1861e+003	1.5968e+003
Cfa4	AR	6.3400e+003	6.7141e+003	6.5560e+003
Cfa5	AR	8.4678e+003	9.9478e+003	8.8443e+003
Cfa6	AR	5.7112e+003	6.0511e+003	5.9434e+003
Cfa8	AR	6.9877e+003	7.2932e+003	7.2278e+003
Cfa11	AR	2.7309e+004	2.9361e+004	2.8732e+004
Cfa12	AR	7.4015e+004	7.6119e+004	7.6967e+004
Cfa13	AR	9.8799e+004	1.0315e+005	1.0259e+005
Cfa14	AR	3.7673e+004	4.0309e+004	3.9162e+004
Cfa15	AR	1.4800e+004	1.5467e+004	1.5323e+004
Cfa16	AR	1.7206e+003	1.7900e+003	1.7596e+003

Table 8. Best model for different Signals
from Sensor 0, laminar flow.

The conclusion which comes out of these experiments, shows that when dealing with signals which come from pipe leaks, the AR model seems to be the best for determining the spectral distribution.

The fact that the AR model is suitable for parametric modeling of signals generated by water leak can be useful when dealing with signal whitening.

We will choose a pair of signals and determine their spectral distribution. The CCF will be calculated before and after the process of whitening in order to see which calculation proves clearer and smoother.

The pair Fa3, are signals acquired at 0.7 l/min leak debit and were presented in Fig.2. The power distribution for the two signals is showed in the following images. These calculations were done before the process of whitening.

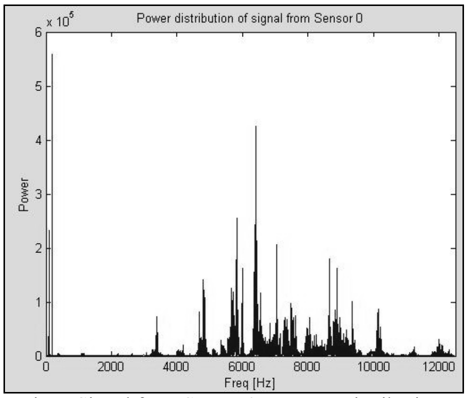

Fig.7. Signal from Sensor 0 – Power Distribution.

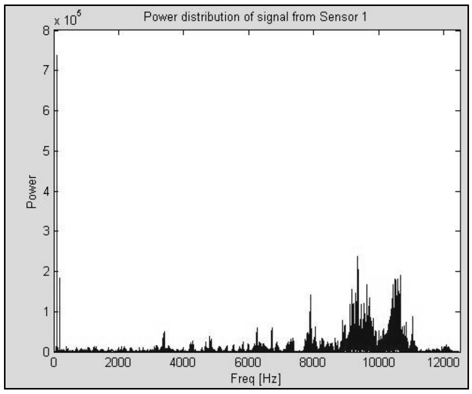

Fig.8. Signal from Sensor 1 – Power Distribution.

The calculation of the biased CCF between the two signals, before whitening, is presented in the following image. We are only concerned with the values around the maximum of the CCF. We need to determine how well the maximum stands out from the other values. The image will show the CCF for 200 points around the maximum value, in both directions. The time delay is not of concern at the moment, but as one can see it indicates that there is little difference in the placement of the sensors.

The maximum value is surrounded by other peaks which in some cases can be a source of error when estimating the time delay.

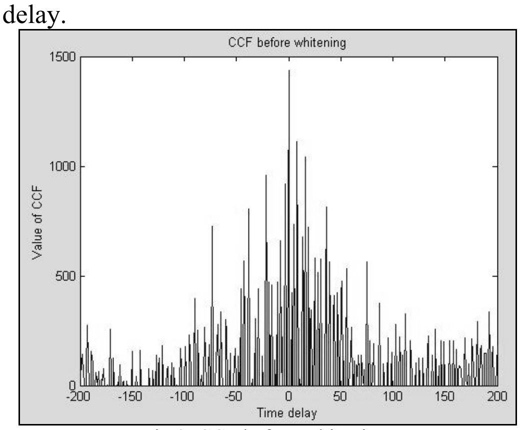

Fig.9. CCF before whitening.

The process of whitening involves the use of inverse filtering. The parametric AR modeling will be used for this purpose with the help of the Arburg Matlab function. We have used only 60 coefficients for the whitening process.

The higher the number of coefficients, the more changes occur in the spectral distribution of the signals.

The following images present the spectral distribution of the signals, after the whitening process.

As one can see there are differences, which appear because of the whitening process. The number of coefficients which needs to be used can be another topic of study. However, with a small number of coefficients, the changes should not be as obvious.

The CCF calculation reveals the fact that the process of whitening helps. Increasing the number of coefficients, increases the quality with which the maximum values stands out.

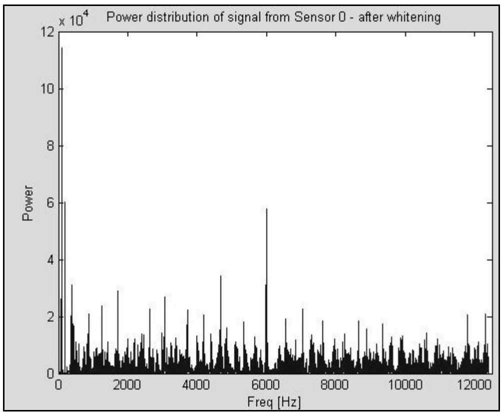

Fig.10. Signal from Sensor 0 – Power Distribution after whitening.

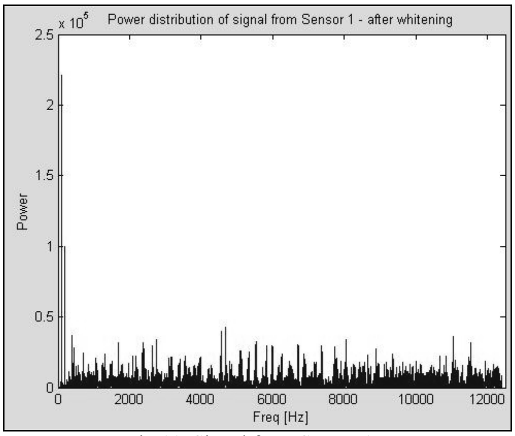

Fig.11. Signal from Sensor 1 – Power Distribution after whitening.

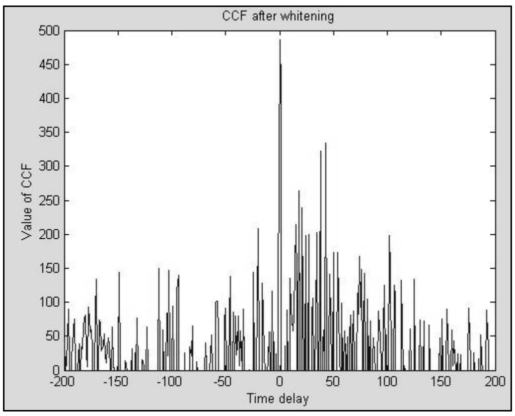

Fig.12. CCF after whitening.

VII. CONCLUSIONS

The process of spectral analysis is an important step in the process of leak detection. The idea of automated spectral analysis means that before implementing such an application, we must establish which methods are most suitable for this purpose. When dealing with random signals, the classical periodogram method of analysis is considered “quick and dirty”. Parametric methods are more suitable.

The length of the signals, but mainly the order of the model have an influence on the performances of the application. All unnecessary calculations should be eliminated from the application, as the increase in the model order significantly rises the execution time of the application.

The preferred model for analyzing this type of signals is the auto regressive (AR) model.

The process of whitening influences the spectral distribution of the signals. The CCF were calculated before and after the process of whitening in order to see which calculation proves clearer and smoother. It was shown that in the case of using the whitening process, the maximum value of the CCF was emphasized. This fact is important when we deal with establishing the position of the maximum value. This position is important for showing the time delay between two signals.

REFERENCES

[1] P. M. T. Broersen, H.E. Wensink, *Autoregressive Model Order Selection By a Finite Sample Estimator for the Kullback-Leibler Discrepancy,* IEEE Transactions on Signal Processing, Vol.46, No. 7, July, 1998.

[2] P. M. T. Broersen, H.E. Wensink, *Order Selection for Vector Autoregressive Models,* IEEE Transactions on Signal Processing, Vol.51, No. 2, February, 2003.

[3] P. M. T. Broersen, *Automatic Spectral Analysis With Time Series Models* , IEEE Transactions on Instrumentation and Measurement, Vol.51, No. 2, April, 2002.

[4] O. Hunaidi, *Detecting Leaks in Water Distribution Pipes* , Published By The Institute for Research in Constructions, Construction Technology Update, No. 40, 2000.

[5] S. Ionel, C. Gavrilescu, *Cercetări experimentale privind localizarea surselor de zgomot cu aplicații în monitorizarea şi diagnosticul tehnic al sistemelor* , Raport de cercetare la tema CNCSU 1270 din contractul 7004/1997.

[6] S. Ionel, *Estimare Spectrală cu Experimente în MATLAB* , Editura “Politehnica”, Timişoara, 2005.

[7] R. Lyons, *Understanding Digital Signal Processing, Second Edition,* Prentice Hall, http://safari.oreilly.com/0131089897, March, 2004.

[8] *** MathWorks Inc®, *MATLAB The Language of Technical Computing,* Copyright The MathWorks Inc., 1999.