Leak Detection in Liquefied Gas Pipelines by Artificial Neural Networks

Salvatore Belsito, Paolo Lombardi, Paolo Andreussi, and Sanjoy Banerjee
Center for Energy and Environmental Technologies, Consorzio Pisa Ricerche, 1 Piazza A. D'Ancona, 56127 Pisa, Italy

A leak detection system for pipelines was developed by using artificial neural networks (ANN) for leak sizing and location and by processing the field data. This system can detect and locate leaks down to 1% of flow rates in pipelines carrying hazardous materials in about 100 s. A reference pipeline was considered for practical implementation of the package. The ability of the package to withstand spurious alarms in the event of operational transients was tested. The compressibility effect, due to "packing" of the liquid in the pipeline, causes many such spurious alarms. Adequate preprocessing of the data was performed by using a computer code in conjunction with the ANN to compensate for the operational variations and to prevent spurious alarms. The package detects leaks as small as 1% of the inlet flow rate and correctly predicts the leaking segment of pipeline with a probability of success that is greater than 50% for the smallest leak. In all cases, the timely response of the system was seen as a major advantage.

Introduction

While pipelines are an efficient and economic means of transporting hazardous fluids over long distances, the risks associated with accidental releases are high. Leaks in pipelines carrying fluids such as oil, ammonia, gasoline, or chlorinated solvents can cause serious pollution, injuries and fatalities, if they are not promptly detected and repaired. For example, a leak in an NGL pipeline was responsible for the death of about 700 persons in Russia in 1988. For the period 1970–1984 in the U. S. alone there were 46 serious accidents associated with natural gas pipelines leading to 86 fatalities. Large leaks cause significant changes in pressure gradients and differences in mass flow rates at measurement points, and therefore are easy to detect. On the other hand, small leaks are more difficult to detect because changes in the usual process measurements are small. However, leaks as small as 1% of the nominal flow rate can cause the discharge of a large amount of dangerous fluid before they are detected, usually by the impact they have on the surrounding environment. The early detection of such small leaks is then the main goal of a leak-detection system.

Correspondence concerning this article should be addressed to S. Banerjee.
Permanent address for S. Banerjee: Department of Chemical Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106.
Current address of S. Belsito: Snamprogepti, San Donato Milanese (MI), Italy.

Many methods for creating leak-detection systems in liquid and gas pipelines have been proposed, mainly based on process variables (pressure, flow rate, and temperature). Usually measured in pipelines. Perhaps the most common is the line volume balance method (Ellul, 1989), based on mass conservation of the fluid in the pipeline. Data for line volume balance come from flowmeters. Usually a short-term and a long-term balance are calculated: the short-term balance provides fast response for large leaks, while it is claimed that the long-term balance will detect a 0.5% leak (of the nominal flow rate) in 3–6 hours. The volume balance can only identify a leak as being located somewhere in the line in which the fluid flow is measured. Since flowmeters are usually installed at input and exit points and seldom anywhere in between, this technique usually does not give any information on leak location along the length of a pipeline.

In his review article Ellul (1989) also describes the so-called deviation method. This method is based on a mathematical model of the pipeline system that can predict flow rates and pressures, which may then be compared with the measured values. Differences may indicate a leak. The system is reported to detect leaks in a natural gas pipeline of 1.6% of the nominal flow rate in 40 minutes. For an ethylene line, 4% leaks were detected with time delays ranging from 1 to 13

minutes, depending on the location of the leak. In another liquid-loaded pipeline the system was able to detect leaks promptly, but could not pinpoint their locations, even with relatively large leaks (4.7%).

Stouffs and Giot (1993) also present some mass-balance-based systems, using a pipeline flow model in order to compute changes in the pipeline inventory during transient flow. They highlight the importance of the packing term, and conclude that the bottom line for leak detection would be approximately 2% for steady-state flow and 3% for transient flow.

Another method, proposed earlier by Siebert (1981), was apparently used to detect and locate leaks in a gasoline pipeline during operation, which was down to 0.2% of the nominal flow rate. The signals needed were recorded with a sampling time of 1.7 s. The method, based on statistical signal analysis (cross-correlation), was also able to detect leaks in gas pipelines; 5% leaks were detected in a few hours, and were located within the pipeline with errors of about 20 km.

Billmann and Isermann (1987) proposed a method based on a nonlinear adaptive way of observing the pipeline dynamics and a special correlation technique for fault detection based on flow and pressure measurements at pipeline inlet and outlet. Nonlinear pipeline models have been used, and the friction coefficient estimated on-line. This leads to an adaptive (nonlinear) observation method. Billmann and Isermann show that the detectable leaks were greater than 2% for liquid and 10% for gas.

Wang et al. (1993) proposed a leak-detection method based on an autoregressive model. Their method requires only four pressure measurements, and is shown to be able to identify a 0.5% leakage in a 120-m-long water pipeline with no significant time delay, using a sampling time of 20 ms. Unfortunately, this sampling rate is unattainable in common acquisition systems installed in long pipelines where data transmission from remote stations and synchronization give a much lower sampling rate.

Hamande et al. (1995) discuss a model-based leak-detection system that was installed on an ethylene pipeline. The system has been in operation since 1989, and is regularly tested by valving product through flares along the pipeline. Alarms are generated when the mass imbalance becomes larger than a given threshold. Measurements of pressures and temperatures are performed at 21 locations along the pipeline, including inlet and outlet, where the mass flow rate is also measured. The system uses a real-time, transient flow model—fed by pressure and temperature measurements—that simulates the detailed design of the pipeline. The smallest detectable leak is about 7% in 1 hour (1 ton/h, the nominal flow rate in the pipeline being 15 ton/h).

Parry et al. (1992) report leak-detection performance attained in a pipeline that features a Compensated Volume Balance system. The system is interfaced with an acquisition system in order to attain real field measurements. Flowmeters are installed in intermediate locations along the pipeline. Leaks of 2% in an LPG pipeline can be detected from 46 min to 9.2 h, depending on the distance of the sections where the leaks occurs.

Zhang (1992) describes a statistical method for measuring and locating leaks in pipelines. The system uses flow and pressure measurements at the ends of a pipeline, and has been applied to detect leaks in a simulated 100-km-long gas pipeline. It has also been field tested on a 37-km-long propylene pipeline. Both the numerical simulation and the field test showed that the scheme can detect leaks as small as 1% of the nominal flow rate as well as pinpoint their locations with high accuracy.

Acoustic methods have been proposed as well. In principle, they can detect very small leaks in a short time, but they do not always work well for large networks, where there may be background noise from compressors and valves (Ellul, 1989). Furthermore, spacing between detection stations must be of the order of 100 m, otherwise the reliability is low.

Sandberg et al. (1989) set up a hydrocarbon sensor system, consisting of a sensing cable and an alarm module. This system can locate leaks very accurately (to about 20 m), with time varying response, depending on the molecular weight of the transported hydrocarbon: for example, 15 minutes are needed for gasoline and 1 hour for toluene. The cable also can be used for solvent detection. Sperl (1991) reports a leak-detection system based on a semipermeable sensor tube that was installed on the onshore part of a natural gas line and that was designed to detect a leak within about 30 minutes.

Finally, pressure waves generated by the leak provide another potential method of leak detection by measuring the pressure disturbances that travel along the line (Silk and Carter, 1995). Signal conditioning is required in order to monitor the temperature and pressure variation in the pipeline (correcting the velocity of the sound for any variation) and to account for process operations, eliminating the pressure disturbances deriving from normal processes.

The performances of some of the leak-detection systems just described are presented in Figure 1, where detection time vs. leak size is reported. The two points representing the systems documented in Siebert (1981) and Wang et al. (1993) used sampling frequencies much higher than those normally used in pipelines. Some curves refer to leak-detection systems not previously described and are not available in the literature (Uguccioni, 1996), usually for reasons of confidentiality. Apart from the aforementioned two points, which lie well below all curves, it is evident that the detection of a 1% leak with delays of the order of 100 s is beyond the capabilities of any of the leak-detection systems reported.

From this discussion of the state-of-the-art leak-detection methods, it appears that at present no single method is universally applicable. Consequently, more than one leak-detec-

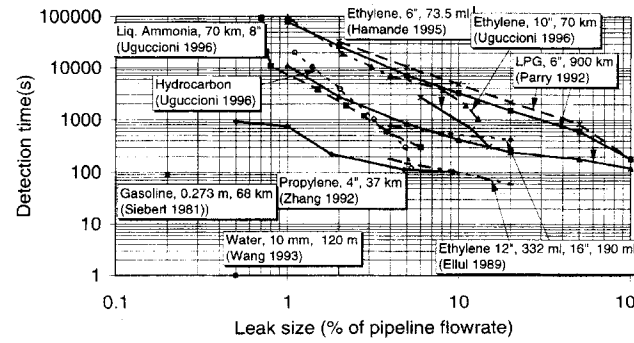


Figure 1. Performances of leak-detection systems.

tion system is generally implemented in the most important pipelines (Stouffs and Giot, 1993).

Artificial neural networks (ANNs) have attributes that can actually make them good for processing routine measurements made in pipelines, and can be used in quite innovative leak-detection systems, without requiring very high sampling frequencies. An ANN can be regarded as a nonlinear mathematical function that transforms a set of input variables into a set of output variables (Bishop, 1994), the transformation function depending on weights that are determined on the basis of a set of training examples. Training can be computationally very expensive and depends on the sizes of the networks and the number of training examples. Once weights have been calculated, processing of new data is fast. In addition to offering very high processing speeds, ANNs are, in principle, capable of learning a general solution to a problem from a limited number of examples. Be that as it may, the use of ANNs does not appear to have received much attention for leak detection, as yet. However, there are many successful applications in tasks of similar complexity. A general review is presented in Bishop (1994). There are some specific reasons why using ANNs for leak detection is quite promising (Bishop, 1994), namely, (1) it is difficult to find an adequate first-principle or model-based solution; (2) new data must be processed at high speed, and (3) the system must be impervious to noise. Another important aspect for the development of ANNs, as outlined by Bishop (1994), is that a large set of data must be available. In general, such information is either available or can be developed for pipelines.

The main goal of the present work was to develop a leak-detection system capable of detecting and locating leaks down to 1% of the nominal flow rate at an acceptable cost. The leak-detection system is based on process variables routinely measured during pipeline operations, and ANNs are utilized in order to process the field data. In the specific case of pipelines for transporting hazardous liquids, the data for ANN development should consist of field measurements of process variables (pressure, flow rate, temperature) performed where there are leaks. Cases where there are no leaks can also be used for part of the training operation, but, ideally, data measured when there are actual leaks should be available. Such data are rarely available from the field in a comprehensive form, however, so alternative strategies must be pursued where necessary. In the present work, the data for network training were generated by a computer code expressly developed for simulating flow in pipelines with and without leaks. Because the data generated must be representative of field conditions, the noise and drift of the instrumentation installed in the field have been included in the data.

Description of the Method

Reference pipeline

An existing pipeline carrying hazardous material (liquefied ammonia) was selected as a reference for the study (Table 1) (Ham and Misuraca, 1995). This pipeline is hereafter referred to as the reference pipeline.

Table 1. Pipeline Characteristics

Pipeline length	74,668 m
Pipeline diameter	8 in. (0.203 m)
Fluid carried	Liquid ammonia
No. of monitoring stations	13
Oper. flow rate (nominal)	10.4 kg/s (max)
Oper. pres. (nominal)	2.22 MPa (outlet); 2.82 MPa (inlet)

Phases of the work

Development of the leak-detection system followed four main steps. During the first phase, a review of the pipeline monitoring technology was performed, with particular attention to four existing pipelines located in northeastern Italy, for which detailed information was available. This task allowed us to determine the types of instrumentation used in those systems. Then, a deterministic pipeline flow model was developed, in the form of a numerical computer code. This numerical code was used to set up a database and develop the artificial neural networks for leak detection and location. In the final phase of the work, the leak-detection system was assessed against both numerical simulations and available field data.

Development of a deterministic model

A wide range of data, covering the entire areas of operating and faulty conditions (i.e., the presence of a leak) should ideally be available for a successful training of ANNs, but field data measured when there are leaks in pipelines are generally not available in the open literature. Due to the absence, in many cases, of these data for ANN training, the database used for the present work was derived from a computer simulator. This required the development and validation of a code, followed by testing the neural networks based on the data generated by the code against field data, to the extent available.

The numerical method that was developed is based on the conservation equations for mass, momentum, and energy, which were solved by using a finite difference scheme. It simulates both single- and two-phase flow, and is written in C++ in order to allow modularity and relative ease in changing models (Belsito, 1995). The actual model is one-dimensional in space and transient in time. It is able to simulate transient behavior of a pipeline after a leak occurs or upon changes of the boundary conditions. The code is based on a homogeneous two-phase flow model, that is, the liquid and vapor phases are well mixed and are essentially at the same velocity and temperature. This leads to one set of equations in terms of "global" variables: pressure, velocity, and enthalpy. The conservation equations that have been implemented follow:

$$\frac{\partial \rho A}{\partial t} + \frac{\partial \rho AU}{\partial z} + M = 0 \qquad \text{(Mass Conservation)} \qquad (1)$$

$$\frac{\partial \rho AU}{\partial t} + \frac{\partial PA}{\partial z} + MU + \frac{\partial \rho AU^2}{\partial z} = -gA\rho \sin\phi - \frac{2f_2 \rho AU|U|}{D} \qquad \text{(Momentum Conservation)} \qquad (2)$$

$$\frac{\partial(\rho HA)}{\partial t} = -\frac{\partial(\rho HAU)}{\partial z} - C\pi D(T - T_G) - MH + \frac{2f_2 \rho AU^2|U|}{D} \qquad \text{(Energy Conservation)}, \qquad (3)$$

where ρ is the homogeneous fluid (mixture) density (kg/m^3); A is the pipeline cross section (m^2); U is the fluid velocity (m/s); M is the leak flow rate per unit length of pipeline (kg/ms); t is time (s); z is the streamwise spatial coordinate (m); P is the fluid pressure (Pa); g is the gravity acceleration (m/s^2); ϕ is the pipeline inclination with respect to the horizontal direction (in radians); D is the pipeline ID (m); H is the mixture enthalpy (J/kg); C is the overall heat-transfer coefficient; T is the fluid temperature (K); T_G is the surrounding ground temperature (K); and f_2 is the two-phase friction factor, which can be calculated as (Sintef, 1990)

$$f_2 = f\Phi_L, \tag{4}$$

where

$$\Phi_L = 1 + \alpha f_\Phi, \tag{5}$$

where α is the void fraction, and

$$f_\Phi = \frac{15.3}{|U|(1-\alpha)}\left(1.18\left(g\sigma\frac{\rho_L-\rho_G}{\rho_L}\right)^{0.25}\sqrt{1-\alpha}\right), \tag{6}$$

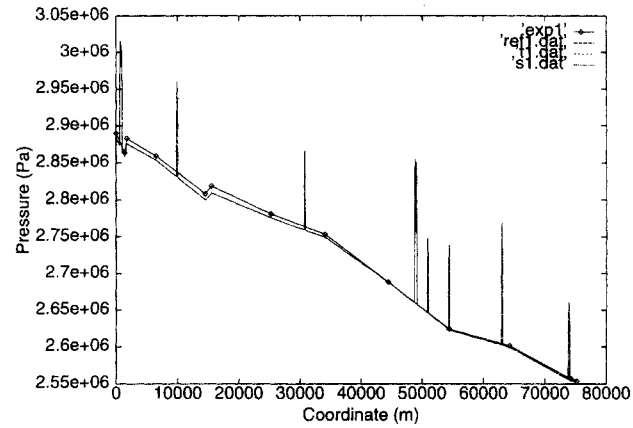
where σ is the fluid surface tension, the subscripts L and G are the liquid and gas phase indicators, respectively; and f is defined by the relationship

$$f = \frac{0.0055}{4}\left[1+\left(2\cdot 10^4\frac{\epsilon}{D}+\frac{10^6}{Re_L}\right)^{1/3}\right] \tag{7}$$

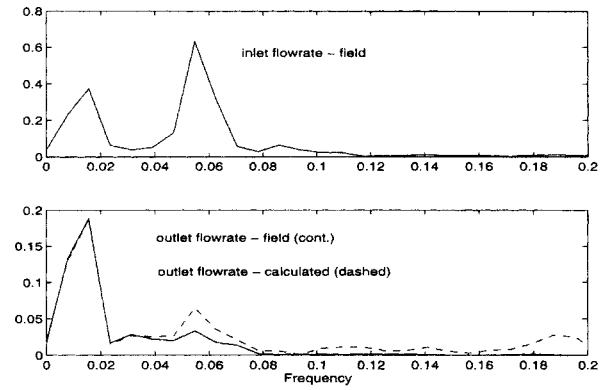
where ϵ is the pipe roughness (m), and Re_L is the liquid Reynolds number.

The code based on this algorithm has been validated against available field data. Several tests have proved the numerical strength of the code and good agreement with the data (Belsito, 1995; Belsito and Campastro, 1997). Figure 2a gives a comparison between calculated and measured pressure profiles, together with a sensitivity study on the number of mesh points utilized for the numerical model. In Figure 2a, the peaks in the calculated pressure profile are due to elevation changes corresponding, for example, to road or river crossings. These peaks are not present in the experimental profile, however, because no transducers are available at such locations. Figure 2b shows spectra of typical inlet and outlet flow rates of the reference pipeline. In the upper part of Figure 2b, a typical (measured) inlet flow-rate spectrum is shown, while in the lower part of the figure, the measured spectrum of the outlet flow rate is compared with the spectrum predicted by the code, when the inlet flow rate is utilized as a boundary condition. The agreement is very good, especially at the characteristic frequencies of the system.

The analyses performed with the code showed that in the reference pipeline two-phase flow does not occur when leaks of the size of interest for the present work occur (the leak flow rate is smaller than 10% of the pipeline flow rate). This is due to the sizable subcooling of ammonia in the pipeline (about 40 K) for the reference case, and also to the small depressurization following the occurrence of leaks (about 5 kPa for a 1% leak). Clearly, for different pipelines with different operating conditions two-phase flow can in principle occur, which the code should be able to simulate, noting that homogeneous equilibrium two-phase flow is to be expected at the relatively high mass velocities encountered in most pipelines.



(a)



(b)

Figure 2. Validation of the deterministic model against field data.

(a) Comparison between calculated and experimental pressure profiles in the reference pipeline (exp1: experimental profile; ref1. dat: calculated profile (1000 grid points); t1. dat: calculated profile (2000 grid points); s1. dat: calculated profile (500 grid points). (b) Flow-rate spectra: measured inlet flow rate (upper) and measured vs. calculated outlet flow rate (lower).

Setup of the database

The numerical code was used to generate a database for training the ANNs. For each of the simulated conditions, the time evolution of inlet and outlet flow rate, and of the fluid pressures at the 13 locations corresponding to the 13 measurement stations along the pipeline, have been calculated. Those data constitute a "pattern." About 1000 runs under different conditions have been calculated with the code, which led to the same number of patterns that constitute the database used. While developing the database, both patterns without leaks and patterns with leaks were considered. The patterns were generated by varying operating conditions (viz., the outlet pressure set to 2.2, 2.4, or 2.6 MPa; the inlet flow

Table 2. Noise Characteristics for Transducers on the Reference Pipeline

Pressure Meter		Flow-Rate Meter	
0 Avg. Noise	Drift	0 Avg. Noise	Drift
12,500 Pa (0.25% span, span 50 bar)	5,000 Pa (0.1% span, span 50 bar)	0.25% of measure	1% of measure

rate remained at 5, 7, 9, or 11 kg/s during one run), leak size (0%, 1%, 2%, 3%, 5%, 10% of the inlet flow rate), and leak location along the line (24 different locations along the line have been considered). Ultimately, a total of 972 patterns were utilized. This database proved to be adequate for successful training of ANNs, and performed well in locating leaks. We determined that a much smaller database is required for building a neural-based system of leak detection and sizing without information on leakage location along the line.

The pattern database just described does not contain instrumentation noise, which needs to be superimposed *a posteriori*. In order to do this, the measurement system used in the reference pipeline was first characterized. The pressure transducers installed in it are Honeywell Smart transmitters, and the flow rate transducers are Honeywell swirl meters (noise characteristics are reported in Table 2).

In order to be conservative while developing the system, worst-case assumptions were made when deciding the operating mode (digital or analog) of the Smart transducers, that is, the largest possible values for noise were assumed. The signals obtained by superimposing noise on the simulated trends are shown in Figure 3, which refers to a 1% leak located 31 km from the pipeline inlet and occurring at $t = 500$ s. The "noisy" flows are compared with code predictions without noise. It appears that, when a leak occurs, there is a variation in both pressure and outlet flow rate, which reach a new steady state in a few hundred seconds.

At the same time, available field data recorded on the reference pipeline (see the section on the reference pipeline) were analyzed as well. All we had, though, were data for non-leak conditions. Typical time histories and spectra are reported in Figures 4 and 5, respectively. The noise part of the signals was extracted from the field measurements by filtering out the low frequencies due to operational variations and then superposed on the code predictions without noise. Typical time histories coming from this superposition are shown in Figure 6.

Development of the Leak-Detection System

The problem has been tackled by using two different nets, one for leak detection and sizing, and one for leak location. The first system would work on-line by monitoring pipeline status and giving an alarm if a leak is detected, thus activat-

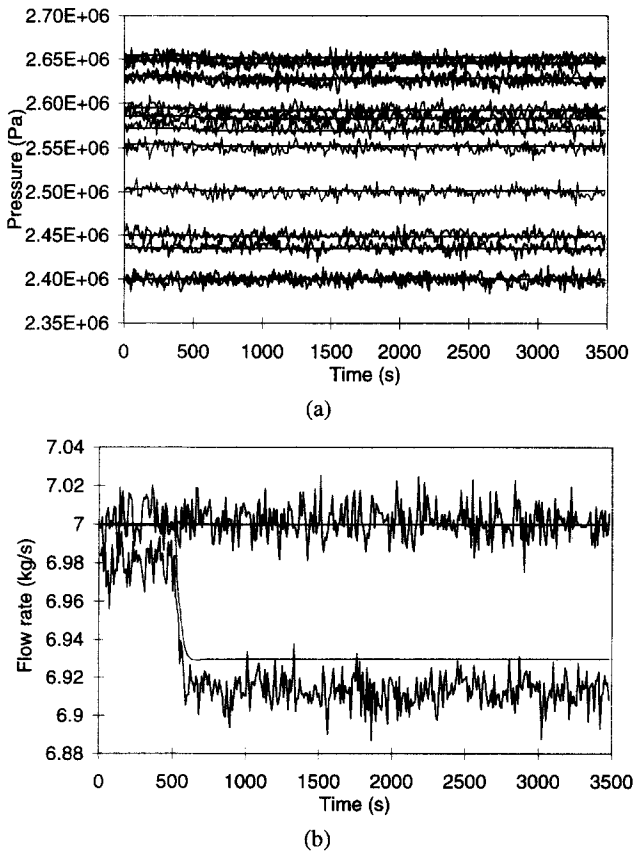


Figure 3. Pressure and flow-rate data with noise.

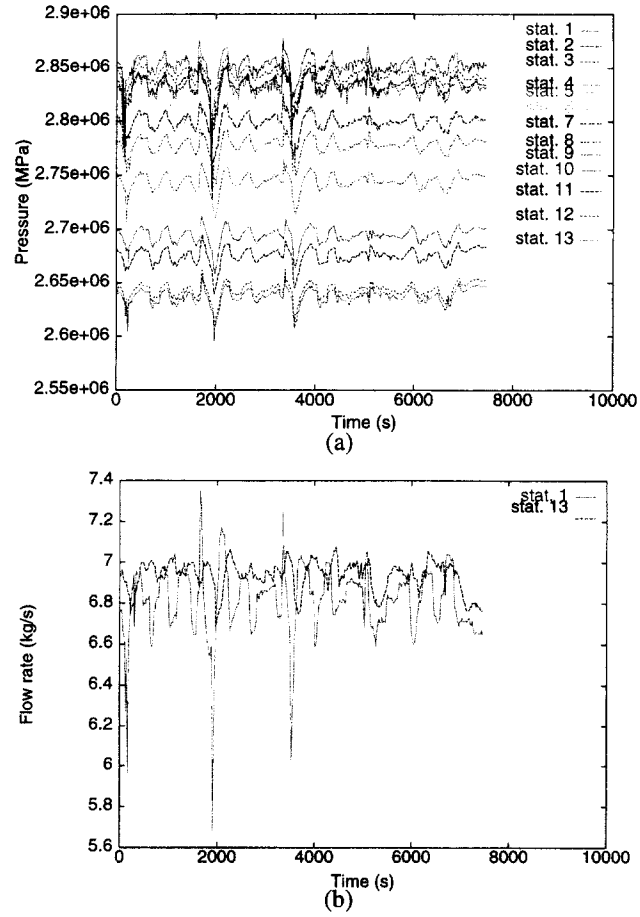


Figure 4. Field data: pressure and flow rate vs. time.

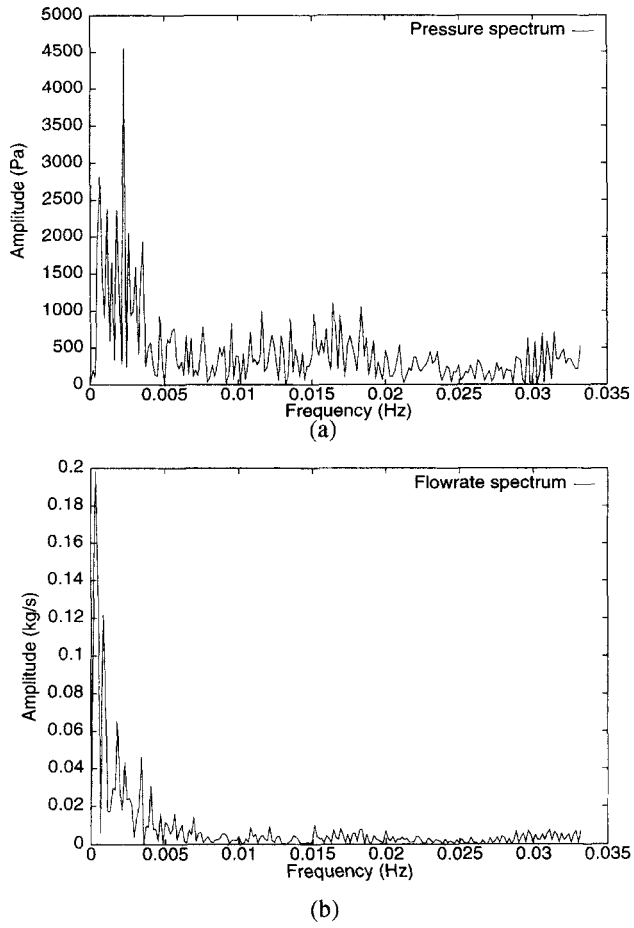


Figure 5. Field data: pressure and flow-rate signal spectra.

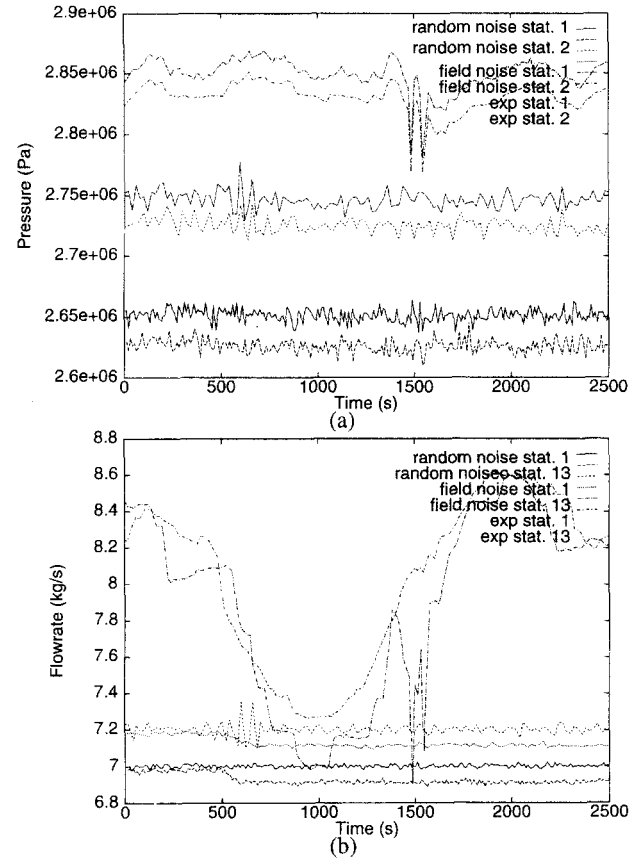


Figure 6. Noisy pressure and flow-rate signals as obtained by superposing code predictions and field noise.

ing the second system, which makes locating the leak along the line possible. For leak sizing, both data with no leaks and data with leaks are used for network training. On the other hand, for leak location the leak is known to have occurred, so data without leaks are not needed.

Development of the leak sizing network

The basic configuration of the leak sizing ANN, such as choice of architecture, learning parameters, and scaling of the input parameters, was set up using patterns without superimposed noise. The patterns used were subdivided into a training set and a test set (2/3 and 1/3 of the total number of patterns, respectively) used for verifying the predictive capability of the network. Field data were then utilized for validation.

Information contained in the patterns was compared to the pipeline's physical steady-state condition, which is attained after leak occurrence. In particular, the values of pressure at the thirteen monitoring stations and the inlet and outlet flow rates are used as input parameters. The transient part of the simulation was neglected, because training the network with these data would create a "one-shot" network that is able to recognize the transient part of the event (when the leak develops), but that is unable to give a continuous alarm when a new steady-state is reached after the leak has formed.

This first stage in developing the sizing ANN was performed because of the shorter computing time required for training a neural network with noise-free data rather than using data with embedded noise. Such a strategy is in line with the techniques suggested by Haykin (1995), and allows rapid development of the main network parameters, besides defining the optimal normalization for the physical quantities contained in the patterns. As a result of this activity, a feed-forward multilayer perceptron, using sigmoidal curves as activation functions, was derived as the first version of the leak-detection and sizing network. Such an ANN consists of three layers: the input layer has 15 nodes, each of them corresponding to one signal measured in the pipeline; the hidden layer has 7 nodes; and the single output gives the size of the leak. The 15 input nodes of the network receive the signals, appropriately scaled, from the two flowmeters and the 13 pressure meters installed along the reference pipeline. All the signals have been zero-averaged and normalized to unit standard deviation. The output of the network corresponds to the size of the leak normalized in the range 0.1–0.9. This is the linearity range within which the sigmoidal function should work (Haykin, 1995). An output of 0.1 therefore corresponds to a condition of no leak, while an output of 0.9 means that there is a 10% leak in the pipeline.

Figure 7 shows the prediction of the network for the training and test sets. The various sets are for different simulated

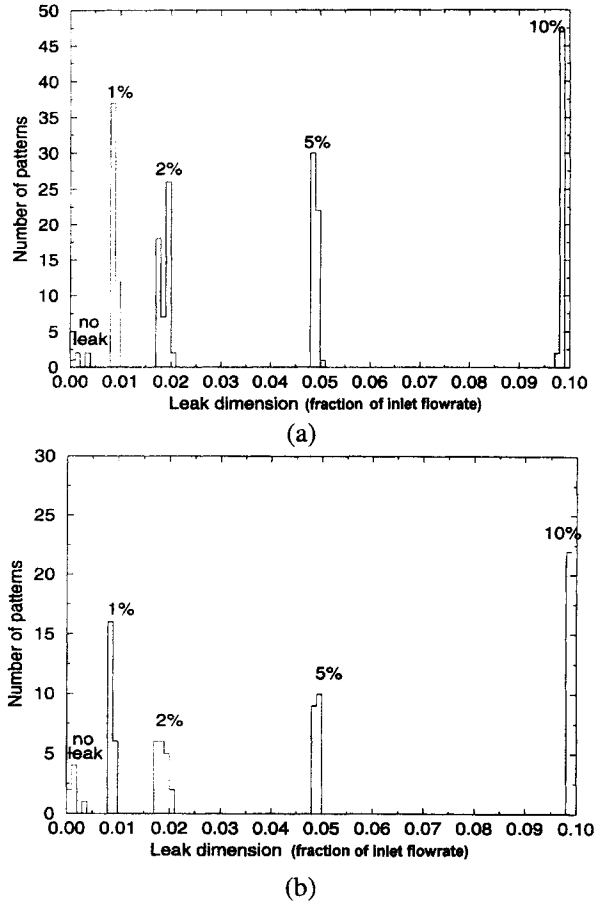


Figure 7. Leak-sizing network (noise-free signals): (a) training set; (b) test set.

leaks (no leak, 1%, 2%, 5%, 10%). The ideal action of the sizing system would be to produce five single bars centered on the value of the imposed leak (from 0% to 10%). As can be seen in the figure, the ANN is able to satisfactorily predict both leak occurrences and sizes. Furthermore, it is able to distinguish cases with the smallest leak (1% of the inlet flow rate) from cases with no leak. Since the histograms of no-leak patterns are separated from the ones with 1% leaks, no spurious alarms would be generated by the system, provided that an adequate selection of the triggering threshold is performed.

In the next step, dealing with the issue of noisy signals (with superimposed randomly generated noise) was tackled. The network architecture that was used before was not capable anymore of recognizing the size of a leak with enough accuracy to distinguish the 1% leak cases from patterns where leak was not present in the line. Since ANNs have been reported to be relatively insensitive to noise (Bishop, 1994), the first step was to obtain a more suitable ANN architecture, but a simple increase in the number of neurons and hidden layers did not yield significant improvements, any more than did applying the radial basis function networks instead of multilayer perceptrons. Using different learning paradigms did not improve the performance of the ANN, either, so the study was refocused on sensitivity analyses of noise sources and identification. Such analyses revealed the strong effect that flowmeter drift has on network performance and led to a number of actions that improved the ANN performance:

- Filtering the signals with moving average techniques: averages calculated over 9 data points (135-s time window) have been utilized;
- Providing ANN with information related to flowmeter drift: this was done by giving as input the difference between the measurements of the two flowmeters some *Dt* seconds before the actual time. Optimization of this time delay vis-à-vis the reference pipeline and flowmeters characteristics, led to $Dt = 1800$ s. In fact, the value of *Dt* is strictly connected to the time constants of the line (in terms of inlet/outlet flow-rate correlation), which does not depend very heavily on operating conditions (e.g., operating variations on pressure and flow rate). The difference between the flowmeter measurements tells us something about the flowmeter's drift, which is herein assumed not to vary significantly during *Dt*.
- Various patterns have been derived by superimposing zero average noise and randomly selected drift, from a single set of data generated by the code. In this way, the ANN faces situations in which different inputs (because of the noise) produce the same leak size. In fact, the number of training data that was needed for the case of noisy signals was about three times higher than for the case with noise-free variables.

The ANN was successfully trained with the same general procedure used for the case without noise, described earlier in this section (e.g., the raw data for averaging were taken only from the stationary parts of the signals), and performing the actions just indicated is a multilayer feedforward neural network made of 23 nodes with 1 hidden layer (16-6-1). It is very similar to the network developed for the case in which noise is not present, except that one more input added. This is the input that represents the difference, appropriately scaled, between the measurements of the flowmeters, at a certain time before the actual time. The performance of this network is summarized by the results reported in Figure 8.

The neural network appears to be able to detect the smallest simulated leak (1% of the inlet flow rate) and appears to distinguish between situations where no leak is present in the line from cases where the smallest leak assumed is present in the line. We believe that widening the flow-rate operating range, as one example, is not likely to affect the ANN performance or architecture, though testing with a wider database would be desirable.

Robust Behavior of the Leak Sizing Network When Using Normal Operation Field Signals. So far, a leak sizing ANN has been successfully trained using noisy, steady-state, numerically generated patterns. Designing a system to be operated in the field requires that routine, relatively slow transient operations be treated by the system without giving rise to spurious alarms. The main difference between such transient operations and the steady-state conditions previously utilized is that, while during steady-state conditions the input and output flow rate differ only by the leak flow rate and the noise, in transient conditions the "packing effect" due to liquid compressibility becomes one more contribution to flow-rate unbalance. This contribution must be included during network training or when preprocessing the transient data. In this phase of the work, then, the leak sizing network was fed with real field data. Typical field data are shown in Figure 9,

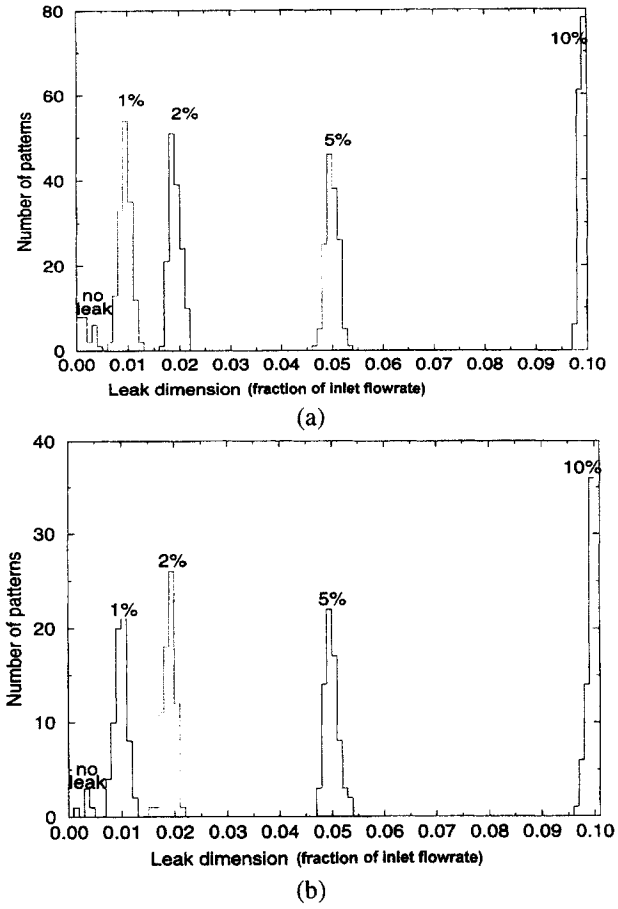


Figure 8. Leak-sizing network (noisy signals): (a) training set; (b) test set.

where about 8000 s of inlet and outlet flow rates for the reference pipeline (see the section on the reference pipeline) are reported. The performance of the ANN fed with the field data is shown in Figure 10, indicating, at several time intervals, the presence of a leak of up to 5% of the actual inlet flow rate in size, whereas in reality no leak is present. In order to avoid so many spurious alarms from the leak sizing system, it was necessary to preprocess the data. The actual (measured) outlet flow rate and code-predicted outlet flow are then fed to the ANN, together with the usual pressure signals (see Figure 11). This allows compensation for the packing effect of the fluid that would otherwise be present in case both measured inlet and outlet flow rates are directly compared. Apart from noise and modeling approximations, measured and calculated outlet flow rates would differ only by the contribution made by the leak, as in the steady-state case. This is the main reason for the improved behavior by the neural network.

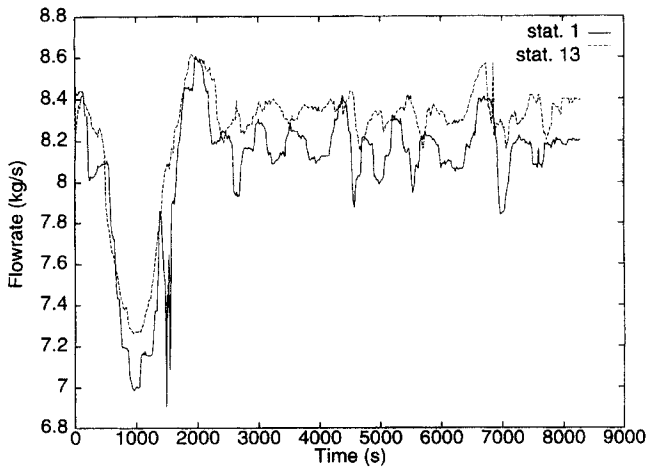


Figure 9. Flow-rate field data showing typical operational transient.

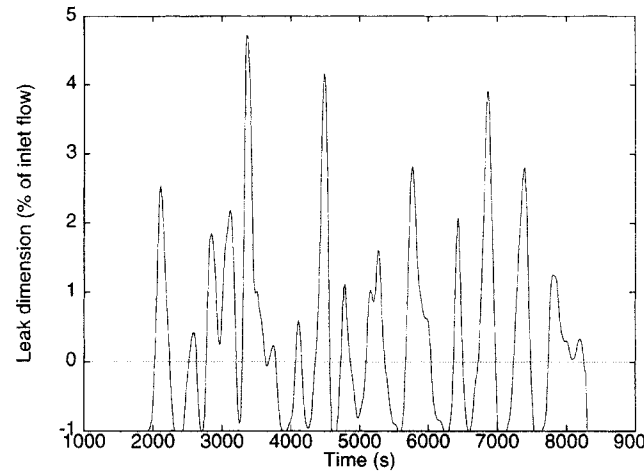


Figure 10. Response of the leak-sizing ANN to field signals with no-leak: nontreated signals.

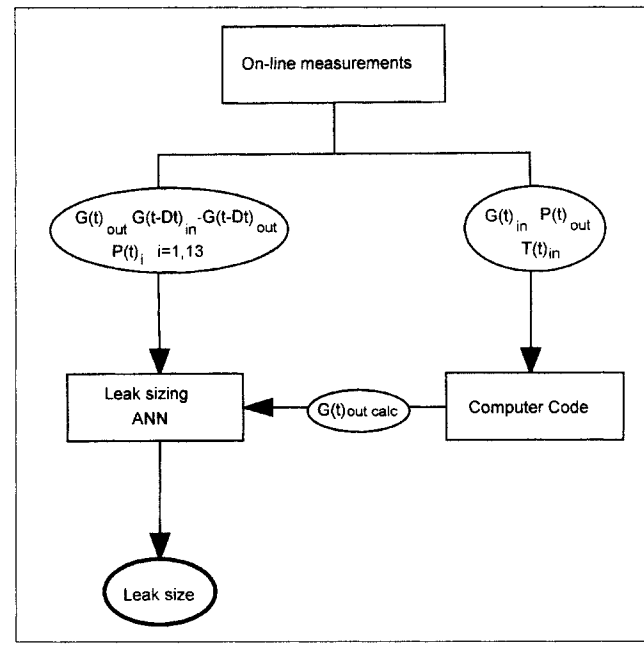


Figure 11. Leak-sizing ANN system.

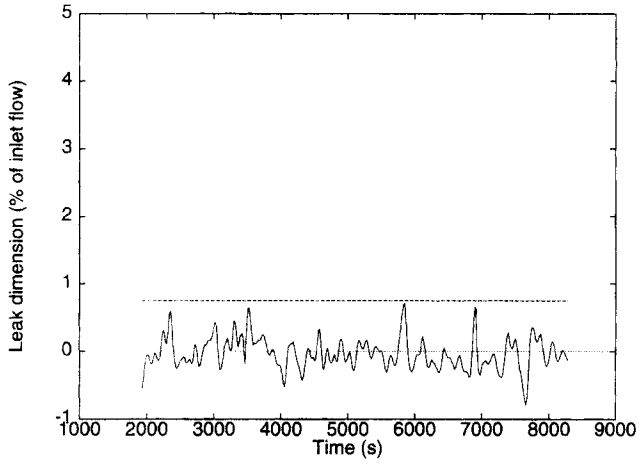


Figure 12. Response of the leak-sizing ANN to field signals with no leak: treated signals.

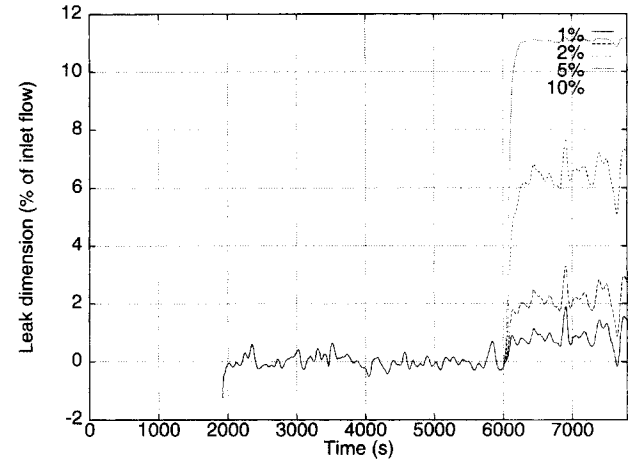


Figure 14. Response of the leak-sizing ANN to field signals with effect of leak simulated.

Figure 11 gives the logic blocks of the leak sizing system (*Dt* being the time delay defined in the section on the development of the leak-detection system). The code that has been developed must run on-line, with the ANN producing one of the inputs. Indeed, this is possible because of computational simplicity and efficiency of the C++ computer code that was developed as part of the present activity (see the section on developing a deterministic model). The new output given by the leak sizing ANN is reported in Figure 12, which shows behavior that is fairly robust vis-à-vis transient field signals. If we think of a warning threshold set at a value of 75% (i.e., with a 25% safety margin with respect to the minimum target leak), no alarms are given in this case, as shown in Figure 12. This threshold was used in the course of the activity: whenever the system gives a leak size larger than 0.75, an alarm for a 1% leak is provided.

Leak Sizing Test. To check system leak-detection reliability, we input representative leak data. When a leak develops, a sudden drop (even if small) occurs in the pressure measured along the pipeline and in the outlet flow rate (Belsito, 1995). The validated numerical code described in the section on the deterministic model was able to calculate this typical pressure and flow-rate drop, and was used to set up the database. In order to get data representative of leak occurrence in field conditions (i.e., transient conditions), the pressure and flow rate reductions, as calculated by the code for various leak flow rates (1%, 2%, 5%, and 10% of inlet flow rate), were superposed on the field data measured on the reference pipeline in real-life operating conditions. The outlet flow rates thus obtained are given in Figure 13 for various leaks developing at 6000 s.

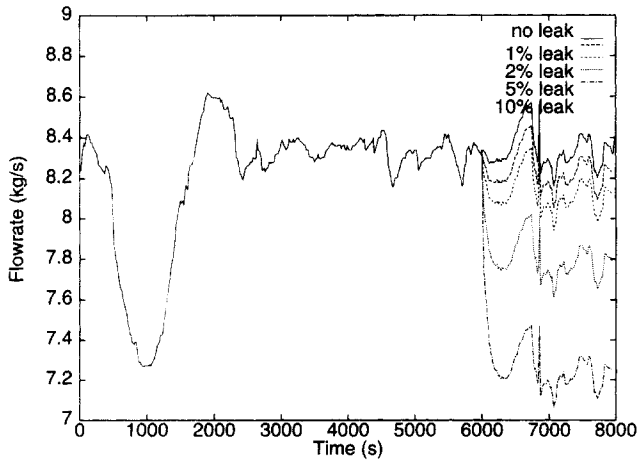


Figure 13. Field-measured flow-rate time history compared with flow-rate time history with simulated leaks superimposed.

The signals derived from field measurements, including the effect of the presence of a leak, were then fed to the system for the leak sizing described earlier. The results of this test are reported in Figure 14, where it can be clearly seen that the system is able to detect the leaks of different sizes, which give signals that exceed the threshold.

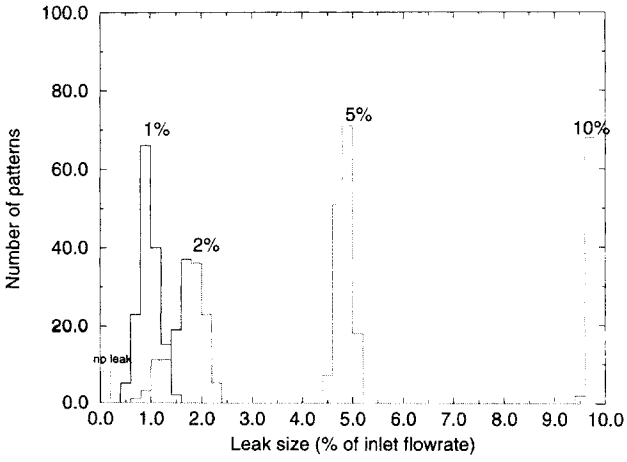
Table 3 contains the time needed for detection of the various leaks, given an alarm threshold of 75%. It can be seen that the smaller leak is actually detected in 120 s, corresponding to eight measurement points. The detection time decreases for larger leaks.

Sensitivity of the Number of Stations. The effect of fewer monitoring stations along the line was also investigated. The activity was carried out to determine the number of measurement stations needed to get acceptable performance of the leak-detection system and to understand what kind of degradation is to be expected after reducing the measurement points.

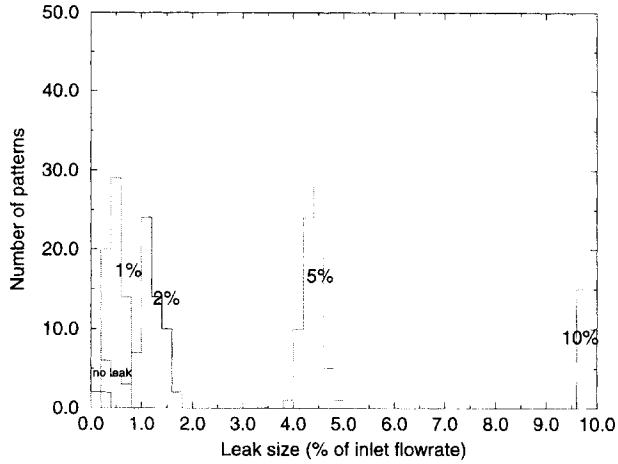
Using a back-propagation algorithm, an ANN was trained for leak sizing, using only 6 pressure measurements out of 13. Results are summarized in Figure 15, where we can see a larger spread of the predicted data around the actual value of the leak size, in comparison with the result obtained using

Table 3. Detection Time for Leaks of Different Size

Leak size (% of inlet flow)	1	2	5	10
Detection time (s/n of measurements)	120/8	90/6	60/4	45/3



(a)



(b)

Figure 15. Performance of ANN for leak detection with reduced number of stations: (a) training set; (b) test set.

all the measurements performed in the pipeline (see Figure 8). Since the data predicted for a 2% leak size overlap data for the 1% leak size, as well as the data for the no-leak condition, they are no longer separated from the events with a 1% leak, even to a small extent. This fact could give rise to spurious alarms or loss of sensitivity in the detection of small leaks.

The analysis confirmed the importance of having as many pressure transducers as possible installed along the line for leak detection, even if the ANN provides quite robust results. Improvement of the performance of the leak-detection system is expected when the number of transducers is increased. An interesting aspect was noted while training this network: training required a significantly larger number of iterations to converge compared to the ANN trained with the full set of measurements (13 pressures and 2 flow rates). This analysis and other investigations of the behavior of the present leak-detection system under conditions where the pipeline pressure transducers may be faulty (not reported in the present article for the sake of brevity), confirm that the pressure transducers provide vital information to the system, and that the operators should closely monitor their effective functioning.

Development of the leak-locating network

The ANN for locating the leak is intended to work in cascade with the ANN for leak detection. When a leak is sized, then the ANN for location is "fired" and the prediction of leak location is made.

For leak-locating purposes, the ANN should at least indicate the segment between sectioning valves where a leak is present. Our reference pipeline is subdivided in 12 parts by sectioning valves, with some of the sections very similar in length (about 10 km), while others are of very short length. To make the prediction simpler, the pipeline was presented to the locating network as if it was organized in eight different parts (nine at the beginning of the activity) of similar length, by including the shortest parts in the neighbor sections to form the database. In this way, the output of the ANN consists of several neurons, one for each of the segments into which the pipeline is divided. The neuron with the highest activation provides leak-location information.

Another key factor in the successful development of the leak-locating network was to provide as input the pressure drop due to a leak, instead of the absolute pressure measured by the transducers along the line. In fact, the leak causes a pressure drop in the various locations where transducers are located, and it generates a drop in the outlet flow rate. The quantities then reach a new steady state where the pressure values and the outlet flow rate are constant. The combination of the pressure drop amplitudes contains the information on leak location. The pressure drop due to the leak is calculated as the difference between the actual pressure value and the pressure before the leak began. In addition, the leak-detection network gives us the size of the leak. This information can also be used as an input to the leak-locating ANN. Finally, moving averages performed on 21 data points are used in order to reduce noise.

The ANN was trained with noisefree signals by using the steady-state pressure drops (13 signals) and flow-rate drops (2 signals). The output is the segment where the leak is located (9 outputs). The ANN was trained with the backpropagation algorithm in about 100,000 iterations. The results of the network are reported in Figure 16 for training and test sets. For each segment the number of patterns correctly and incorrectly classified is reported for each segment. It can be seen that the performance of the ANN is very good: 97.5% (194/199) of the patterns are correctly classified in the test set.

Because an attempt at utilizing the spatial pressure gradients for dealing with noisy signals did not give good results—which can be explained by the fact that variation in spatial gradient pressure due to leaks is so small, and so similar for leaks in different locations, that the ANN is not able to distinguish between the various patterns—the pressure-differences signals were kept as inputs and one further input was added to the network. This input is representative of the leak size, which is calculated by the network for sizing the leak and which showed very good performance in leak sizing. Moreover, the pipeline was divided into eight segments (in-

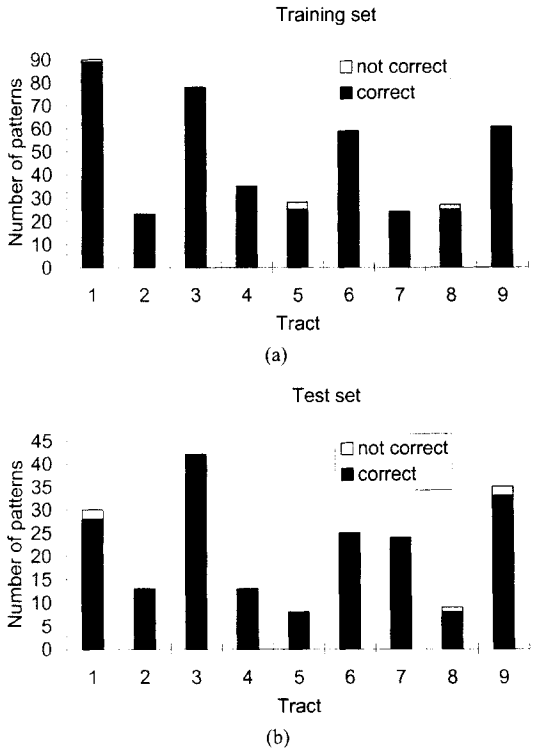
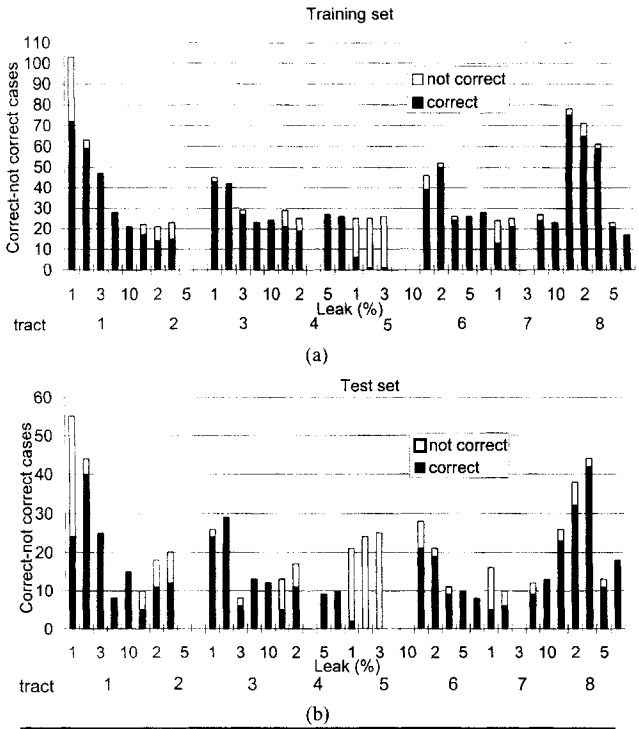


Figure 16. Response of the leak-locating ANN using signals without noise: (a) training set; (b) test set.

stead of nine) in order to get a more uniform distribution of the segment lengths. Using the transducer characteristics, random noise was superimposed and the input was filtered, with 21 time points being used to calculate each input value by moving average. The system was conceived as a feedforward 16-9-8 network and was trained with backpropagation. The results of this ANN are reported in Figure 17, where the number of correct classifications is now reported per segment and per leak size. The general performance of the ANN is quite good. Nonetheless, it can be seen that the lack of some examples for certain classes of leaks (for example, in segment 5, 5 and 10% leaks are not used for training; see Figure 17) is causing a generally bad performance of the ANN for the other leaks represented in the same segment.

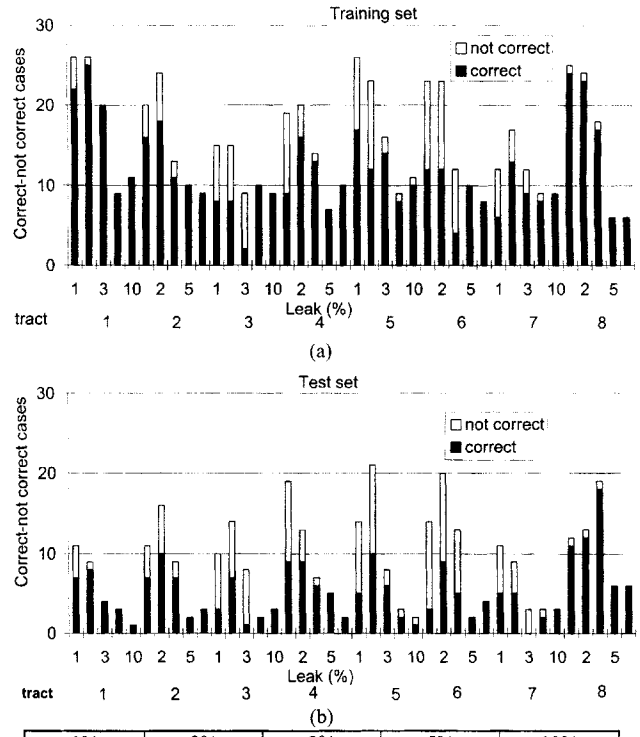
Actually, extensive field data were available only at this point in the research, and they were used to get direct information on field instrumentation behavior. The ANNs developed after this point were trained with data with "field" noise superimposed, so the database used was the most complete one, with the largest number of patterns. The examples are now more uniformly distributed to cover the possible situations in terms of leak sizes and locations. The results are reported in Figure 18. Even if large improvements in the results have not been obtained, the errors are now more uniformly distributed over the various classes of the output. An interesting result for this ANN is given in Figure 19, which shows the results when a segment is identified if the leak is located in the correct segment or in one of the adjacent ones.



1%	2%	3%	5%	10%
90/135 (67%)	123/145 (85%)	75/95 (79%)	72/77 (94%)	87/92 (95%)

Figure 17. Leak location using signals with noise: (a) training set; (b) test set.

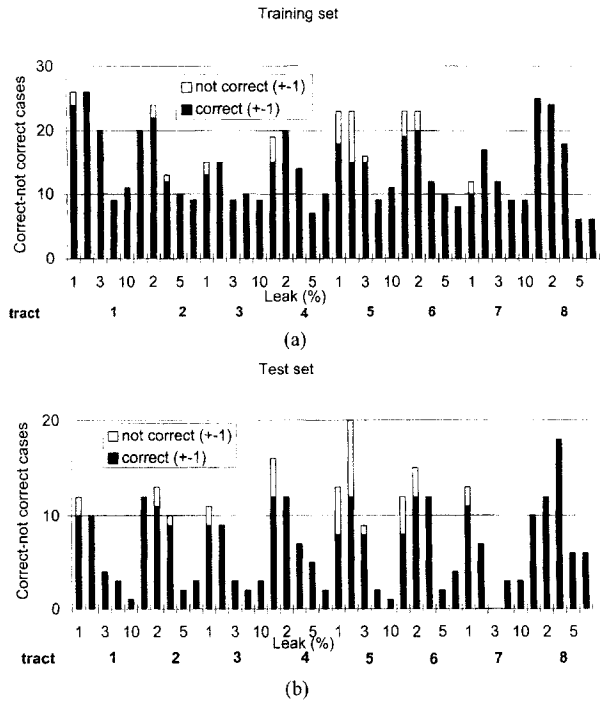
The annexed table summarizes the number of test-set correct patterns for the various leak sizes.



1%	2%	3%	5%	10%
42/75 (56%)	61/73 (84%)	40/56 (71%)	27/29 (93%)	31/32 (97%)

Figure 18. Leak location using signals with noise: (a) training set; (b) test set.

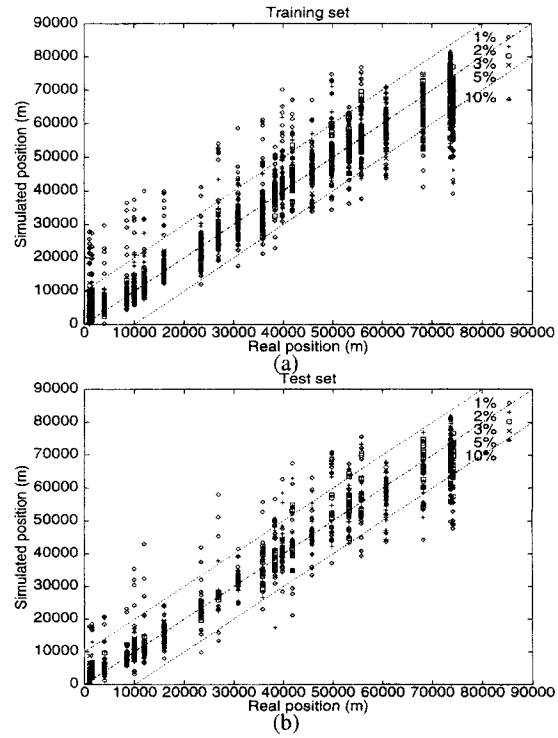
The annexed table summarizes the number of test-set correct patterns for the various leak sizes. Final database.



1%	2%	3%	5%	10%
62/75 (83%)	69/73 (95%)	53/56 (95%)	28/29(97%)	32/32(100%)

Figure 19. Leak location using signals with noise assuming the data predicted in the exact tract or in one of the adjacent segments are correct: (a) training set; (b) test set.

The annexed table summarizes the number of test-set correct patterns for the various leak sizes. Final database with field data included.



Leak size (% of inlet flow)	1	2	3	5	10
Average error (m)	8020	5134	3438	2227	1727
Standard deviation (m)	6973	4923	3460	1950	1645
95th percentile (m)	21204	17002	11133	6207	4685

Figure 20. Leak location along the pipeline: (a) training set; (b) test set.

The annexed table reports performance of ANN.

As can be seen, when the leak location is not correctly classified, it is usually located in an adjacent segment.

Some further analyses of noisy signals have been carried out. From these analyses it appears that the effect of noise on the results is not substantial, in the sense that although the noise somewhat degrades the response of the ANN, inputs with a doubled noise level have a similar performance overall. Moreover, the results are quite insensitive to the effect of the number of points used for calculating the moving averages.

Other attempts have been made to refine leak-location prediction. To this end, an ANN has been set up with the same input and hidden architecture and scaling layers as in the ANN previously developed for leak location in the tract. Only one neuron was used in the output layer, for predicting the leak-location coordinate along the pipeline. The results for training and test sets are reported in Figure 20. The figure reports the bands identifying an error range of ± 10 km. In addition, some statistics on the predictions given by the leak-locating network, such as the average, the standard deviation, and the 95th percentile of the error, are reported in Table 4 for different leak sizes. The range of the data is fairly large for the smallest leaks, although it decreases for bigger leaks. For the 1% leaks, the 95% of the predicted locations have an error less than 21.2 km. The performance clearly improves for larger leaks.

Leak-locating Test. The leak-locating capabilities of the ANN have been tested using the data used before as input. The data refer to a leak of different sizes (1%, 2%, 5%, 10%) located at 46 km from the start of the pipeline. The location of the leak corresponds to tract number 6, referring to the subdivision of the pipeline done for leak location.

The results of the system are reported in Figure 21. The segment where the leak is located is reported vs. time. After a transient that is also due to the fact that the system uses 21 points for the moving average (315 s with the sampling time of 15 s used), quite good location is obtained. The location is given exactly for a 5% leak (a similar result was obtained for a 10% leak), while a neighboring segment was indicated in the case of a 1% leak (similar results were obtained for a 2% leak). Some oscillations are present in this case. This confirms the earlier result: when a misclassification in leak location along the pipeline is made, the leak is usually found to be located in a tract adjacent to the one that was identified.

Table 4. Prediction Statistics on the Leak-Locating ANN

Leak size (% of inlet flow)	1	2	3	5	10
Average error (m)	8,020	5,134	3,438	2,227	1,727
Std. deviation (m)	6,973	4,923	3,460	1,950	1,645
95th percentile (m)	21,204	17,002	11,133	6,207	4,685

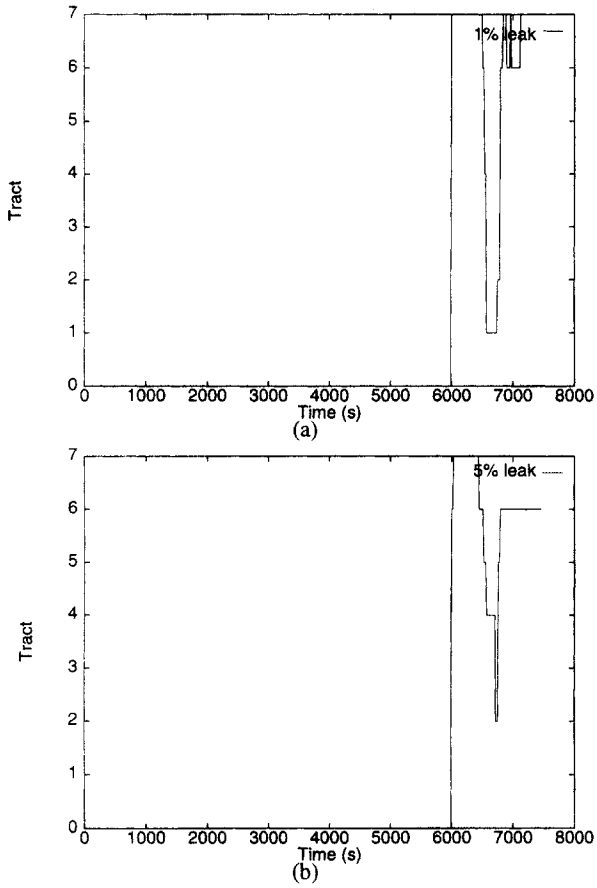


Figure 21. Response of the leak-locating ANN to field signals with effect of leak simulated: (a) 1% leak; (b) 5% leak.

Multiple Leak Detection. The behavior of the system that was set up was tested for the effect of multiple leaks. The results obtained when occurrence of two leaks in the pipeline is simulated (with the superposition of no-leak field data and the calculated time histories during leak occurrence, as described in the section on the leak sizing test) are shown in Figure 22. Two tests of consecutive leak openings were performed, the first one with a 0.5% leak developing at $t = 6000$ s in the seventh segment, followed, at $t = 7000$ s, by another 0.5% leak that opens in the fifth segment. In a second test, the exact sequence of events was repeated, with two 1% leaks opening one after the other.

From what is in Figure 22, the leak sizing and locating systems behave consistently with the performances shown in the previous sections, when just one leak was found in the pipeline. In fact, in the first test, a 0.5% leak did not cause an alarm, as the output signal is always smaller than the threshold chosen for the leak-detection system, but after the second leak opens, an alarm was indeed recorded, as expected. Consequently, leak locating was performed after the second leak occurred, and the ANN system actually predicted one leak in the sixth segment (see Figure 22b). In a second multiple-leak test, the system correctly identifies and positions a leak immediately after the first 1% leak opened (at $t = 6000$ s). When the second 1% leak opened, the overall performance was still adequate.

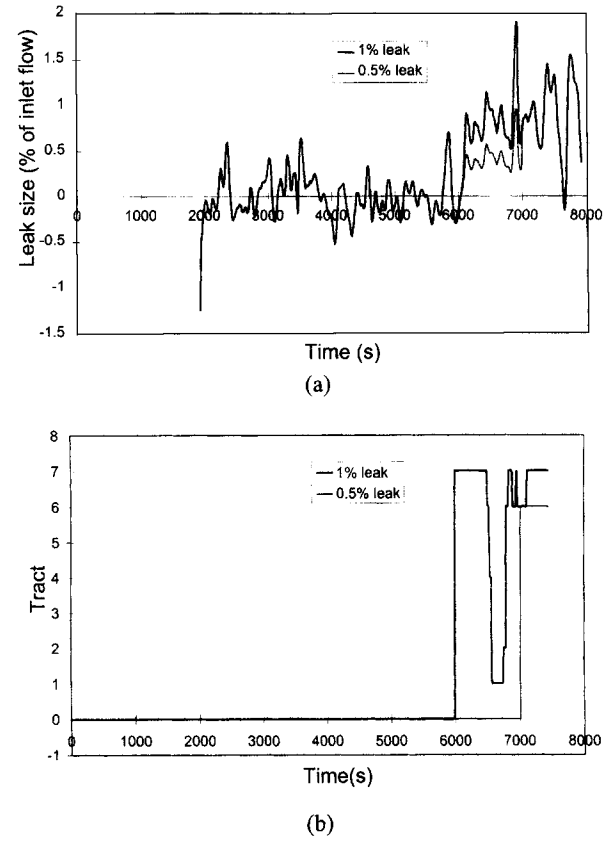


Figure 22. Simulation of multiple leaks: (a) leak sizing; (b) leak locating.

Flow Sheet of the Global Detection/Location System

The general flow sheet of the system that was developed and tested is shown in Figure 23. The outlet flow rate at the actual time, measured Dt seconds before, the inlet flow rate measured at $(t - Dt)$, and the actual pressure signals are fed to the leak-sizing network, which also receives the outlet flow rate calculated by the code. The measured inlet flow rate and outlet pressure are used as boundary conditions to the code to provide the calculated outlet flow rate, and hence compensate for the compressibility effect. If a leak is detected, the leak-sizing network provides the leak size to the leak-location system, which also receives the signals of flow rate and pressure measured at time t and before leak occurrence. In this way, the pressure drop due to the leak can be calculated and used for leak location.

Conclusions

An ANN-based system for leak detection and location has been developed. The main database was generated by using a well-validated computer code, the effect of noise being taken into account, by starting from the transducer characteristics and then using available field data.

The leak-detection neural networks were trained and successfully tested with noisy signals. The system was able to detect leaks as small as 1% of the flow rate without generat-

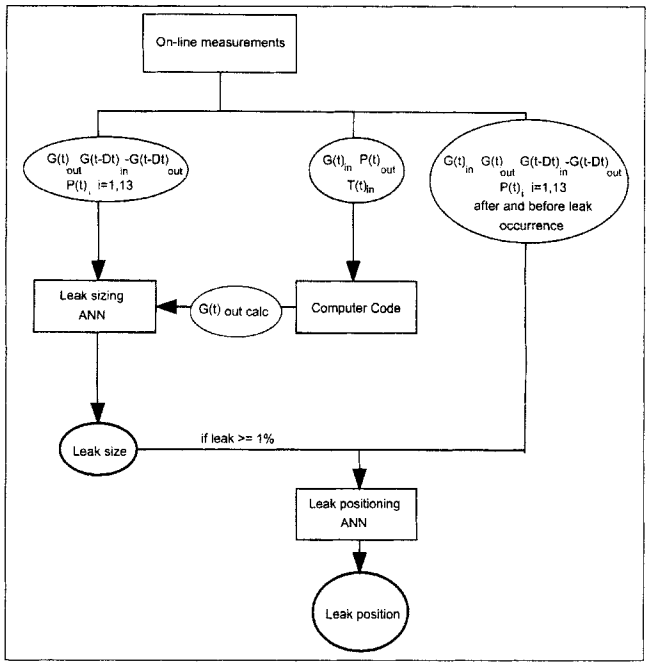


Figure 23. Logic diagram of the complete system for leak sizing and locating.

ing spurious alarms. It was found that it is necessary to compensate for the effect due to fluid compressibility to avoid spurious alarms as a consequence of a simple transient flow condition in the line. The coupling between the code and the ANN allows data preprocessing and compensation of the compressibility effect due to variations in the boundary conditions in the system.

The effect of the leak was added to field-measured signals (without leaks) in order to test the system's capabilities. The resulting signal was fed to the leak-detection system. The performance of the system was very good, as it was able to detect even the 1% leak in transient conditions.

The ANNs performed very well in leak location when noise was not present. The location of large (5%, 10%) leaks was predicted very accurately even when noisy signals were used. This is particularly interesting because fast location is very important for such leaks. The locations of about 50% of 1% leaks were correct. The percentage of correct analysis increased to about 75% with regard to location of 2% leaks. Usually the misclassification resulted in the leak being found in an adjacent segment of the line.

Acknowledgments

The activity described was performed in the framework of the Research and Technological Development Project "Detection and mitigation of risks associated with pipeline releases" (DEPIRE) sponsored by the European Commission under the ENVIRONMENT Programme (Contract EV5V-CT94-0419). The international consortium was composed by Snamprogetti (Italy), TNO (The Netherlands), AEA Technology (United Kingdom), and Consorzio Pisa Ricerche (Italy). Close collaboration from all partners is gratefully acknowledged.

Literature Cited

Belsito, S., "Development of a Model for the Simulation of Leaks in Pipelines," EC Project DEPIRE Report, Pisa, Italy (1995).

Belsito, S., and G. Campastro, "Assessment of the Capabilities of the ANN-Based Leak Detection System," EC Project DEPIRE Report, Pisa, Italy (1997).

Billmann, L., and R. Isermann, "Leak Detection Methods for Pipelines," *Automatica*, **23**, 3 (1987).

Bishop, C. M., "Neural Networks and Their Applications," *Rev. Sci. Instrum.*, **65**, 6 (1994).

Ellul, I., "Pipeline Leak Detection," *Chem. Eng.*, N461 (1989).

Ham, J. M., and F. Misuraca, "Basic Project Data," EC Project DEPIRE DGXII Meeting, Apeldoorn, The Netherlands (1995).

Hamande, A., V. Condacse, and J. Modisette, "New System Pinpoints Leaks in Ethylene Pipeline," *Pipeline Gas J.* (1995).

Haykin, S., *Neural Networks, A Comprehensive Foundation*, Maxwell Macmillan, London (1995).

Parry, B., R. Mactaggart, and C. Toerper, "Compensated Volume Balance Leak Detection on a Batched LPG Pipeline," *1992 OMAE*, Vol. V-B, *Pipeline Technology*, ASME, New York (1992).

Sandberg, C., J. Holmes, K. McCoy, and H. Koppitsch, "The Application of a Continuous Leak Detection System to Pipelines and Associated Equipment," *IEEE Trans. Ind. Appl.*, **IA-25**, 5 (1989).

Siebert, H., "A Simple Method for Detecting and Locating Small Leaks in Gas Pipelines," *Process Automation* (1981).

Silk, M. G., and P. Carter, "A Review of Means of Pipeline Leak Detection Independent of Flow Measurement," EC Project DEPIRE DGXII Meeting, Harwell, UK (1995).

Sintef-Ife, *OLGAs89, Model Numerics and Users Guide*, Internal Publication (1990).

Sperl, J. L., "System Pinpoints Leaks on Point Arguello Offshore Line," *Oil Gas J.*, **89**, N36 (1991).

Stouffs, P., and M. Giot, "Pipeline Leak Detection Based on Mass Balance: Importance of the Packing Term," *J. Loss Prev. Process Ind.*, **6**, 5 (1993).

Uguccioni, G., EC Project DEPIRE DGXII Meeting, Pisa, Italy (1996).

Wang, G., D. Dong, and C. Fang, "Leak Detection for Transport Pipelines Based on Autoregressive Modeling," *IEEE Trans. Instrum. Meas.*, **IM-42**, 1 (1993).

Zhang, X. J., "Statistical Method for Detection and Localization of Leaks in Pipelines," *1992 OMAE*, Vol. V-B, *Pipeline Technology*, ASME, New York (1992).

Manuscript received Aug. 4, 1997, and revision received Aug. 19, 1998.