An inverse approach for piping networks monitoring

Antonio C. Caputo, Pacifico M. Pelagagge *

University of L’Aquila, Faculty of Engineering, Monteluco, 67040 L’Aquila, Italy

Abstract

Spills and leakages of hazardous fluids from piping networks may pose a significant safety risk to population, industrial plants and the environment. Therefore in fluid distribution the problem of monitoring the network status in order to identify abnormal conditions and locate leakages arises. In the paper an inverse approach resorting to a multi-layer perceptron back-propagation Artificial Neural Network (ANN) is proposed, in order to locate leakages based on pressure and flow rate information. Strategies for generating input data and for correlating by ANN such data to the fluid distribution system status are presented. A two-level architecture is selected, composed by a main ANN at the first level and several branch-specific second-level ANNs in cascade to the main one. The branch in which the leakage occurs is identified, resorting to the ANN operating at the first level, while the specific second-level ANN is activated to estimate accurately the magnitude and location of the leakage in the selected branch. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Artificial neural networks; Safety; Hazardous substance transportation; Fluid distribution piping; Pipeline systems; Leakage monitoring

1. Introduction

Piping networks have found wide diffusion in industrial and civil applications where fluids have to be distributed to several consumers or processes. In order to carry out such a task in an effective and safe manner, the occurrence of unauthorized withdrawal, spills and leakages should be identified and promptly eliminated, especially when flammable or toxic liquids are involved which may cause accidents having relevant consequences.

Many causes of piping accidents have been classified by the industry with reference to the oil and gas sector (Papadakis, 1999; ECIG, 1993; US Dept. of Transportation, 1991; CONCAWE, 1996, 1977) with the main causes being: external interference or third party activities (i.e. the use of machinery), corrosion, construction defect and mechanical or material failure, ground movements and natural hazards in general, and operational errors. Currently there are about 31,000 km of onshore pipelines throughout the EU conveying crude oil or refinery products, while approximately 10,000 km of chemical pipelines within the EU are used for carriage of such hazardous substances as ethylene, propylene, butadiene, chlorine, styrene and ammonia (Papadakis, Porter & Wettig, 1999). In Europe the high pressure gas transmission grid totals, instead, a length of over 180,000 km to which must be added over one million kilometres of low pressure distribution lines. Gas transmission lines are buried underground and necessarily routed through cities and densely populated areas. Montiel, Vilchez, Arnaldos and Casal (1996) report that 70.8% of overall accidents which occurred with natural gas utilization took place in piping systems.

Even if it is widely recognized that pipelines are one of the safest modes of transporting large volumes of hazardous products, the failure of pipelines conveying dangerous substances can pose major risks. Release of flammable and toxic materials can be the initiating event of accidents with catastrophic effects. Fluids distribution systems may therefore be considered a source of industrial major accident hazards.

In recognition of this issue the European Union has recently started an initiative on the control of the major accidents hazards from pipelines given that hazardous fluids piping networks and pipelines are not subjected to risk assessment or safety studies by national regulations in most European Countries (Papadakis, 1999; Papadakis, Porter & Wettig, 1999). However, while measures...
Nomenclature

D diameter;
f Friction factor;
k Loss coefficient;
L Length;
p Pressure;
q Flow rate;
Re Reynolds number;
v Velocity;

Subscripts

i i-th node;
j j-th node;
F Fanning;

Superscripts

in Inlet;
ex Outlet;

Symbols

ε Roughness;
Δp Pressure loss;
ρ Density.

to reduce risk may be better implemented in newly commissioned pipelines (increased burial depths, lower operating pressures, increased pipe thickness etc.), leak detection approaches become more relevant for aged pipelines, especially those with a record of past accidents.

Therefore, the problem arises of monitoring the network status, in order to identify abnormal conditions and locate leakages. In the literature, despite the significant efforts made, an effective solution to the networks monitoring problem has not been found and it still presents a significant challenge.

Current technologies for fault detection and identification (FDI) are based on methods which are both dependent or not-dependent on process parameters. Not-dependent technologies include:

- injection of tracing substances in the fluid stream;
- analysis of acoustic emissions from leaks;
- pressure waves transmission;
- visual inspection;
- detection of leak-generated temperature variations resorting to infrared thermography;
- identification of radar or radio frequencies emitted from transmitters located inside the pipes and permeating outside through pipe cracks.

At the moment, however, such classes of techniques are expensive, often intrusive and perturbing the whole fluid network, or difficult to monitor by telemetry.

Methods dependent on process parameters, instead, are based on establishing mass and energy balances at different network nodes using measured values of flow rates (q) and pressure levels (p).

However, it is well known that computation of the network boundary conditions starting from measured q and p values at the moment presents mathematical difficulties that are far from being solved. On the contrary, evaluation of p and q at each node of the network is straightforward when boundary conditions are known. Therefore the FDI in piping networks based on process parameters results in a classical inverse engineering problem, that is determination of causes (leaks) starting from knowledge of their effects (p–q maps) (Ozisik & Orlande, 2000).

Control of fluid distribution networks requires moni-
toring of the current state of the system. Although telemetry systems represent a promising solution for evaluating state variables in the network (pressure and flowrates), their widespread diffusion is impaired by the high costs. State estimation techniques are therefore utilized to ascertain system status from limited information. However, both metering and telemetry systems are subject to errors having a negative impact on the accuracy of the state estimation calculation. Results may therefore be very inaccurate when compared to the actual system state. Moreover, the analytical techniques adopted are rather complex and computationally intensive. Therefore, a robust, efficient, and low-time-consuming means to evaluate the entire fluid network status, starting from a limited amount of state knowledge, is required.

In this paper, a recursive mathematical model of a generic network, enabling the direct computation of the system status in any condition, has been presented first. The inverse problem is successively stated, discussing difficulties involved in its analytical solution, also presenting the techniques available to correlate state variables during perturbed operations of the network. A solution approach, based on the utilization of an ANN, is then presented, examining possible ANN architectures and selecting a multi-layer perceptron back-propagation network. The strategy for coding of input and output data, with reference to the real system and the utilization of this method, are discussed. A two-level architecture is selected composed by a main ANN at the first level, and several branch-specific second-level ANN, in cascade to the main one. The branch in which the leakage occurs is identified, resorting to the global ANN operating at the first level, while the specific second-level ANN is activated to estimate precisely the magnitude and location of the leakage in the selected branch.

2. Literature review

A considerable amount of research effort has focused on the identification of leaks in fluid distribution networks, addressing either the compressible or incompressible fluid flow cases. As already mentioned, available techniques rely either on process-dependent or process-independent variables. Silk and Carter (1996) presented an in-depth review of available techniques related to process-independent methods.

As far as process-dependent methods are concerned, the most common is the line volume balance technique, utilizing data supplied from flow rate meters located at the extremity of each branch. Usually such approaches prevent the exact pinpointing of the leak location. Ellul (1989) proposed a method in which measured pressure and flow rate data are compared with those computed on the basis of a mathematical model of the system, and the leak amount is made proportional to the obtained difference. Leaks lower than 5% may be discovered with processing times in the order of minutes. Stouffs and Giot (1993) present methods based on mass balances able to treat transient states of compressible fluids highlighting how the practical detectable limit is 2–3% for compressible fluids. Billman and Isermann (1987) propose an adaptive non-linear observation method of the pipeline dynamic behaviour, along with a special correlation technique, based on the measurement of pressure and flow rate data at the inlet and exit of a branch. Leaks of around 2% for liquids and 10% for gases have been detected. Siebert (1981) developed a method based on the statistical analysis (cross-correlation) of signals sampled over a time interval of 1.7 s, being able to locate leaks in the order of 0.2% for liquids, while for gases the value of 5% was reached. Also Zhang (1992) adopted statistical signal analysis techniques to address the problem, detecting 1% leaks on a 37 km pipeline. Whang, Dong and Fang (1993) developed a method able to identify leaks of 0.5% in 120 m long water pipes in negligible time delay utilizing an autoregressive modeling approach requiring only four pressure measurements with a sampling time of 20 ms. Hamande, Condace and Modissette (1995) tested a leak detection system on an actual ethylene pipeline operated since 1989. The system relied on 21 pressure measurement stations and two flowrate meters associated with a mathematical model of the pipeline, enabling detection of 7% leaks in a 60 min time span. Parry, MacTaggart and Toerper (1992) describe another system adopted for LPG pipelines interfaced with a measurement instrumentation enabling detection of 2% leaks over times ranging from 46 minutes to 9 h depending on the distance of the leak from the nearest measuring station. Another leak detection and location system for pipelines was developed by using artificial neural networks (Belsito, Lombardi, Andreussi & Barerje, 1998). This system can detect and locate leaks down to 1% of flow rate with a probability of success that is greater than 50% for the smallest leak. In the paper, specific signal processing techniques able to reject spurious alarms due to compressibility effects subsequent to operational transients are developed. However, this method too involves leaks identification in single-branch pipelines only. Up to now, detection of leaks lower or equal to 1% requires monitoring times ranging from about 100 to 1000 s according to the technique utilized.

Generally speaking, besides such approaches, usually aimed at long pipelines or single branches of a network, in case of inverse problem solution, calculation of all pressures or flowrates in a fluid distribution system may be accomplished by formulating and solving the mass and/or energy conservation equations. Such balances have to be consistent with the measurements carried out in the system provided that there are as many independent equations as there are variables that are to be
In this paper, instead, a simplified approach to leak detection and location is suggested using ANN as global classifiers, among the possible pressure/flowrate patterns, in order to correlate directly state variables’ values to network states.

3. The proposed methodology

The methodology is based on two main phases (Fig. 1):

- evaluation of pressure/flowrate conditions (effects) by simulation imposing the piping network’s boundary conditions (causes);
- correlation of effects with causes by Neural Networks.

In order to carry out the training step of the ANN, the direct problem is solved at first resorting to the mathematical model of the network. In particular, variables at the nodes are computed in the case of either normal operation and when artificial leakages are introduced, enabling us to determine a set of patterns characterizing the system status in several conditions of normal or abnormal operation. Normal operating data could, in fact, be obtained on the field but only for existing networks—moreover this is a time consuming task. Furthermore no data are readily available for atypical events (i.e. leakages) or operation in non standard situations, while it is not feasible to obtain such data from physical simulation on the real network. Therefore a mathematical model of the network has been utilized to predict the status of the network consequent to normal and abnormal flow situations, and to generate learning data sets. Successively the ANN is trained on these data sets, representing the system status (i.e. pressures and flow rates) in different operating conditions either with or without leakages. In this way a robust correlation between leak effects (p–q maps) and location-amount of leaks (causes) may be defined. However, the ANN should be able to reject the noise represented by the inevitable measurement errors affecting the sensors. In order to increase robustness of the ANN, the input data sets may therefore, be altered in respect of the accurate values exactly representing the system status, by superimposing random deviation to the pressure drop and the

![Diagram](image-url)
leak flow readings, comprised in the instruments’ uncertainty range. During the running phase the ANN is fed with actual measurements of pressure and flow rate from a limited number of nodes, and performs clustering and classification tasks, giving in output the location (branch) and magnitude of the leak.

In more detail the following main steps have to be performed in the proposed methodology:

1. development of a mathematical model of the network for p and q evaluation starting from the actual boundary conditions;
2. simulation of leakages occurring and computation of resulting p and q values in the network (data generation);
3. correlations of leaks patterns to p and q data resorting to ANN.
4. integration of the ANN system in the piping network in order to carry out real-time and on-line monitoring and leak identification tasks based on actual fluid network measured operational status.

The entire process is schematized in Fig. 2.

Implementation of the ANN-based pattern classification system in particular involves:

- ANN architecture definition;
- selection of data structure and ANN input and output layer;
- ANN training;
- ANN testing;
- running phase.

4. Estimation of the network status

Several well proven methods for the solution of a direct fluid distribution network problem are available in the literature (Eggener & Polkowski, 1976; Roy, 1988).

Pressure drop due to friction losses of the flow rate \( q_{ij} \) across a generic branch connecting nodes \( i \) and \( j \) may be expressed as

\[
\Delta p_{ij} = k_{ij}(q_{ij})^2
\]

where \( k_{ij} \) is the loss coefficient

\[
k_{ij} = \frac{32 f_F \rho L_{ij}}{\pi^2 D_{ij}^5}
\]

The Fanning friction factor \( f_F \), a function of flow regime and wall roughness, has been expressed by the following empirical correlation:

\[
f_F = \frac{1}{16 \left( \log \frac{\varepsilon}{3.7 D} - \frac{5.2 \log \left( \frac{\varepsilon}{3.7 D} + \frac{14.5}{\text{Re}} \right)}{\text{Re}} \right)^2}
\]

It follows that the branch flow rate is

\[
q_{ij} = \frac{p_i - p_j}{k_{ij} |q_{ij}|} = \frac{p_i - p_i}{\sqrt{k_{ij}} |p_i - p_i|}
\]

However, the mass balance asks that

\[
\sum_i (q_{ij} + q_{in} + q_{ex}) = 0 \quad \forall j
\]

meaning that both the mass balances at the nodes

\[
\sum_i q_{ij} = 0 \quad \forall j
\]

and the overall external flow rate balances

\[
\sum_i (q_{in} + q_{ex}) = 0
\]

have to be satisfied. Therefore Eq. (5) may be rewritten as

\[
\sum_i \left( \frac{p_i}{\sqrt{k_{ij}} |p_i - p_j|} + q_{in} - q_{ex} \right) = p_i \sum_i \frac{1}{\sqrt{k_{ij}} |p_i - p_j|}
\]

Indicating by \( \Sigma I \) the summatory at left hand member and with \( \Sigma II \) the summatory at right hand member one gets the expression for the pressure at the \( j-th \) node as

\[
p_j = \frac{\Sigma I}{\Sigma II}
\]

The algorithm is as follows:

**Initialization phase**

- input of tentative inlet flowrate values at each network
nodes \( q_i^{in} = (0,1) \), the value is set to 1 to indicate that the node receives an input external flowrate, otherwise it is set to 0;

- input of actual exit flowrate values \( q_i^{out} \);
- input of tentative values of pressure \( p_i \) at nodes. Pressure at the network input nodes are considered to be known and to remain constant, while pressures at any other node is the simulation output;
- definition of network topology. This is carried out resorting to the pressure loss coefficient \( k_{ij} \); it is initially assigned value 1 if nodes \( i \) and \( j \) are connected by a branch, or value 0 otherwise;
- definition of length \( L_{ij} \) and diameter \( D_{ij} \) for each network branch.

**Network simulation phase**

1. branch flow rates \( q_{ij} \) are computed by Eq. (4);
2. the average fluid velocity in each branch \( v_{ij} = q_{ij}/\pi D_{ij}^2/4 \) is computed as well as the Reynolds number;
3. the Fanning friction factor is computed (Eq. (3));
4. values of loss factor \( k_{ij} \) are updated through Eq. (2);
5. pressures at the nodes \( p_i \) are updated through Eq. (9);
6. flowrates for the nodes receiving an external flow rate (those where the initial value was \( q_i^{in} = 1 \) are also updated through Eq. (9) by maintaining constant the assigned \( p_i \) value and solving Eq. (9) for \( q_i^{in} \);
7. flow rates \( q_{ij} \) for the other generic branches are updated by Eq. (4).

The entire procedure is iterated until convergence of computed flow rates \( q_{ij} \) is obtained.

The algorithm shows fast convergence and enables an easy configuration of the network topology during the data initialization phase. In fact it is not required to define a different set of equations to be solved according to the specific network analysed. Program outputs are pressure values at each node and flow rate values at each branch. This algorithm was compared with other traditional solution approaches (the Hardy–Cross iterative method, and the solution of a set of equations describing conservation of flow rate at the nodes and of the algebraic sum of pressure loss across any pipe loop (Roy, 1988)) obtaining very good agreement as shown in Table 1 for a simple 5-nodes network.

The simulation of a leaking network may be carried out by simply inserting in the desired branch, fictious nodes from where the leakage flowrate spills. The user may specify the leak amount, its location and the interested branch. The program then automatically rearranges the network configuration, consequent to the increase in the number of nodes and branches, in respect of the original network, and computes the new values for pressures and flow rates. Such values may therefore represent the readings supplied under degraded network operation, by measuring instruments distributed on the networks that will be analysed to locate the leaks. According to the control method hypothesized the input flow rate will increase in order to compensate for the leaking flowrate, by maintaining the fixed pressure at the input nodes. Besides generating training data for the ANN, the program is a useful tool to explore the amount of pressure perturbation consequent to leakages, and to assess the required sensitivity of the measuring instruments, so that the flow anomaly can be positively detected. The effect of signal noise may also be evaluated. On the contrary, given a set of measuring instruments having a specified sensitivity, the minimum leak amount able to be detected may be computed. Optimal placement of sensors along the network may also be studied. Such issues are of great importance for the practical utilization of this method. In fact, the quality of available pressure sensors determines the accuracy that can be obtained in determining the location of the leak along the pipe.

5. Utilization of ANN for FDI

In order to solve the leak detection and identification problem in a generic fluid distribution network, the use of ANNs is proposed by exploiting their classification abilities. The main goal is to recognize patterns among the pressure and flowrate data supplied by a set of measuring instruments distributed across the piping network, and as a consequence, select the branch where a leak occurs, location along the branch and amount (Fig. 3).

ANN are basically a non linear transfer function applied to a set of input variables. Physically the ANN is composed of a set of elementary computing units, the neurons, connected each other according to the peculiar network structure adopted. Each neuron’s output is the value of the non-linear activation function computed on the weighted sum of the input signals to the neuron itself. Assigned to the neuron’s connections are weights which are generally chosen during the network training phase in order to minimize the error between the output values computed by the ANN, and the true values corresponding to the set of input data processed by the ANN. The error function may be considered as a surface in the space of weights leading to the utilization of gradient-based minimization techniques. Computational performances may be enhanced if multiple neuron layers are introduced between the input and output layer. ANN have proven successful in approximating nonlinear multivariable functions, and in classification tasks. Provided that a sufficiently large training data set is available, ANN may be useful when it is difficult or impossible to obtain a mathematical model of the system to be solved by analytical methods, when new data have to be processed at high speed in real time, and when the computation method needs to be robust and fairly insen-
Table 1
Comparison of computation methods

<table>
<thead>
<tr>
<th>Branch</th>
<th>Hardly-Cross</th>
<th>Mass/energy balance</th>
<th>This work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–2</td>
<td>47.30</td>
<td>47.14</td>
<td>47.17</td>
</tr>
<tr>
<td>1–3</td>
<td>52.70</td>
<td>52.85</td>
<td>52.83</td>
</tr>
<tr>
<td>2–3</td>
<td>10.10</td>
<td>10.08</td>
<td>10.01</td>
</tr>
<tr>
<td>2–4</td>
<td>57.50</td>
<td>57.22</td>
<td>57.18</td>
</tr>
<tr>
<td>3–5</td>
<td>42.70</td>
<td>42.77</td>
<td>42.81</td>
</tr>
<tr>
<td>4–5</td>
<td>7.50</td>
<td>7.22</td>
<td>7.19</td>
</tr>
</tbody>
</table>

Fig. 3. Leak identification through ANN.

5.1. Selection of neural network type

Characteristics of the aforementioned ANNs will be discussed briefly in the following in order to make some consideration about the type of network which is more likely to represent the best candidate for this application. However, the final choice will depend on extensive experimentation, and the characteristics of the actual fluid network being studied.

A PNN is basically a Bayesian classifier into neural network architecture. This kind of ANN does not update the internal weights during the training phase. Rather, it stores the training sets and compares them with the actual input during the working phase, in order to select the weight set which corresponds to the closest training set in respect of the actual data input (Nearest Neighbor method) and compute the output accordingly. PNN output, in this case, is simply the class of the leak, i.e. a code identifying one of the original training set that represents a known and specific status of the network (characterized by a given leak amount and location).

PNN generally result in the best classifier when inputs very similar to those utilized for the training phase are utilized. Adding a noise to the input data in respect of the training data, may result in some input cases not correctly classified. In case of incorrect classification the error is unlikely to be on the leak amount or position on a branch, but on the branch itself, making the computation output useless for the purposes of the study. It follows that the main drawback of the PNN lies in the fact that no estimation of the ANN output uncertainty can be made. Therefore even if the PNN shows perfect sensitivity to noise in the input data. Topologies of ANN which show best classification abilities are the Probabilistic Neural Network (PNN), the Radial Basis Function (RBF) network and the Multi Layer Perceptron (MLP).
correspondence between known inputs and corresponding classes it is rather seriously affected by inevitable disturbances in the sensor reading, making it quite unreliable in this application. In fact a leak or an alteration of sensor signals generating a p/q pattern much different from those the ANN has been trained on, or very similar to the pattern generated by an entirely different leak condition, may completely deceive the ANN which could indicate in output a leak class entirely different from the real one. The output layer may consist simply of a neuron indicating a number directly corresponding to one of the classes of training data. However, the required output coding (consisting of the association of the input to a class among those utilized for training) prevents ready recognition of an erroneous result.

RBF networks are made up by two external layers, with the hidden layer composed of neurons having a gaussian activation function, while the output layer is composed of simple neurons carrying out a linear combination of the output of hidden neurons. This type of ANN showed interesting performances during preliminary trials, including significant robustness to noise, when applied to a single branch. However, the required uncertainty in leak location remained the same, but the flow rate noise was nearly directly transferred to the new leak estimation value, making the RBF network even more robust than the MLP. However, it was highly unreliable when applied to a fluid distribution network. The ANN could not classify inputs altered by a noise while, also in absence of noise, it was not possible to minimize the ANN errors, which involved indistinctly any component of the output vector.

MLP are multilayer networks where the neuron activation function is sigmoidal. The MLP also shows good robustness with respect to the alteration of input values. During the performed test MLP correctly identified the branch where the leak occurred in all the examined cases. Therefore the MLP may represent the best compromise among good previsional capability and good noise rejection.

This latter type of ANN resulted to offer the best performance compromise and has been adopted in the following of the study.

5.2. Information coding

Different input–output coding schemes and number of neurons in the external layers may be utilized when MLP are utilized to correlate state variables in fluid networks. In particular, inputs may be the absolute or relative pressure loss in respect of the nominal value (better if normalized in the 0–1 range) at the nodes where such measurements are available, as well as the boundary conditions (flow rate measurement at the inlet nodes and head at the fixed heads nodes).

Outputs may be coded as binary 0–1 to identify the leaking branch (if an output neuron is associated with each branch) or as values comprised in the 0–1 range if the degree of leak severity (as a predefined percentage of nominal branch flow rate) and the leak location along a branch (as a predefined fraction of the branch length) need to be estimated.

The kind of output coding may be selected in order to suit specific requirements, but the preferable input coding is the one based on relative variations of the input variables in respect of the value at nominal operating conditions.

5.3. Overall neural network architecture

MLP is the better suited architecture in this kind of application, but a single MLP presents some output coding limitations that may prevent our obtaining good accuracy in real applications. On the other hand, given its proven function approximation capability, on a single branch application the MLP may guarantee a satisfying accuracy, together with a proper presentation of output data. Such consideration leads to the utilization of a hybrid network structure having two levels (Fig. 4).

A first-level MLP is utilized to locate the branch where the leak occurs, in order subsequently to activate specific second-level branch-dedicated MLPs having the task of estimating precisely the leak flow rate and location. In this way the complex task of identifying leaks in realistic fluid distribution networks may be subdivided into two sub-problems that may be fairly easily solved separately by dedicated MLP artificial neural networks. In the case of unstationary distribution networks, where periodic variations of the overall state are observed, the entire structure is repeated and possibly selected by another ANN as suggested by Gabrys & Barjela, 1999b).

![Fig. 4. Two-level neural network architecture.](image-url)
6. Conclusions

In this paper the feasibility of utilizing Artificial Neural Networks in locating leaks over a fluid distribution network, as a means of enhancing system safety, has been discussed also by analysing the relevant literature. A suitable overall two-level ANN architecture is also presented where a first-level ANN determines the branch where the leak occurs and a specific second-level ANN estimates leak amount and location. The proposed architecture has been satisfactorily tested in a simplified case, obtaining promising results. However, details of the implementation phase on a real test case and a discussion of performances will be carried out in a forthcoming paper. As a future research work, the complexity of problems which can be solved with this approach will be assessed and the robustness of the ANN to measurement errors of the sensors will be evaluated, in order to define the conditions of applicability of this method to actual fluid distribution networks. More specifically, some practical issues having economic relevance will be dealt with, such as the determination of the sensors’ required number and accuracy level which minimizes the total monitoring system cost, maintaining a required level of provisional capability.

References


