Leak Detection and Topology Identification in Pipelines Using Fluid Transients and Artificial Neural Networks

Jessica Bohorquez1; Bradley Alexander2; Angus R. Simpson, M.ASCE3; and Martin F. Lambert, M.ASCE4

Abstract: Condition assessment of water pipelines using fluid transient waves is a noninvasive technique that has been investigated for the past 25 years. Approaches to identify different anomalies and to identify elements of the topology of a pipeline have been proposed but often require detailed modeling and knowledge of the system. On the other hand, artificial neural networks (ANN) have become a useful tool in a range of different fields by enabling a computer to solve a problem without being explicitly programmed to do so, but rather by learning from a series of known examples. This paper presents a new methodology that uses ANNs to predict the presence of features in a pipeline. First, the location and characteristics of a junction have been predicted as a way to identify elements of the topology of a pipeline followed by identification of the location and sizing of a leak. The ANN characteristics and training approaches have been determined for both the junction and the leak example. Results show that the ANN that has been designed for this research is able to accurately predict the location of a junction with an error in this estimation of 2.32 m (out of a 1,000 m long pipeline) or less in 95% of the tested cases. The prediction of the two different diameters on either side of the junction was extremely accurate with only one misidentification of one of the diameters in the 5,000 tested examples. When the ANN was trained and tested to locate and size a leak, the results were also successful. A total of 95% of the tested examples located the leak with an error equal or less than 3.0 m (out of a 1,000 m pipe length) and the leak size was predicted with an average absolute error of only 0.31 mm. The results presented in this paper demonstrate the potential of combining the use of both fluid transient pressure waves and ANNs for the detection of features in pipelines. DOI: 10.1061/(ASCE)WR.1943-5452.0001187. This work is made available under the terms of the Creative Commons Attribution 4.0 International license, http://creativecommons.org/licenses/by/4.0/.

Author keywords: Junction identification; Leak detection; Water pipelines; Artificial neural networks (ANNs); Machine learning; Water distribution systems.

Introduction

Water is a vital resource to society and its continuous distribution for all types of use is becoming a challenge, including time and spatial inequality in its availability, and the lack of responsible management and poorly maintained infrastructure. In addition, pipes in water distribution system networks are often underground, which complicates their monitoring and maintenance. To overcome this, various noninvasive techniques have been developed to monitor and detect faults in pipelines for accomplishing an efficient and cost-effective maintenance. These methods include visual inspection (Guo et al. 2009), electromagnetic methods (Wang et al. 2012), acoustic methods (Juliano et al. 2013), ultrasonic, radiographic, thermographic methods (Zheng and Yehuda 2013), and more recently transient-based inspection (Lee et al. 2008; Gong et al. 2014, 2018).

Of these different techniques, transient-based methods have received special attention in the past two decades given that they allow the inspection of large sections of a pipe with a relatively simple set up (Gong et al. 2013b), and results can be obtained quickly (Lee et al. 2006; Shi et al. 2017). These methods are based on the interpretation of the effect that any feature in a pipeline has on the transient head trace, when a small controlled artificial transient pressure event is generated. To detect faults by using transient pressure signals, several methods have been proposed and can be organized in three groups according to Colombo et al. (2009): (1) inverse transient techniques, (2) frequency domain techniques, and (3) direct transient methods. However, whereas each of these approaches has been moderately successful, they also have associated disadvantages in terms of processing time or required knowledge of the analyzed system.

The current paper presents an innovative transient-based technique that uses artificial neural networks (ANNs) to identify topological elements such as junctions in water pipelines networks and to locate and characterize leaks. Unlike previous studies, the proposed technique is data-driven because it does not need any detailed information with regard to the analyzed water network system. When the ANN is tested, it gives successful results using just the transient head trace measured at the transient generation location. Previous techniques of transient-based pipe condition assessment have been model driven because they have always had to prepare a detailed simulation model of the pipeline to be analyzed. The technique proposed in this paper uses transient head data for the training of the ANN that can be obtained either from historical...
data or from an available model and is data-driven in its application because it only requires measured transient head data to locate and characterize the desired features. In addition, the computational effort of the proposed technique is concentrated in the ANN training stage, but once this stage is complete, the technique can process and find leaks and/or topological features from different transient head traces almost immediately, without the need of retrain the ANN. Considering that this is the first reported joint application of fluid transients and ANNs, the examples and the functioning of the technique are demonstrated only with numerically derived data. However, the promising results of this first application prove the promising potential of this approach and offer insights for future validations (with experimental and field data) and applications in more complex systems.

The methodology proposed in this paper is applied to the location and characterization of different features in a single pipeline. It is shown that the technique is successful in accurately predicting the location of the feature within the pipeline and its size, and it proves the potential of exploring the joint use of fluid transients and ANNs in more realistic scenarios. The ANNs were trained with numerical data generated using a method of characteristics (MOC) transient model and its performance has been tested also with different sets of numerical data showing that the ANN is able to identify new features that were not used in its training.

Two hydraulic systems have been considered. The first one is a pipeline with two segment lengths of different diameter (referred to as a junction system), and the second one was a single pipeline with the presence of a leak. The junction system was used to determine the most appropriate structure and characteristics for the ANN given that it is a simpler system and the response of the transient head trace is easier to identify in comparison to a leak in a pipeline. In addition, the identification of a junction was selected as an example of topology identification because its location and the diameters of the two segments are predicted.

The present paper provides a background in transient-based methods for identifying topological elements and for locating leaks based on transient pressure traces, highlighting certain limitations of each method that can potentially be overcome with the use of ANNs. A summary of the key elements of ANNs that were used for the present technique are then discussed, including the different configurations considered. The hydraulic systems selected are defined next, and the ANN training settings are described. Finally, results for the location and sizing of a junction (sizing referring to the determination of the two diameters at the junction in the pipeline) and for the location and sizing of a leak are presented. The proposed method is shown to be accurate and general for the considered systems; however, in addition to some concluding remarks, some challenges of the future application of this technique are briefly discussed.

**Background**

Fluid transient waves are an interesting mechanism for detecting existing faults, features in pipelines, and different elements of the topology of a water system (Bohorquez et al. 2018). However, this technique is highly sensitive to multiple system characteristics, and understanding what the pressure signal response should look like when a specific fault is present in realistic cases is often challenging (Xu and Karney 2017). Different authors have applied techniques using fluid transients to detect topological elements and to locate leaks in pipelines.

First, different techniques have been proposed to identify junctions, branches, and partially closed valves in pipelines. Brunone et al. (2008) applied time reflectometry to confirm the existence of a Y junction in a water main pipe in Italy. By interpreting the arrival of the transient wave reflection at the Y junction when it arrives at the measurement point, the precise position of the Y was computed accurately. Although successful, this technique requires the visual analysis of the transient pressure trace to determine the arrival of the reflected wave and depending on the system, can fail in terms of both precision and reproducibility.

Meniconi et al. (2011b) developed a laboratory test to locate branches on a system using a wavelet analysis and a reflection coefficient for estimating the size of the branch. Their results showed that a visual inspection of the wavelet transform is a good mechanism to locate branches, but the damping of the transient pressure waves affected the use of the reflection coefficient due to the pipe viscoelasticity and unsteady friction. In addition, to determine the condition of the branch (active or inactive), this approach used a MOC model to compare with the measured pressure traces.

Second, detection of leaks has received special attention given that water losses affect the performance, diminishes the operational age of a water supply system, and represents economic and environmental losses for a water utility (AL-Washali et al. 2016). The possibility of detecting single and distributed leaks by analyzing a transient pressure trace has been analyzed by different authors that explored techniques that can be divided in three groups: (1) inverse transient techniques, (2) frequency domain techniques, and (3) direct transient methods (Colombo et al. 2009).

Inverse transient analysis (ITA) techniques are based on the calibration of a numerical model in terms of the existence and characteristics of leaks to match the measured and the numerical transient pressure trace (Colombo et al. 2009). The first application of ITA and transient pressure was proposed by Liggett and Chen (1994) but a number of researchers have used ITA for detecting leaks in pipelines by modifying the optimization problem, the pressure measurements, and the selected algorithms (Covas et al. 2001; Vítkovský et al. 2003; Soares et al. 2011; Kim 2014; Capponi et al. 2017). Despite the accuracy of ITA methods in the location of leaks under different scenarios, they require a detailed numerical model of the analyzed system, they can report discrepancies in leak size estimations (Covas et al. 2001), and an optimization model needs to be executed each time a transient pressure trace is analyzed.

A second group of methods for locating leaks are the frequency domain techniques, which are often associated with determining the frequency response function of the system and comparing it with the one in a pipe without any anomalies (Gong et al. 2016). A first numerical application of these techniques was proposed by Mpesha et al. (2001) and since then, different approaches have been developed emphasizing the theory, procedure, and application of this technique (Lee et al. 2005; Sattar and Chaudhry 2008; Duan et al. 2010; Ghazali et al. 2010; Gong et al. 2013a; Duan 2017; Gong et al. 2016). However, frequency domain techniques require extensive measurements in the field and the frequency response function of the intact pipe, which can be obtained numerically using a detailed model (Colombo et al. 2009).

A third group of methods has been defined by Colombo et al. (2009) as direct transient methods. These methods aim to identify special features in the transient pressure trace that are induced by the presence of the leak. This group includes techniques such as time reflectometry (Brunone 1999; Lee et al. 2007; Brunone et al. 2008; Guo et al. 2012; Lazhar et al. 2013), transient signal damping analysis (Wang et al. 2002; Nixon et al. 2006; Brunone et al. 2018), and wavelet analysis (Ferrante and Brunone 2003; Ferrante et al. 2007, 2009; Meniconi et al. 2011a; Brunone et al. 2013). These methods have been successful in the detection and characterization of leak in pipelines; however, the physical
characteristics of the pipeline are required, and some approaches depend on knowing the transient pressure trace of an intact pipeline. In addition, experience is required to interpret the resulting pressure trace due to the existence of vibrations, background transients, and instrument noise (Colombo et al. 2009).

Although the majority of previous techniques have achieved satisfactory results in the task of detecting elements of the system topology or locating a leak using pressure transients, these methods require information about the analyzed pipeline (physical characteristics and resulting transient pressure traces after installation) or a detailed numerical model (model-based techniques). In addition, to obtain results extensive field tests may be needed, or a significant computational time is required for the analysis of each transient pressure trace obtained from a test. Thus, there is an advantage in using data-driven techniques that can provide accurate results fast and with the potential of applications to different water pipelines configurations.

ANNs have had a broad set of applications in different fields, and they have become a powerful tool in machine learning. The use of ANNs has proven to be effective, computationally efficient, and robust, provided enough information for its training exists (Caputo and Pelagagge 2003). In different applications, it has been demonstrated that results from ANNs are robust and are not sensitive to background noise in the analyzed system (Carlini and Wagner 2017; Mangal et al. 2019). In addition, ANNs have the ability to incorporate new information to previous training results provided that these results come from the same underlying distribution (Shamir et al. 2010). In water-related research, ANNs have been successfully applied to a variety of problems including water availability under climate change scenarios (Swain et al. 2017), asset failure prediction (Harvey et al. 2014), water demand prediction (Guo et al. 2018), cyber-physical attacks location (Taormina and Galelli 2018), and burst detection in water distribution systems (Romano et al. 2014), among many others.

Nonetheless, the use of ANNs for the identification of elements in the topology of a system, for leak detection and in general, for pipeline condition assessment using fluid transients, have not been previously explored to the knowledge of the authors. Only one application was reported by Belzito et al. (1998) where the location of leaks was predicted using ANNs in liquefied gas pipelines for different leak sizes. Pressure signals were used in their research as input information for the ANN, but transient waves were not considered. The results showed that the ANN was able to locate the leak accurately in more than 50% of the cases, for a leak size of 1% of the base flow. The research presented in the current paper is the first combined use of ANN and fluid transient waves to identify topological aspects and to detect and characterize leaks in water distribution system pipelines.

Artificial Neural Networks

An ANN implements a mathematical function (model) from $n$ inputs to $m$ outputs, and this function is represented by a mathematical graph of connections and nodes linking the input to the output. An example network is shown in Fig. 1. The inputs to the ANN are a vector of numerical values; these values are transmitted through the links of the graph to activation functions, which represent neurons. All links in the graph have an associated weight which is used to scale the value traveling on that link. Each activation function transforms the sum of the weighted values it receives, to an output value that is then propagated through the network. Thus, the input values are transformed by traversing the weighted links and the activation functions in the graph until they reach the output links.

To produce the desired behavior, an ANN needs to be trained. The training process modifies the weights associated with each of the links in the network to improve the accuracy of the model represented by the network. In theory, through modification of weights alone it is possible for a network of at least three layers, with enough links, to approximate an arbitrary function (Cybenko 1989). However, deep networks (with more than three layers) tend to be more practical to train, more versatile, and less prone to over-fit (Urban et al. 2016). The extent to which this approximation succeeds is determined by the interaction between the network architecture and its training process. In practice, training the network is a process of mathematical regression where a gradient search is used to adjust the weights in the network to minimize the error between the actual outputs of the network and the desired output. To be useful in the desired application domain, a network will approximate the required function to a high level of accuracy on both the data it was trained with and any new test data that it is presented with.

In the past decade, the technology used to train, specify, and implement ANNs has advanced rapidly. As a result, modern ANNs present the designer with a very broad range of design decisions for the ANN relating to topology, scale, activation functions, regularization strategies, and training methodology. As a general rule, the designer of an ANN has to use a network design that captures the behavior of the desired function without having so many weights (parameters) as to over-fit the data used in the training process.

In the application described in this paper, the functions that are approximated by the ANNs transform an input in the form of a transient head time series into continuous scalar values representing the location and the sizes of features in a pipeline as the output, as shown in Fig. 2. Thus, in this application, the input data (the transient head time series) may be quite large, and the resulting relationships to be learned are then quite complex. Two architectural options for the ANN have been explored in this paper. The most general available ANN structure was evaluated first, which corresponds to a fully connected dense network. Such networks are, in theory (Hornik et al. 1989), able to capture arbitrary functions. Subsequently, the use of a 1D convolutional network was explored. A-priori convolutional networks are known to be easier to train and capture structural invariants in their input signal more reliably than fully connected networks (Lecun et al. 1998).

A dense network, an example of which is shown in Fig. 3(a), connects each neuron in a layer to every neuron in the subsequent layer. Dense networks embed a lot of information in their weights and can express quite complex functions. However, the large number
of weights in dense networks create the risk the network will over-fit the training data and not generalize adequately to new data.

The second architectural option used in this paper is a 1D convolutional network [Fig. 3(b)]. These networks link each neuron in a layer to neurons in the corresponding neighborhood in the subsequent layer. Convolutional networks capture local interactions between data points and can work well in applications where there is potential to exploit spatial locality of features in input data. For a given number of input nodes, convolutional networks have many less weights and thus are less prone to over-fitting. Additionally, these networks have been successful in problems that require a deep configuration with a faster and easier training when compared with a dense network.

Hydraulic Systems Configuration

An ANN requires data as a training set to determine its characteristics and the functions behind its performance before being used for predicting a result. To generate the input data for detecting a junction or a leak (depending on the case), repetitive numerical simulations of the transient response of a system to a valve closure have been conducted using a MOC numerical simulation for transient behavior while changing the location and characteristics of the analyzed feature. This section describes the systems used for obtaining the input data for: (1) identifying junctions as a topological element, and (2) for detecting the presence of leaks.

Junction Model

The system considered for applying an ANN to identify topological elements is a junction in a single pipeline with two different diameters as described in Fig. 4. The pipeline is connected at the upstream end to a reservoir with a fixed head \( H_0 \) and on the downstream end to a side discharge valve. The length of the pipeline is fixed \( (L_T) \), the length of the upstream segment of the pipeline of diameter \( (D_1) \) is defined as \( x \), and the length of the downstream pipe segment of diameter \( (D_2) \) is then \( (L_T - x) \).

Steady-state conditions of the system were fixed for all the transient simulations. An initial head and an initial velocity in the pipeline (at the valve) were defined as \( H_0 = 70 \) m and \( V_0 = 0.15 \) m/s, respectively. Only steady-state friction was considered by using a Darcy–Weisbach friction factor, \( f \), with a pipeline roughness height of \( \varepsilon = 0.01 \) mm. The total length of the pipe was assumed to be \( L_T = 1,000 \) m.

The detection of the junction includes the prediction of both its location (by predicting the length of the upstream pipe segment) and the combination of diameters on either side of the junction in the pipeline. To accomplish this using the ANN, input training data included simulations of 10 different combinations of diameters on either side of the junction. These combinations were defined according to the Australian/New Zealand standard for ductile iron pipes with cement mortar lining and are presented in Table 1. Different wall and cement mortar lining thicknesses were considered for the different diameters. It is important to highlight that in the 10 combinations of diameters, five correspond to flow going from a larger to a smaller \( (D_1 > D_2) \) diameter, whereas five correspond to flow going from a smaller to a larger diameter \( (D_1 < D_2) \).

As it was mentioned previously, generation of the ANN input data (transient head traces at the valve after its closure) to train and test the ANN was accomplished by running multiple numerical simulations of the system using a conventional MOC. For each of the diameter combinations, the length of the upstream segment of the pipeline was changed along the complete length of the pipeline to simulate different locations of the junction. The number of generated locations, selection of these locations, the time step, and length of each reach used in the MOC are described later in this paper.

Fig. 2. ANN applied to pipe condition assessment.

Fig. 3. Architectural options of ANNs: (a) dense network; and (b) 1D convolutional network.
Table 1. Outer diameter combinations for detection of junctions

| Combination | \(D_1/D_2\) | \(D_1\) (mm) | \(D_2\) (mm) |
|-------------|-------------|-------------|-------------|
| 1           | 1.25        | 750         | 600         |
| 2           | 1.50        | 750         | 500         |
| 3           | 1.2         | 600         | 500         |
| 4           | 1.33        | 600         | 450         |
| 5           | 1.11        | 500         | 450         |
| 6           | 0.9         | 450         | 500         |
| 7           | 0.75        | 450         | 600         |
| 8           | 0.83        | 500         | 600         |
| 9           | 0.67        | 500         | 750         |
| 10          | 0.80        | 600         | 750         |

Source: Data from AS 2280 (AS 2014).

Table 2. Leak diameters considered for training and testing the ANN

| \(D_1\) (mm) | \(D_2\) (\%) |
|-------------|-------------|
| 13          | 1.79        |
| 14          | 1.92        |
| 15          | 2.06        |
| 16          | 2.20        |
| 17          | 2.34        |
| 18          | 2.47        |
| 19          | 2.61        |
| 20          | 2.75        |
| 21          | 2.89        |
| 22          | 3.02        |
| 23          | 3.16        |
| 24          | 3.30        |
| 25          | 3.44        |
| 26          | 3.57        |
| 27          | 3.71        |
| 28          | 3.85        |
| 29          | 3.99        |
| 30          | 4.12        |
| 31          | 4.26        |
| 32          | 4.40        |
| 33          | 4.54        |
| 34          | 4.67        |
| 35          | 4.81        |
| 36          | 4.95        |
| 37          | 5.09        |
| 38          | 5.22        |
| 39          | 5.36        |
| 40          | 5.50        |
| 41          | 5.64        |
| 42          | 5.77        |
| 43          | 5.91        |
| 44          | 6.05        |
| 45          | 6.19        |
| 46          | 6.32        |
| 47          | 6.46        |
| 48          | 6.60        |
| 49          | 6.74        |
| 50          | 6.87        |
| 51          | 7.01        |
| 52          | 7.15        |
| 53          | 7.29        |
| 54          | 7.42        |
| 55          | 7.56        |
| 56          | 7.70        |
| 57          | 7.84        |
| 58          | 7.97        |

Artificial Neural Network Training Settings

This section summarizes the ANN and input data set configuration that was required to develop a satisfactory application of ANNs for the location of junctions and leaks. Different settings needed to be defined including the type and characteristics of the ANN, and the characteristics of the input head data in terms of their nature, time resolution, and size.

Two different ANN architectures were considered, a dense network and a 1D convolutional network. Preliminary tests for the determination of the location of a junction, which are not included in this paper for brevity, showed that the prediction in the location of this feature with a 1D convolutional network were smoother and more uniform for junctions located across the analyzed pipeline with an average error of 1.14 m, whereas the results of the dense network predicted junction locations with errors up to 30 m. This ANN architecture was then used for both the location of the junction and the leak.
The 1D convolutional network used in this paper has been designed by defining a set of characteristics for it, by training the ANN for the location of a junction and by testing it to analyze its performance. These characteristics were defined to ensure that the final ANN was successful in finding the location of the junction, without over-fitting to the training data. The final set of characteristics of the 1D convolutional network include a network with: (1) the use of leaky rectified linear unit (Leaky ReLU) as an activation function, (2) three convolutional layers of size 1,200, 600, and 300, (3) 10 filters in each layer, and (4) three dense layers of size 21, 9, and 2 (or 3 depending on the analyzed feature). With this configuration, 32,409 weights have been trained.

To train and test the ANNs, multiple locations for the considered features (junction or leak, depending on the case) were selected for generating the input data. The number and the characteristics of these examples were defined in terms of the distribution of these locations (random or uniform locations along the pipeline), the time resolution of each transient head trace (original resolution or down sampled), and the number of locations along the pipeline (ANN input data size). To define these characteristics, preliminary training and testing runs were developed for the junction model and then applied to the leak model.

For determining the junction locations distribution along the pipeline (to create the ANN input data set of transient head traces), two options were considered: generating transient head traces for junctions at either uniform spacing or randomized spacing along the pipeline. A random distribution of the location of the junction proved to be more efficient when compared with a uniform distribution given that the ANN predictions of the preliminary tests were more accurate in terms of the median of the error and the maximum errors.

On the other hand, depending on the number of junctions considered to generate the input transient head traces, the distance between locations varied between 0.1 and 2 m and the numerical results of MOC needed to reflect different transient head traces for each location. To achieve this, the time resolution of the MOC is required to match the spatial separation of each location to apply the method along the characteristic lines. This implies that, for instance, for a spatial resolution of the junction location of 0.1 m and a total simulation time of 3 s, 10,000 junction locations would be necessary to cover a 1,000 m pipeline and each of these locations would have 20,000 head values. In this case, the complete input data set would include 200 million head values.

Considering the potential size of the ANN input data set, a time-wise down-sampling process was explored. To evaluate the performance of this process, two input data sets were created: one with transient head traces with 26,800 head values (for 10,000 different junction locations) and one with a uniform down sampling from the first data set to 1,200 head values, for the same 10,000 different junction locations. This indicates that the down-sampled transient head traces preserve the information from the original junction location even though they only contain 1,200 head values. Two 1D convolutional networks were then trained and tested 20 times, and the obtained results were analyzed in terms of the distribution of the average and the maximum absolute error for all the junctions across the pipeline.

Results from this sensitivity analysis are presented in Fig. 6. This figure shows that the median average absolute error is slightly larger for the down-sampled input data set; however, the total variation of these errors is smaller. In addition, the maximum absolute error is notably better when the down-sampling process is carried out. This behavior is explained primarily because the training process of the ANN is easier when fewer weights are required to describe the same head variation traces. Therefore, by down-sampling the transient head trace variation, the size of the input data decreases without significantly compromising the information contained within it.

Finally, the number of training and testing examples (ANN input sample size) was defined. The selection of this parameter is directly involved with the computational effort required to generate the input data and subsequently, to train the ANN. The most convenient input sample size was determined separately for the location of the junction and the leak because the transient head deviations induced by the presence of a leak are more subtle than for a junction, resulting in the necessity of using more training examples.

For the location of a junction, four different sample sizes were tested: 500, 1,000, 5,000, and 10,000 locations along the 1,000 m pipe. After performing the preliminary training and testing of the ANN, a sample size of 5,000 was selected as the most appropriate. It provided accurate results for the location of the junction with significantly less computational effort (the training process took 248 min) because the junction only needs to be located each 0.2 m for a sample size of 5,000 and not each 0.1 m as for a sample size of 10,000. Using a sample size of 5,000 indicated that to simulate 2.5 s, 13,400 head values would be calculated with a $\Delta t = 1.866 \times 10^{-4}$ s. After the down-sampling process, 1,200 head values were obtained with a $\Delta t = 0.0021$ s.

For the location of a leak, more samples were required. Four different input data size sets were tested: 5,000, 10,000, 25,000 and 50,000. Results of preliminary predictions made with ANNs trained with 25,000 and 50,000 were significantly better than the first two sample sizes. However, given the results of these tests it was not possible to choose a preferred sample size; therefore, results are presented for two ANNs, one trained with 25,000

![Fig. 6. Input data down sampling for junction location. Box whiskers plots of: (a) average absolute error; and (b) maximum absolute error.](image-url)
locations and the other with 50,000. In both of these sample sizes, the separation between leaks was 0.2 m (preserving the Δt described above for the location of the junction). Therefore, five (or 10 depending on the case) different locations were randomly generated each 0.2 m.

Results

Once the different settings for training an ANN were defined as described above, results for the prediction of the location and size of a junction and a leak have been obtained. For both cases (junction location and leakage), the results are presented in terms of the training and testing location errors, followed by testing of the size errors (diameters of both pipes on either side of the junction in the first case and diameter of the leak in the second case). In addition, a distribution of the error, represented as a percentage exceedance plot, for the location of the features is presented.

Junction Location and Sizing

Results for the topology identification application by locating and sizing a junction along a pipeline are presented first. Based on the hydraulic system configuration described in Fig. 1, the location of a junction was predicted using the ANN by the length of the upstream segment of the pipe x. Sizing of pipes on either side of the junction model refers to the determination of the two diameters associated with the two segments of the pipe. Those diameters are included in the combinations described in Table 1. The ANN used was a 1D convolutional network using a sample size of 5,000 (half for training and half for testing) with examples generated randomly (in 0.2 m intervals) and down sampled to 1,200 time steps.

Location errors are shown in Fig. 7(a) for the 2,500 examples used for the training stage of the application. The average absolute error for the training data set was 0.79 m with a maximum error of 7.21 m when the junction is located at 987 m downstream (within 13 m of the end of the pipe) of the reservoir (the total length of the pipe is 1,000 m). From this figure, it is possible to observe that near the extreme ends of the pipe, results tend to present larger errors in comparison to the results in the central part of the pipe. This behavior is due to the effect that those extreme locations have on the transient head trace at the measurement point. When close by, the reflections from the junction interact with the reflection at the reservoir (if the junction is close to the upstream end) or with the initial transient head rise (if the junction is close to the downstream end of the pipe). For these locations, instead of having a head drop (or increase depending on the ratio of diameters D₁/D₂), the presence of the junction causes a short spike at the beginning of the trace or close to a time of 2L/α s that makes it difficult for the ANN to use the same weights that represent this behavior in the rest of the pipeline.

On the other hand, testing location errors for the 2,500 examples used for this stage of the application are presented in Fig. 7(b). By comparing this figure with Fig. 7(a) it is possible to conclude that the ANN trained is not over-fitted because the error behavior is similar for the training and the testing data sets. Over-fitting is one of the key issues when working with ANNs because depending on its parameters and the input data, the ANN can perform satisfactorily for the training set but poorly for new unknown data as the test data set. The testing data set had an average absolute error of 0.81 m (slightly larger than the average absolute error for the training data set) and a maximum error of 8.3 m when the junction was located at 420 m from the reservoir.

Errors in the estimation of the values of both diameters of the pipe junction (D₁ and D₂) are shown in Fig. 8. In general, errors in diameter were extremely small, reaching a maximum of 27.1 mm for D₁ and only 7.25 mm for D₂ for the worst estimation. Average absolute errors were 0.31 and 0.27 mm, respectively. However, when predicting diameters, it is important to consider that pipe diameters cannot take on continuous values due to commercial production restrictions; therefore, results from Fig. 8(a) were rounded to the closest diameter in the list defined in Table 1 and are shown in Fig. 8(b). By performing this rounding procedure, only in one diameter prediction case (out of the 5,000 diameter predictions, two per each testing example) was the diameter not correctly identified.

Finally, Fig. 9 presents the percentage exceedance associated with the absolute error in the junction location prediction. This type of plot is useful to analyze the distribution of errors in detail. The percentage exceedance can be interpreted as the proportion of time that the junction location surpassed a certain error size. For instance, according to Fig. 9, in 5% of the testing examples, the junction location was predicted with an error of 2.32 m or larger. For reference, the maximum location error for the training and the testing data sets is also shown in the figure.

The distribution of the errors shows that the extreme values of error (larger than 3 m) are relatively rare in occurrence, and these correspond to only 2% of the total testing examples. In addition, the similarity in the distribution of error for the training and the testing data sets is validation of the adequate performance of the 1D convolutional network to predict the location of a junction in a single pipeline.

Leak Location and Sizing

Results for the prediction of the location and size of a leak are presented in this section. As described above, the preliminary tests
developed for determining the most adequate input training sample size for the location of the leak were inconclusive between using 25,000 and 50,000 examples for the training and testing of the ANN. Therefore, results are presented for both input sample sizes.

Fig. 10. Percentage exceedance for absolute junction location error.

Based on these findings, the ANN obtained from an input sample size of 50,000 was selected because it can predict the location of the leak with an average error of 1.15 m, and it shows a similar behavior when compared with the training errors (showing that less over-fitting issues are present). The maximum error for the 25,000 examples used for the testing (which corresponds to a total input data set of 50,000 because half of the data is used for training) was 98.21 m; however, this misleading result was predicted for an example where the leak was located only 1.01 m from the reservoir. Those errors were expected given the fact that the effect on the transient head trace of a leak located close to either of the extreme ends of the pipe can be difficult to distinguish from the reflections at the boundary conditions. In addition, it was observed that for leaks close to the reservoir or close to the closed valve, the down-sampling process in time could cause the loss of information on the reflection from the leak. However, in a real application, if the leak was located close to either of the extremes, it would be manually detected in the process of obtaining the transient data. In general, the time down-sampling process does not affect the accuracy of the technique because even though some information is lost in this process, the ANN is sensitive to the changes in pressure that are not lost in the down sampling.

Errors for the leak size are presented in Fig. 11, only for the ANN obtained from an input sample size of 50,000. This figure shows that errors in the prediction of the leak size are highly satisfactory because most of them are within the range of −0.2, 0.2 mm. The maximum error is 1.11 mm when the leak is located only 61 cm away from the reservoir, and the average absolute error is 0.03 mm. Considering that in a real pipeline, leak size could take any continuous value, a rounding process was not performed as was done for the diameter size in the junction sizing. In general, results from Fig. 11 show that the prediction of the leak size using ANNs is successful. Fig. 12 presents the percentage exceedance for the absolute leak location error for both the training and the testing data set. To improve the analysis of the data, the first and the last 12 m (which would correspond approximately to two segments of pipe in the field) were eliminated before compiling the figure because it allows us to see the behavior of the error once the extreme and unrealistic errors are discarded. By doing this, the maximum error for both the testing and the training is about 12 m, and errors larger than 4 m are only found in 1.89% of the examples. Likewise, in 95% of the examples, the ANN prediction for the location of the leak in the pipe is at least or more accurate than 3.0 m, which represents 0.3% of the total length of the 1,000-m-long analyzed pipe. Finally, a similar distribution of location errors for the training and the testing data sets demonstrate the adequate performance of the ANN when it is predicting locations in new examples.
Conclusions

A novel technique to identify topological elements such as junctions and to detect, locate, and size leaks has been proposed in this paper using ANNs as a tool based on the generation of a transient event. Two ANN architectures have been compared, and a 1D convolutional network proved to predict more accurately the location of a pipeline junction in comparison with a dense network. The transient head data required for the training and testing of the designed ANNs have been obtained numerically using the MOC by changing the position of the analyzed feature randomly along the length of the pipeline. A time down-sampling process has been conducted to reduce the time resolution of the transient head traces, which reduced the computational time and enhanced the performance of the ANN in predicting the location of the features. The number of training examples required to obtain accurate results has been defined. For the location and sizing of a junction, 2,500 examples are enough to train an ANN with accurate results. However, for the location and sizing of a leak, 25,000 examples are necessary to train the ANN because the effect of a leak in the transient head variation trace is more subtle and is more difficult for the ANN to learn from the input data.

The numerical application of the technique shows that an ANN can accurately predict the location of a junction with an error in the

![Fig. 10. Location errors for leak location: (a) training data set errors with 25,000 training examples; (b) testing data set errors with 25,000 training examples; (c) training data set errors with 50,000 training examples; and (d) testing data set errors with 50,000 training examples.](image)

![Fig. 11. Leak size errors for testing data set.](image)

![Fig. 12. Percentage exceedance for absolute leak location error.](image)
location smaller than 2.32 meters in 95% of the examples tested with a virtually perfect prediction of the diameter sizes on each side of the pipeline junction. In addition, for the location and sizing of a leak, the trained ANN can also accurately predict the leak location with an error smaller than 3.0 mm in 95% of the examples tested with an average absolute error of 0.03 mm in the prediction of the leak size. These results demonstrate the outstanding potential of using machine learning techniques and fluid transient waves for the location of topological elements and anomalies in pipelines. The approach proposed is fast, accurate, and data-driven because no previous information of the system or a hydraulic transient numerical model is required for the testing stage; only a transient head variation trace is needed.

Considering this is the first application of fluid transient waves and ANNs, the proposed approach has been tested in simple hydraulic systems with data obtained numerically; however, some challenges arise when the same technique is applied to more complex situations, and the impact of these challenges will be addressed in future field tests of the performance of this technique. Detection of topological elements and anomalies such as leaks is, in essence, a complex problem. Pipelines and water distribution systems have different elements that can affect the transient head signal obtained in the field, and changes in background conditions can make the detection of anomalies challenging. However, the background fluctuations are typically more gradual with lower frequency content allowing them to be differentiated from other anomalies. In addition, for subtle anomalies (such as small leaks), the deteriorating condition of the pipeline itself can hide the transient head response of the anomaly. Considering these challenges, it would be expected that the accuracy of the use of the proposed technique would be reduced when applied to real pipelines. Nonetheless, ANNs are highly amenable to retraining on real data as it is harvested, and ANNs are robust to noise. The results here offer promise that the use of ANNs can be extended to field settings.

The application presented in this paper demonstrates the promising potential of the proposed technique, yet it should be expanded to the analysis of hydraulic systems with different dimensions and configurations, and validated with laboratory and field data to explore its usefulness and accuracy in comparison to other existing techniques.

Data Availability Statement

Some or all the data, models, or code generated or used during the study are available from the corresponding author upon request.

References

AL-Washali, T., S. Sharma, and M. Kennedy. 2016. “Methods of assessment of water losses in water supply systems: A review.” Water Resour. Manage. 30 (14): 4985–5001. https://doi.org/10.1007/s11269-016-1503-7.

AS (Standards Australia). 2014. Ductile iron pipes and fittings. AS 2280. Sydney, Australia: AS.

Belsito, S., P. Lombardi, P. Andreussi, and S. Banerjee. 1998. “Leak detection in liquefied gas pipelines by artificial neural networks.” AIChE J. 44 (12): 2675–2688. https://doi.org/10.1002/aic.690441209.

Bohorquez, J., A. R. Simpson, and M. F. Lambert. 2018. “Characterization of transient pressure traces due to the effects of different anomalies and features in water pipelines.” In Proc., 13th Pressure Surges Conf., 151–169. Cranfield, UK: BHR Group.

Brunone, B. 1999. “Transient test-based technique for leak detection in outfall pipes.” J. Water Resour. Plann. Manage. 125 (5): 302–306. https://doi.org/10.1061/(ASCE)0733-9496(1999)125:5(302).

Brunone, B., M. Ferrante, and S. Meniconi. 2008. “Portable pressure wave-maker for leak detection and pipe system characterisation.” J. Am. Water Works Assn. 100 (4): 108–116. https://doi.org/10.1002/j.2151-8833.2008.tb09607.x.

Brunone, B., M. Ferrante, S. Meniconi, and C. Massari. 2013. “Effectiveness assessment of pipe systems by means of transient test-based techniques.” Procedia Environ. Sci. 19: 814–822. https://doi.org/10.1016/j.proenv.2013.06.090.

Brunone, B., S. Meniconi, and C. Capponi. 2018. “Numerical analysis of the transient pressure damping in a single polymeric pipe with a leak.” Urban Water J. 15 (8): 760–768. https://doi.org/10.1080/1573062X.2018.1547772.

Capponi, C., M. Ferrante, A. C. Zecchin, and J. Gong. 2017. “Leak detection in a branched system by inverse transient analysis with the admittance matrix method.” Water Resour. Manage. 31 (13): 4075–4089. https://doi.org/10.1007/s11269-017-1730-6.

Caputo, A., and P. M. Pelagagge. 2003. “Using neural networks to monitor piping systems.” 22 (2): 119–127. https://doi.org/10.1016/j.proenv.20022008.

Carlini, N., and D. Wagner. 2017. “Towards evaluating the robustness of neural networks.” In Proc., 2017 IEEE Symp. on Security and Privacy (SP), 39–57. New York: IEEE.

Colombo, A., P. Lee, and B. Carney. 2009. “A selective literature review of transient-based leak detection methods.” J. Hydro-environ. Res. 2 (4): 212–227. https://doi.org/10.1016/j.jher.2009.02.003.

Covas, D., I. Stoianov, D. Butler, C. Maksimovic, N. Graham, and H. Ramos. 2001. “Leakage detection in pipeline systems by inverse transient analysis-from theory to practice.” In Proc., Int. Conf. on Computing and Control for the Water Industry (CCWI 2001), 3–16. Hong Kong: Hong Kong Polytechnic Univ.

Cybenko, G. 1989. “Approximation by superpositions of a sigmoidal function.” Math. Control Signals Syst. 2 (4): 303–314. https://doi.org/10.1007/BF02551274.

Duan, H. F. 2017. “Transient frequency response based leak detection in water supply pipeline systems with branched and looped junctions.” J. Hydroinf. 19 (1): 17–30. https://doi.org/10.2166/hydro.2016.008.

Duan, H. F., P. J. Lee, M. S. Ghidaoui, and Y. K. Tung. 2010. “Essential system response information for transient-based leak detection methods.” J. Hydraulic Res. 48 (5): 650–657. https://doi.org/10.1080/00221686.2010.507014.

Ferrante, M., and B. Brunone. 2003. “Pipe system diagnosis and leak detection by unsteady-state tests. 2: Wavelet analysis.” Adv. Water Resour. 26 (1): 107–116. https://doi.org/10.1016/S0309-1708(02)00102-1.

Ferrante, M. B. Brunone, and S. Meniconi. 2007. “Wavelets for the analysis of transient pressure signals for leak detection.” J. Hydraulic. Eng. 133 (11): 1274–1282. https://doi.org/10.1061/(ASCE)0733-9429(2007)133:11(1274).

Ferrante, M., B. Brunone, and S. Meniconi. 2009. “Leak detection in branched pipe systems coupling wavelet analysis and a Lagrangian model.” J. Water Supply Res. Technol. AQUA 58 (2): 95–106. https://doi.org/10.2166/aqua.2009.022.

Ghazali, M. F., W. J. Staszewski, J. D. Boxall, and S. B. M. Beck. 2010. “Instantaneous phase and frequency for the detection of leaks and features in a pipeline system.” Struct. Health Monit. 10 (4): 351–360. https://doi.org/10.1177/1475921710373958.

Gong, J., M. F. Lambert, S. T. N. Nguyen, A. C. Zecchin, and A. R. Simpson. 2018. “Detecting thin-walled pipe sections using a spark transient pressure wave generator.” J. Hydraulic. Eng. 144 (2): 06017027. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001409.

Gong, J., M. F. Lambert, A. Simpson, and A. Zecchin. 2015a. “Single-event leak detection in pipeline using first three resonant responses.” J. Hydraulic. Eng. 139 (6): 645–655. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001020.

Gong, J., M. F. Lambert, A. R. Simpson, and A. C. Zecchin. 2014. “Detection of localized deterioration distributed along single pipelines by reconstructive MOC analysis.” J. Hydraulic. Eng. 140 (2): 190–198. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000806.
Gong, J., M. F. Lambert, A. Zecchin, and A. Simpson. 2016. “Experimental verification of pipeline frequency response extraction and leak detection using the inverse repeat signal.” J. Hydraul. Res. 54 (2): 210–219. https://doi.org/10.1080/00221686.2015.1116115.

Gong, J., A. R. Simpson, M. F. Lambert, A. C. Zecchin, Y. I. Kim, and A. Tijskering. 2013b. “Detection of distributed deterioration in single pipes using transient reflections.” J. Pipeline Syst. Eng. Pract. 4 (1): 32–40. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000111.

Guo, G., S. Liu, Y. Wu, J. Li, R. Zhou, and X. Zhu. 2018. “Short-term water demand forecast based on deep learning method.” J. Water Resour. Plann. Manage. 144 (12): 04018076. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000992.

Guo, W., L. Soibelman, and J. H. Garrett. 2009. “Visual pattern recognition supporting defect reporting and condition assessment of wastewater collection systems.” J. Comput. Civ. Eng. 23 (3): 160–169. https://doi.org/10.1061/(ASCE)0887-3801(2009)23:3(160).

Lee, P., M. F. Lambert, A. R. Simpson, J. Liggett, and J. Vítkovský. 2003. “Optimal measurement site locations for inverse transient analysis in pipe networks.” J. Water Resour. Plann. Manage. 129 (6): 480–492. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:6(480).

Mangal, R., A. V. Nori, and A. Orso. 2019. “Robustness of neural networks: A probabilistic and practical approach.” In Proc., 41st Int. Conf. on Software Engineering: New Ideas and Emerging Results, 93–96. Montreal: IEEE Press.

Meniconi, S., B. Brunone, M. Ferrante, and C. Massari. 2011a. “Small amplitude sharp pressure waves to diagnose pipe systems.” Water Resour. Manage. 25 (1): 79–96. https://doi.org/10.1007/s11269-010-9688-7.

Meniconi, S., B. Brunone, M. Ferrante, and C. Massari. 2011b. “Transient tests for locating and sizing illegal branches in pipe systems.” J. Hydroinf. 13 (3): 334–345. https://doi.org/10.2166.hydro.2011.012.

Mpesh, W., S. Gassman, and M. H. Chaudhry. 2001. “Leak detection in pipes by frequency response method.” J. Hydraul. Eng. 127 (2): 134–147. https://doi.org/10.1061/(ASCE)0733-9429(2001)127:2(134).

Nixon, W. M. S, Ghiadaou, and A. A. Kolyshkin. 2006. “Range of validity of the transient damping leakage detection method.” J. Hydraul. Eng. 132 (9): 944–957. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:9(944).

Romano, M., Z. Kapelan, and D. A. Savić. 2014. “Automated detection of pipe bursts and other events in water distribution systems.” J. Water Resour. Plann. Manage. 140 (4): 457–467. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000339.

Sattar, A. M., and M. H. Chaudhry. 2008. “Leak detection in pipelines by frequency response method.” Supplement, J. Hydraul. Res. 46 (S1): 138–151. https://doi.org/10.1080/00221686.2008.9521948.

Shamir, O., S. Sabato, and N. Tishby. 2010. “Learning and generalization with the information bottleneck.” Theor. Comput. Sci. 411 (29): 2696–2711. https://doi.org/10.1016/j.tcs.2010.04.006.

Shi, H., J. Gong, A. C. Zecchin, M. F. Lambert, and A. R. Simpson. 2017. “Hydraulic transient wave separation algorithm using a dual-sensor with applications to pipeline condition assessment.” J. Hydroinf. 19 (5): 752–765. https://doi.org/10.2166.hydro.2017.146.

Soares, A. K., D. Covas, and L. F. Reis. 2011. “Leak detection by inverse transient analysis in an experimental PVC pipe system.” J. Hydrow. 13 (2): 153–166. https://doi.org/10.2166.hydro.2010.012.

Swain, E. D., J. Gómez-Fragsoso, and S. Torres-Gonzalez. 2017. “Projecting impacts of climate change on water availability using artificial neural network techniques.” J. Water Resour. Plann. Manage. 143 (12): 04017068. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000844.

Taormina, R., and S. Galelli. 2018. “Deep-learning approach to the detection and localization of cyber-physical attacks on water distribution systems.” J. Water Resour. Plann. Manage. 144 (10): 04018065. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000983.

Urban, G., K. J. Geras, S. Ebrahimi Khou, O. Aslan, S. Wang, R. Caruana, A. Mohamed, M. Philipose, and M. Richardson. 2016. “Do deep convolutional nets really need to be deep and convolutional?” Preprint, submitted March 17, 2016. https://arxiv.org/abs/1603.05691.

Vítkovský, J. P., J. A. Liggett, A. R. Simpson, and M. F. Lambert. 2003. “Optimal measurement site locations for inverse transient analysis in pipe networks.” J. Water Resour. Plann. Manage. 129 (6): 480–492. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:6(480).

Wang, W., J. Boon, and X. Kong. 2012. “Condition assessment of live PCCP lines with free-swimming electromagnetic inspection system.” In ICPTT 2012: Better pipeline infrastructure for a better life. Reston, VA: ASCE.

Wang, X. J., M. F. Lambert, A. R. Simpson, J. Liggett, and J. Vítkovský. 2002. “Leak detection in pipelines using the damping of fluid transients.” J. Hydraul. Eng. 128 (7): 697–711. https://doi.org/10.1061/(ASCE)0733-9429(2002)128:7(697).

Xu, X., and B. Karney. 2017. “An overview of transient fault detection techniques.” In Modeling and monitoring of pipelines and networks: Advanced tools for automatic monitoring and supervision of pipelines, edited by C. Verde and L. Torres, 13–37. Cham, Switzerland: Springer.

Zheng, L., and K. Yehuda. 2013. “State of the art review of inspection technologies for condition assessment of water pipes.” Measurement 46 (1): 1–15. https://doi.org/10.1016/j.measurement.2012.05.032.