A Model-Based Bayesian Framework for Pipeline Leakage Enumeration and Location Estimation

Juan Li (✉ ljuan@jlu.edu.cn)

Jilin University

Ying Wu

Jilin University

Changgang Lu

Jilin University

Research Article

Keywords: Leakage localization, Leakage enumeration, Bayesian inference, Nested sampling, Transient pressure wave

Posted Date: June 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-401326/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A Model-Based Bayesian Framework for Pipeline Leakage Enumeration and Location Estimation

Juan Li*, Ying Wu, Changgang Lu

Abstract

Leak detection in pipelines is an important issue, because leakages pose financial losses, environmental pollution and even health risks. The paper considers the problem of detecting multiple leaks based on transient wave theory for a reservoir pipeline valve system. The given measured data under consideration may contain one or multiple leaks originating from different locations when estimating the leak locations. This leads to two problems to be solved: first determine the correct number of leaks, and then identify the actual location of each leak. Thus, a probabilistic method of model-based Bayesian analysis is applied to this paper. This work employs a model to describe various scenes, individually defined by a specific number of leaks and their locations. Bayesian inference is used to select which model that is the most appropriate to fit the measured data. Through the process, the number of leaks is first estimated, and then the leak locations are extracted from the model that the measured data prefers. This paper presents different experimental setups and scenarios to demonstrate the availability of the proposed method, demonstrating that this model-based Bayesian analysis is an accurate tool for leakage enumeration and location estimation.

Keywords Leakage localization · Leakage enumeration · Bayesian inference · Nested sampling · Transient pressure wave

Juan Li (corresponding author): College of Communication Engineering, Jilin University, Changchun, China. E-mail: ljuan@jlu.edu.cn

Ying Wu: College of Communication Engineering, Jilin University, Changchun, China.

E-mail: wuying18@mails.jlu.edu.cn

Changgang Lu: College of Automotive Engineering, Jilin University, Changchun, China.

E-mail: lucg@jlu.edu.cn

1 Introduction

With the rapid development of economy, pipeline transportation is playing an increasingly important role in the national economy, national defense industrial and people's daily life. Oil, natural gas, and water pipelines have become the lifeblood of national economic development (Li et al. 2019). Pipelines are the safest means of transportation, but this does not mean that they are risk-free. In recent years, pipeline leakage accidents are the main source of water losses and occur frequently (Gupta and Kulat, 2018). The impact of water loss in urban water supply systems on water and energy resources as well as the quality of public services has become a continuous and global challenge (Colombo and Karney 2002; Duan 2018; Del Teso et al. 2019). In order to ensure the safe work of pipelines and minimize the harm caused by leakage accidents, it is necessary to study leakage detection technology to raise the sensitivity of detection.

Various leak detection methods have been developed for decades. Recent researches have focused on transient-based leak detection methods. Transient-based technology utilizes the hydraulics of transient flow and measured pressure response at specified location(s) to detect leaks in the pipeline (Wang and Ghidaoui, 2018). The reason that such methods are expected to work is that the pressure response signal in fluid conduits measured at a specific location is changed by its interaction with the physical system as it propagates and reflects throughout the system as a whole. It contains useful information of the conduit's properties and state. Specific methodological examples of this approach are (i) inverse transient-based method (ITM) (Soares et al. 2011; Stephens et al. 2013; Vitkovsky et al. 2007); (ii) frequency response-based method (FRM) (Duan 2016a; Sattar and Chaudhry 2008; Kim 2016); (iii) transient damping-based method (TDM) (Wang et al. 2002); and (iv) transient reflection-based method (Covas et al. 2004; Sun et al. 2016).

In recent researches (Li et al. 2020; Wang and Ghidaoui 2018; Wang et al. 2019), the authors respectively formulated a transient leak detection method and found that the approach can locate leaks for the cases of one leak and two leaks. However, the simulation results also indicated that the localization

efficiency of two leaks depends on the leak locations. Moreover, these methods cannot simultaneously determine the number and locations of leaks. These are precisely the shortcomings that this paper addresses.

Model-based Bayesian inference is a methodology that is capable of incorporating information related to the system under study, including a sufficiently accurate model derived from the observed phenomena. These methods have been used more and more in recent years. Previous researches (Bush and Xiang 2018; Escolano et al. 2014) applied Bayesian inference to DoA analysis. Knuth et al.(2014) applied Bayesian analysis to the context of signal processing. Bayesian model selection has also been employed to room-acoustic modal analysis (Beaton and Xiang 2017). Inspired by the above research, the paper presents a transient model-based Bayesian analysis for pipeline leak localization using nested sampling, which is a method that can infer the number of leaks and their locations through probabilistic analysis.

The following contents of this paper are as follows. This paper begins by describing the pipe system and the frequency-domain transient wave model. Then, the Bayesian framework is introduced: the basis of Bayesian parameter estimation and model selection, followed by the introduction of nested sampling. Numerical simulations are then presented, where cases of two leaks and three leaks are both considered, and the performance of the proposed methods is evaluated and the results are shown. Finally, conclusions are drawn.

2 Linearized Model of Wave Propagation in Pipeline

In this section, the model of transient wave propagation in a pipeline is introduced. The configuration of the considered pipe system is shown in Fig. 1. A single, horizontal pipeline is connected by an upstream and a downstream node with locate at $x = x_U = 0$ and $x = x_D = l$. A valve is positioned at $x = x_D$ to generate transient waves. A pressure measurement station is assumed to be placed near the downstream node whose coordinate is $x = x_M$. The locations $(x_{L1}, \dots, x_{LN})$ of N leaks are parameters to be estimated. Q_o^L and H_o^L denote the steady-state discharge and head at the leak, respectively. The lumped leak parameter $s_L = C_d A_L$ stands for the leak size, in which C_d is the discharge coefficient of the leak and A_L denotes the flow area of

the leak opening (orifice). The steady-state discharge of the leak is related to the lumped leak parameter by $Q_0^L = s_L \sqrt{2g(H_0^L - z_L)}$, where z_L denotes the elevation of the pipe at the leak and g is the gravitational acceleration.

Assuming that a single pipe has equal section, equal wall thickness and the same pipe material, its field matrix can be expressed as follows (Chaudhry 2014)

$$F(x) = \begin{bmatrix} \cosh(\mu x) & -\frac{1}{Z} \sinh(\mu x) \\ -Z \sinh(\mu x) & \cosh(\mu x) \end{bmatrix} \quad (1)$$

where $\mu = a^{-1} \sqrt{-\omega^2 + i g A \omega R}$ is the propagation function and $Z = \mu a^2 / (i \omega g A)$ is the characteristic impedance, where a is the wave speed, ω denotes the angular frequency, A is area of pipeline, R is steady-state friction resistance term by linearization, equal to $(f Q_0) / (g d A^2)$ for turbulent flow (Duan et al. 2018), in which f is the Darcy-Weisbach friction factor, Q_0 is the steady-state discharge, and d is the pipe diameter.

The flow discharge and head oscillations due to a rapid change in flow setting are represented by q and h . Given the discharge $q(x_U)$ and head $h(x_U)$ at the upstream node x_U , the quantities at x_M can be obtained via the transfer matrix method after linearization

$$\begin{Bmatrix} q(x_M) \\ h(x_M) \end{Bmatrix} = F^{NL}(x_M - x_{L_N}) \prod_{n=1}^N \begin{Bmatrix} 1 & -\frac{Q_0^{L_n}}{2(H_0^{L_n} - z_{L_n})} \\ 0 & 1 \end{Bmatrix} F^{NL}(x_{L_n} - x_{L_{n-1}}) \begin{Bmatrix} q(x_U) \\ h(x_U) \end{Bmatrix} \quad (2)$$

where the superscript NL stands for no leak.

By rewriting Eq. (2), the head measurement at $x = x_m$ ($m = 1, \dots, M$) near the downstream for a given angular frequency ω_j ($j = 1, 2, \dots, J$) is assumed to follow the theoretical expression from Eq. (2) plus a noise term (Wang et al. 2018):

Fig. 1 Setup of the considered pipeline system

$$h(\omega_j, x_m) = h^{NL}(\omega_j, x_m) + \sum_{n=1}^N s_{L_n} G(\omega_j, x_{L_n}, x_m) + n_{jm} \quad (3)$$

wherein

$$h^{NL}(\omega_j, x_m) = -Z \sinh(\mu x_m) q(x_U) + \cosh(\mu x_m) h(x_U) \quad (4)$$

and n_{jm} is assumed to follow additive independent Gaussian random distribution with zero mean and covariance σ^2 .

The corresponding head difference due to leaks can be approximates via a linearized model:

$$\Delta \mathbf{h} = \mathbf{G}(x_L) s_L + \mathbf{n} \quad (5)$$

in which $\Delta h_{jm} = h(\omega_j, x_m) - h^{NL}(\omega_j, x_m)$ represent the head difference between the head measurement in the presence of leaks and the theoretical head that does not include the leak terms at the measurement station x_m and at the frequency ω_j . The data $\Delta \mathbf{h}$ is used in the following section.

In Eq. (5), the n -th column of matrix $\mathbf{G}(x_L)$ is

$$\mathbf{G}(x_{L_n}) = (G(\omega_1, x_{L_n}, x_1), \dots, G(\omega_J, x_{L_n}, x_1), \dots, G(\omega_1, x_{L_n}, x_M), \dots, G(\omega_J, x_{L_n}, x_M))^T \quad (6)$$

$$\Delta \mathbf{h} = (\Delta h_{11}, \dots, \Delta h_{J1}, \dots, \Delta h_{1M}, \dots, \Delta h_{JM})^T \quad (7)$$

$$\mathbf{n} = (n_{11}, \dots, n_{J1}, \dots, n_{1M}, \dots, n_{JM})^T \quad (8)$$

where

$$G(\omega_j, x_{L_n}, x_m) = \begin{cases} -\frac{\sqrt{g} Z \sinh(\mu(x_m - x_{L_n}))}{\sqrt{2(H_0^{L_n} - z_{L_n})}} (Z \sinh(\mu x_{L_n}) q(x_U) - \cosh(\mu x_{L_n}) h(x_U)), & x_m > x_{L_n} \\ 0, & x_m \leq x_{L_n} \end{cases} \quad (9)$$

Here, The upstream is connected to a reservoir, thus $h(x_U) = 0$ is reasonably assumed. The discharge $q(x_U)$ can be estimated by adding a pressure sensor near the upstream boundary denoted by x_{M0} . Assuming there is no leak between x_U and x_{M0} and using the pressure head measurement $h(x_{M0})$ at x_{M0} and applying the boundary $h(x_U)$ at x_U , the discharge $q(x_U)$ at x_U can be solved via

$$\begin{pmatrix} q(x_U) \\ h(x_U) \end{pmatrix} = M^{NL}(x_U - x_{M0}) \begin{pmatrix} q(x_{M0}) \\ h(x_{M0}) \end{pmatrix} = \begin{pmatrix} \cosh(\mu(x_U - x_{M0})) & -\frac{1}{Z} \sinh(\mu(x_U - x_{M0})) \\ -Z \sinh(\mu(x_U - x_{M0})) & \cosh(\mu(x_U - x_{M0})) \end{pmatrix} \begin{pmatrix} q(x_{M0}) \\ h(x_{M0}) \end{pmatrix} \quad (10)$$

that is,

$$\hat{q}(x_U) = \frac{\cosh(\mu(x_{M_0} - x_U))h(x_U) - h(x_{M_0})}{Z \sinh(\mu(x_{M_0} - x_U))} = -\frac{h(x_{M_0})}{Z \sinh(\mu(x_{M_0} - x_U))} \quad (11)$$

3 Bayesian Inference

3.1 Parameter Estimation

Bayesian inference is extensively based on the Bayes's theorem. This section introduces the Bayesian analysis for pipeline leakage localization problem. Let $\Theta = \{x_{L1}, x_{L2}, \dots\}$ represent a vector of model parameters (leak locations). For a given dataset $\mathbf{D}$ and a given model $\mathbf{M}$, the Bayesian inference formulated in this specific problem begins with Bayes's theorem (Escolano et al. 2014; Landschoot and Xiang 2019)

$$P(\Theta | \mathbf{D}, \mathbf{M}) = \frac{P(\mathbf{D} | \Theta, \mathbf{M})P(\Theta | \mathbf{M})}{P(\mathbf{D} | \mathbf{M})} \quad (12)$$

The term $P(\Theta | \mathbf{D}, \mathbf{M})$ represents the posterior probability of model parameters, Θ , which quantifies the state of information for the parameters. $P(\mathbf{D} | \Theta, \mathbf{M})$ is referred as the likelihood function, which indicates the resemblance of the data $\mathbf{D}$ for a given model parameter to the model $\mathbf{M}$. In the paper, the dataset $\mathbf{D}$ is the pressure difference $\Delta \mathbf{h}$. Since the vector of head difference $\Delta \mathbf{h}$ follows a JM -dimensional complex-valued Gaussian distribution, its probability density function is

$$P(\mathbf{D} | \Theta, \mathbf{M}) \equiv L(\Theta) = (\pi\sigma^2)^{-JM} \exp\left(-\frac{1}{\sigma^2} \|\Delta \mathbf{h} - \mathbf{G}(x^L)\mathbf{s}^L\|^2\right) \quad (13)$$

The term $P(\Theta | \mathbf{M})$ represents the prior distribution of the parameters given the model, $\mathbf{M}$. It represents any prior knowledge we may have about the likely values of the parameters. According to the principle of maximum entropy, this should be uniformly distributed to avoid any preference. Finally, the conditional probability $P(\mathbf{D} | \mathbf{M})$ of the observed data for a particular model can be considered the likelihood of the model given the data, referred to as the marginal likelihood or the Bayesian evidence or just evidence for $\mathbf{M}$.

The posterior probability of the parameter satisfies the following conditions,

$$\int_{\Theta} P(\Theta | \mathbf{D}, \mathbf{M}) d\Theta = \int_{\Theta} \frac{P(\mathbf{D} | \Theta, \mathbf{M})P(\Theta | \mathbf{M})}{P(\mathbf{D} | \mathbf{M})} d\Theta = 1 \quad (14)$$

This equation can be rearranged to

$$P(\mathbf{D} | \mathbf{M}) \equiv Z = \int_{\Theta} P(\mathbf{D} | \Theta, \mathbf{M}) P(\Theta | \mathbf{M}) d\Theta \quad (15)$$

The marginal likelihood value, $P(\mathbf{D} | \mathbf{M})$ is evaluated over the whole parameter space by integrating the product of the likelihood and prior distribution.

3.2 Model Selection

In a Bayesian formulation, estimating the number of sources is an application of model selection. Bayesian model selection is a probabilistic method of evaluating a finite set of models, given the observed data, and then seeking the model that best describes the data. The idea behind the model selection is to compare the posterior probability of a set of competing models (Escolano et al. 2014; Landschoot and Xiang 2019; Sivia and Skilling 2006). This can be determined by the probability of the model, $\mathbf{M}_i$, given the data, $\mathbf{D}$, represented as $P(\mathbf{M}_i | \mathbf{D})$. Bayes's theorem states that

$$P(\mathbf{M}_i | \mathbf{D}) = \frac{P(\mathbf{D} | \mathbf{M}_i) P(\mathbf{M}_i)}{P(\mathbf{D})} \quad (16)$$

$P(\mathbf{M}_i | \mathbf{D})$ is the posterior probability of the model, $P(\mathbf{D})$ is the probability of the measured data, which is independent of $\mathbf{M}_i$, and for this research it will be used as a constant that will not be of interest. $P(\mathbf{M}_i)$ is the prior probability of the model, and should be assigned based on any previous knowledge we may have about the likely values of the models. In the research, each model will be given equal prior probability in order to avoid preference for any model.

For convenience, the posterior ratio of two models $\mathbf{M}_i$ and $\mathbf{M}_j$ is defined as

$$F_{ij} = \frac{P(\mathbf{M}_i | \mathbf{D})}{P(\mathbf{M}_j | \mathbf{D})} = \frac{P(\mathbf{D} | \mathbf{M}_i) P(\mathbf{M}_i)}{P(\mathbf{D} | \mathbf{M}_j) P(\mathbf{M}_j)}, \quad i \neq j \quad (17)$$

when assigning the competing models equal prior probability, i.e., $P(\mathbf{M}_i) = P(\mathbf{M}_j)$; the model selection is determined in terms of the likelihood function $P(\mathbf{D} | \mathbf{M}_i)$. If the numerator is greater than the denominator, the data prefers model $\mathbf{M}_i$ over $\mathbf{M}_j$. The likelihood function in model selection is exactly the Bayesian

evidence term in parameter estimation task. Therefore, model selection can be carried out just by comparing evidences obtained in the process of parameter estimation.

4 Nested Sampling

4.1 Nested Sampling Algorithm Theory

At the heart of Bayesian inference is calculating the evidence of each model for comparison, so different sampling methods must be put in place. A numerical approach is to calculate the Bayesian evidence using a sampling algorithm, termed nested sampling. The chosen approach exploits the prior and likelihood as inputs and generates samples from the posterior as a secondary output.

Nested sampling utilizes the close relationship between the likelihood function $L(\Theta)$ and the prior mass $\xi(\lambda)$. In the terminology of the subject, mass denotes an accumulated amount of probability and the prior mass can be accumulated from its elements $d\xi = P(\Theta|\mathbf{M})d\Theta$ in any order (Skilling 2004), so let us define

$$\xi(\lambda) = \int_{L(\Theta) > \lambda} P(\Theta|\mathbf{M})d\Theta \quad (18)$$

as the cumulant prior mass covering all likelihood values greater than λ . As λ increases, the enclosed mass ξ decreases from 1 to 0. Thus the evidence becomes

$$Z = \int_0^1 L(\xi)d\xi \quad (19)$$

and the cumulative form of numerical integration in Eq. (15) can be expressed as

$$Z = \sum_{i=1}^K L_i \Delta \xi_i = \sum_{i=1}^K L_i (\xi_{i-1} - \xi_i) \quad (20)$$

where $0 < \xi_K < \dots < \xi_2 < \xi_1 < 1$, $\xi_0 = 1$ and K is the total number of samples taken. For practical implementations, the prior mass will tend to shrink exponentially by one part in Q at each iteration for a population of Q samples, leading to (Skilling 2006)

$$\xi_k = \exp\left(-\frac{k}{Q}\right) \quad (21)$$

Fig. 2 displays one sequence sampled from the experimental data. The likelihood value from each

sample is recorded when the parameters change under the constraint that each accepted parameter increases the likelihood value of the sample. The curve grows monotonically and may be considered to be completed when following increase can be ignored as the sample population converges on the most likely parameter values. Here, we choose the appropriate number of iterations, K .

4.2 Implementation

The main steps in the implementation of the nested sampling are summarized as follows:

1. Identify a model for evaluation and generate a population, Q , of sample models with random parameter values according to the assigned prior distribution; in this case, $Q = 100$.
2. Calculate the likelihood values of each sample with Eq.(13) inside the Q populations.
3. Store the lowest likelihood value.
4. Randomly perturb a parameter of the lowest-likelihood-sample and reevaluate its likelihood value. If the new sample has a high likelihood value, replace the previous least-likely sample by the new sample. If not, repeat this step until its likelihood exceeds the lowest value in the population.
5. Repeat steps (3) and (4) until the number of iterations is met.
6. Use the recorded likelihood values to calculate the evidence according to Eq. (20).
7. Repeat steps (1)–(6) for all competing models and select which model best represents the data in the investigation. This is the model selection step.
8. Select the parameters from the sample with maximum likelihood in the final population of the selected model. This is the parameter estimation step.

5 Numerical Simulation

In this section, numerical simulations are introduced to illustrate the performance of the Bayesian analysis. Model selection is used to estimate the number of leaks. After model selection, each leak location is also extracted from given the selected model. The multiple-leak cases with two leaks and three leaks are respectively considered.

Fig. 2 An example of sampling results for leaks at 300m and 1200m. Natural log of the likelihood is plotted against the sample number.

5.1 Numerical Setup

The setup of the considered pipe system is shown in Fig. 1. The pipe is connected to two reservoirs; the heads of the upstream and downstream reservoirs are $H_1 = 25m$ and $H_2 = 20m$, respectively. The pipe length is $l = 2,000m$ and the wave speed is $a = 1,000m$. The Darcy-Weisbach friction factor is $f = 0.02$ and the pipe diameter is $d = 0.5m$. The steady-state discharge $Q_0 = \sqrt{2gdA^2(H_1 - H_2)/(lf)} = 0.0153m^3/s$. A valve is just set at the downstream of the pipeline and its role is to generate the desired transient wave. Assuming that an impulse wave is generated by rapidly closing and opening the valve, the boundary conditions $h(x_U) = 0$ and $q(x_D) = 1$ are given. The given dataset in the frequency domain is accomplished by using the transfer matrix method in Eq. (1). Here, there are three pressure measurement stations located at $x_{M1} = 2,000m$, $x_{M2} = 1,800m$ and $x_{M3} = 1,600m$, respectively. Another pressure sensor at $x = x_{M0} = 50m$ is used to estimated $q(x_U)$ according to Eq. (7). The resonant and antiresonant frequencies $\omega = \{(1 + \alpha)\omega_{th}, \alpha = 0, 0.02, 0.04, \dots, 25\}$ are used, where $\omega_{th} = \pi/2l$. The dimension of the dataset is $JM = 3603$. Different noise levels are considered to study the performance of the proposed method. Zero-mean independent and identically distributed Gaussian white noise is added to all the pressure sensors. The noise level is divided by signal-to-noise ratio in decibel, which is defined as

$$SNR = 20 \log_{10} \left(\frac{|\mathbf{E}(\Delta \mathbf{h})|}{\sigma} \right) \quad (22)$$

where $\overline{E(\Delta h)}$ denotes the average head difference and σ represents the standard deviation of the Gaussian white noise.

5.2 Two-Leak Example

Here, a numerical example in Li et al. (2020) is revisited. A single pipe which is connected by two reservoirs in a horizontal plane is considered. The configuration is described in the above section. Fig. 3 shows the output function of multiple signal classification (MUSIC)-like algorithm with two-leak assumption, i.e., two leaks at $x_{L1} = 300m$ and $x_{L2} = 1,200m$ with actual sizes $s_{L1} = 1.0 \times 10^{-4} m^2$ and $s_{L2} = 1.2 \times 10^{-4} m^2$. The SNR is set to 10dB. In this case, the output function reaches maximum at the actual leak position. It is clear that there is a local maximum near each actual leak, but there is a higher lobes at around 1,500m, which will prevent obtaining correct estimates of leak locations. However, the method can locate two leaks in other locations (e.g., two leaks at $x_{L1} = 600m$ and $x_{L2} = 1,200m$). Therefore, it can be seen that the location accuracy of two-leak case is related to the location of the leak from the results.

In the following, this section presents simulation results to demonstrate the outcomes of the Bayesian analysis method. The conditions used are the same as the previous case. The evidence results are shown as a line chart in order to compare models of differing number of leak sources side-by-side. Each model has a different number of sound sources. The model selection results for 2-leak case with leaks at $x_{L1} = 300m$ and $x_{L2} = 1,200m$ with actual sizes $s_{L1} = 1.0 \times 10^{-4} m^2$ and $s_{L2} = 1.2 \times 10^{-4} m^2$ are shown in Fig. 4. The evidence for each model is evaluated over 10 individual runs using nested sampling. Fig. 4 illustrates the log-evidence estimations over different models for the number of leaks is 1, 2, 3, and 4. It is observed that the position of the maximum indicates that there is most evidence for two leaks. Note that, there is a fall-off from the maximum on the left side, because there is not enough structure in the proposed model to adequately illustrate the data; there is a decrease on the right side, as the models become increasingly, and unnecessarily, complicated. According to the results presented in Fig. 4, one come to the conclusion that the experimental data prefer the model, $\mathbf{M}_2$, indicating the presence of two leaks. After model selection,

select the parameters from the sample with maximum likelihood in the final population of the two-leak model as the actual leak locations. The results of leak locations are displayed in Fig. 5, wherein the solid lines and circles represent the actual and estimated leak locations, respectively. It is clear that the parameter estimation process can accurately localize the leaks. Two estimates are close to the actual leaks, with locations $\hat{x}_{l_1} = 300.16m$ and $\hat{x}_{l_2} = 1199.75m$. Since the results are affected by random error, in order to observe the statistical properties of the leak localization results, the root-mean-square error(RMSE) of each leak location estimation with respect to various SNR is plotted in Fig. 6. Here, the RMSE is defined as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_{L_i} - x_L)^2} \quad (23)$$

Fig. 3 Localization of double leaks using MUSIC-Like by plotting the spatial power spectrum. The actual leak positions are 300m and 1,200m.

Fig. 4 Log evidence among competing models for the 2-leaks case (300m and 1,200m)

Fig. 5 Leak localization using Bayesian analysis, the solid lines and circles represent the actual and estimated leaks, respectively.

in which x_L is the actual leak location and $\hat{x}_{L_i}$ represents the estimated value of the i th trial. When the number of leaks is determined, each simulation is repeated 10 times, i.e., $N = 10$. The SNR is varied from -15dB to 25dB. The results clearly shows that as SNR increases, the localization error (RMSE from 10 simulations) of each leak location decreases and the error is within the acceptable range.

Next, the other two sets of results are also considered to demonstrate the performance of the proposed method. Two sets of experimental data were obtained in two cases, the case with leak locations are $x_{L1} = 600m$ and $x_{L2} = 1,200m$ with actual sizes $s_{L1} = s_{L2} = 1.2 \times 10^{-4} m^2$ and the case with two close leaks whose coordinates are $x_{L1} = 1,000m$ and $x_{L2} = 1,040m$ with actual sizes $s_{L1} = 1.0 \times 10^{-4} m^2$ and $s_{L2} = 1.2 \times 10^{-4} m^2$. According to the results of model selection in Fig. 7a and Fig. 7b, it is determined that the given experimental data prefer the model M_2 , indicating that there are two leaks. Fig. 7c and Fig. 7d respectively plot the localization results of two cases. It can be seen that the proposed method can accurately localize the two leaks. In summary, this method can locate any two leaks. Fig. 7e displays the RMSE of leak localization for the former case. Fig. 7f shows the leak localization results with respect to various SNR for the latter case. As indicated in the above two figures, the estimation error increases as SNR decreases for each leak location. By comparing Fig. 7e and Fig. 7f, it is observed that for the former case, the estimation error is smaller than the latter case for the low SNR environment. For a low SNR being -15dB, reducing the distance between the two leaks increases the localization error.

Fig. 6 RMSE(m) of each leak location for $x^L = 300m$ and $1,200m$

(a)

(b)

(c)

(d)

(e)

(f)

Fig. 7 (a and b) Log evidence among competing models for the 2-leaks case; (c and d) leak localization using Bayesian analysis; and (e and f) RMSE(m) of each leak location

5.3 Three-Leak Example

In this section, a more complicated case with three leak is considered. The leak locations are $x_{L1} = 400m$, $x_{L2} = 1,000m$ and $x_{L3} = 1,400m$; the leak sizes are $s_{L1} = 1.4 \times 10^{-4}m^2$, $s_{L2} = 1.4 \times 10^{-4}m^2$ and $s_{L3} = 1.2 \times 10^{-4}m^2$. Other configuration parameters are the same as the previous section. In the following, the Bayesian inference is justified to be able to select the model that most appropriately represents experimental data and identify the leak locations.

Fig. 8 shows the localization results using the MUSIC-Like method. The SNR is set to -10dB. There are higher side lobes at around 200m and 600m, which may disturb the determination of leak locations. Other smaller lobes may be mistakenly identified as leaks especially when the number of leaks is unknown. Then, the Bayesian inference is used to detect the leaks. Firstly, the evidence obtained by nested sampling is used for model selection in order to determine the number of leaks. Fig. 9 illustrates the log-evidence with respect to various models for the number of leaks is 1, 2, 3, and 4. The highest evidence is at the case of three leaks. The position of the maximum indicates that there is most evidence for three leaks. According to the results illustrated in Fig. 9, the following conclusions can be drawn: the experimental data prefer the model, $\mathbf{M}_3$, indicating the presence of three leaks. After the selection of the three-leak model, adopt the parameters from the sample with maximum likelihood in the final population of the selected model as the actual leak locations. Note that the each estimated leak location is close to its actual value, with locations $\hat{x}_{l_1} = 416.1376m$, $\hat{x}_{l_2} = 1002.5304m$ and $\hat{x}_{l_3} = 1401.9674m$. The left half of Table 1 lists the RMSE of leak location with various noise level. It is clear that as the SNR decreases, the RMSE(obtained

Fig. 8 Localization of double leaks using MUSIC-Like by plotting the spatial power spectrum. The actual

leak positions are 400m 1,000m and 1,400m.

from 10 simulations) of each leak increases. However, the average localization error of each leakage exceeds $200m$ for a low SNR being -15 dB.

In the following, the three-leak case including close leaks is also considered. The leak locations are $x_{L1} = 1,000m$, $x_{L2} = 1,040m$ and $x_{L3} = 1,600m$; the leak sizes are $s_{L1} = 1.4 \times 10^{-4}m^2$, $s_{L2} = 1.4 \times 10^{-4}m^2$ and $s_{L3} = 1.2 \times 10^{-4}m^2$. Fig. 10 displays the log-evidence estimations over the different models. The results clearly implies that the model, $\mathbf{M}_3$, most appropriately represents the experimental data, indicating the presence of three leaks. Similarly to the previous case, each leak location estimation is close to its actual value and the estimated values are respectively $\hat{x}_{l_1} = 999.7778m$, $\hat{x}_{l_2} = 1040.2151m$ and $\hat{x}_{l_3} = 1606.5884m$. In addition, The right half of Table 1 lists the localization error with various SNR. It is observed that the localization performance improves in high SNR environment.

6 Conclusions

In this paper, the problem of detecting multiple leaks by using transient wave theory for a reservoir pipeline valve system is considered. A probabilistic method of model-based Bayesian inference is applied to detect leaks. The transient wave model deduced by the momentum equation and continuity equation, is used to describe various scenes that is defined by the combinations of a specific number of leaks and their locations. The Bayesian inference using nested sampling is utilized to first

Fig. 9 Log evidence among competing models for the 3-leaks case ($400m$, $1,000m$ and $1,400m$)

Fig. 10 Log evidence among competing models for the 3-leaks case ($1,000m$, $1,040m$ and $1,600m$)

Table 1 RMSE(m) for the case with leaks at $x^L = 400m$ 1,000m and 1,400m and the case with leaks at $x^L = 1,000m$ 1,040m and 1,600m

SNR[dB]	RMSE[m]			RMSE[m]		
	first leak	second leak	third leak	first leak	second leak	third leak
	[m]	[m]	[m]	[m]	[m]	[m]
-15	291.8471	293.6523	207.7569	475.9810	26.8399	363.8399
-10	19.5528	13.9665	13.8780	14.9233	7.6296	15.3998
-5	6.0114	11.2823	10.2427	8.9075	7.0047	7.1330
0	7.1526	6.9057	6.1713	6.0928	8.0747	4.0017
5	3.7026	4.1099	4.1854	4.6097	3.1630	2.3346
10	3.4821	4.7112	2.1194	3.7241	4.1255	2.0714
15	2.1246	2.2458	1.4416	4.5530	5.0753	2.1612
20	2.4197	1.6683	1.8293	4.8596	4.6408	2.5016
25	2.2231	2.3906	2.0619	3.0120	4.1739	3.3174

select which model the given experimental data prefer from competing models, i.e., determine the number of leaks, and then estimate the model parameters, including the particular leak locations.

The technique has been validated by two numerical studies with two leaks and three leaks, respectively. The information about pipeline leakages can be identified from the results for both cases. The number of leaks firstly is determined through the model selection process, which uses the Bayesian evidence calculated by nested sampling for each model to determine which model best fits the data under investigation. Then, adopt the parameters that maximize the likelihood value in the final population of the selected model as leak location estimations. The results for two-leak case illustrate that the localization efficiency of Bayesian inference is independent of leak locations and the localization error is acceptable even in noisy environment, which solves the defect that the localization ability of the MUSIC-Like method is limited by leak locations. Besides, the proposed method for the case of three leaks situation is investigated. It is also able to determine the number and locations of leaks even for the cases including

close leaks. However, the MUSIC-Like method does not work for three-leak case. Through the analysis of simulation results, Bayesian inference can not only determine the number of leaks, but also identify the location of each leak. To better apply transient-based leak detection method to practical applications, future work requires efforts to study various uncertainties that could affect the leak enumeration and localization, such as imprecise measurement of friction factor, wave speed and steady-state discharge.

Acknowledgements

This paper is supported by the key Science Foundation of the Department of Science and Technology of Jilin Province (Grant No. 20180201081SF, 20190303082SF), science and technology project of The Education Department of Jilin Province (Grant No. JJKH20200983KJ), and the Fund project of The Science and Technology Department of Jilin Province (Grant No. 20200201046JC). Thanks for the permission to publish this paper.

Authors Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Juan Li, Ying Wu and Changgang Lu. The first draft of the manuscript was written by Ying Wu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Competing Interests

None.

Availability of data and materials

All authors make sure that all data and materials as well as software application or custom code support the published claims and comply with field standards.

References

Beaton D, Xiang N (2017) Room acoustic modal analysis using Bayesian inference. *Journal of the Acoustical Society of America* 141(6):4480–4493

- Bush D, Xiang N (2018) A model-based Bayesian framework for sound source enumeration and direction of arrival estimation using a coprime microphone array. *Journal of the Acoustical Society of America* 143(6):3934–3945
- Chaudhry MH (2014) *Applied hydraulic transients*. 3rd ed. New York: Springer.
- Colombo AF, Karney BW (2002) Energy and costs of leaky pipes: toward a comprehensive picture. *Journal of Water Resources Planning and Management, ASCE* 128(6):441–450
- Covas D, Ramos H, Graham N, Maksimovic C (2004) Application of hydraulic transients for leak detection in water supply systems. *Water Science and Technology, IWA* 4(5-6):365–374
- Del Teso R, Gómez E, Estruch-Juan E, Cabrera E (2019) Topographic energy management in water distribution systems. *Water Resources Management, EWRA* 33(12):4385-4400
- Duan Huanfeng (2016a) Transient frequency response based leak detection in water supply pipeline systems with branched and looped junctions. *Journal of Hydroinformatic, IWA* 19(1):17-30
- Duan Huanfeng (2018) Accuracy and sensitivity evaluation of TFR method for leak detection in multiple-pipeline water supply systems. *Water Resources Management, EWRA* 32(6): 2147-2164
- Duan Huanfeng, Che Tongchuan, Lee PJ, Ghidaoui MS (2018) Influence of nonlinear turbulent friction on the system frequency response in transient pipe flow modelling and analysis. *Journal of Hydraulic Research, IAHR* 56(4): 451-463
- Escolano J, Xiang N, Perez-Lorenzo JM, Cobos M, Lopez JJ (2014) A Bayesian direction-of-arrival model for an undetermined number of sources using a two-microphone array. *Journal of the Acoustical Society of America* 135(2): 742-753
- Gupta Aditya, Kulat KD (2018) A selective literature review on leak management techniques for water distribution system. *Water Resources Management, EWRA* 32:23
- Kim S (2016) Impedance method for abnormality detection of a branched pipeline system. *Water Resources Management, EWRA* 30(3):1101–1115
- Knuth KH, Habeck M, Malakare NK, Mubeen AM, Placek B (2015) Bayesian evidence and model selection. *Digital Signal Process.* 47:50–67.
- Landschoot CR, Xiang N (2019) Model-based Bayesian direction of arrival analysis for sound sources using a spherical microphone array. *Journal of the Acoustical Society of America* 146(6): 4936-4946

- Li Juan, Wu Ying, Lu Changgang (2020) Pipeline leak detection using the multiple signal classification-like method. *Journal of Hydroinformatics*, IWA 22(5):1321-1337
- Li Juan, Zheng Qiang, Qian Zhihong, Yang Xiaoping (2019) A novel location algorithm for pipeline leakage based on the attenuation of negative pressure wave. *Process Safety and Environmental Protection* 123: 309-316
- Sattar AM, Chaudhry MH (2008) Leak detection in pipelines by frequency response method. *Journal of Hydraulic Research*, IAHR 46(EI1):138–151
- Sivia DS, Skilling J (2006) *Data analysis: A Bayesian tutorial* second edition
- Skilling J (2004) Nested sampling. *AIP Conference Proceedings*. 735:395-405
- Skilling J (2006) Nested sampling for general Bayesian computation. *Bayesian Analysis* 1(4): 833-859
- Soares AK, Covas DIC, Reis LFR (2011) Leak detection by inverse transient analysis in an experimental PVC pipe system. *Journal of Hydroinformatic*, IWA 13 (2):153–166
- Stephens ML, Lambert MF, Simpson AR (2013). Determining the internal wall condition of a water pipeline in the field using an inverse transient. *Journal of Hydraulic Engineering*, ASCE 139(3): 310-324
- Sun JL, Wang RH, Duan HF (2016) Multiple-fault detection in water pipelines using transient-based time-frequency analysis. *Journal of Hydroinformatic*, IWA 18(6):975-989
- Vitkovsky JP, Lambert MF, Simpson AR, Liggett JA (2007) Experimental observation and analysis of inverse transients for pipeline leak detection. *Journal of Water Resources Planning and Management*, ASCE 133(6):519-530
- Wang XJ, Lambert MF, Simpson AR, Liggett JA, Vitkovsky JP (2002) Leak detection in pipelines using the damping of fluid transients. *Journal of Hydraulic Engineering*, ASCE 128(7):697-711
- Wang X, Ghidaoui, MS (2018) Pipeline leak detection using the matched-field processing method. *Journal of Hydraulic Engineering*, ASCE 144(6):04018030
- Wang X, Palomar DP, Zhao Licheng, Ghidaoui MS, Murch RD (2019) Spectral-based methods for pipeline leakage localization. *Journal of Hydraulic Engineering*, ASCE 145(3):04018089

Figures

Figure 1

Setup of the considered pipeline system

Figure 2

An example of sampling results for leaks at 300m and 1200m. Natural log of the likelihood is plotted against the sample number.

Figure 3

Localization of double leaks using MUSIC-Like by plotting the spatial power spectrum. The actual leak positions are 300m and 1,200m.

Figure 4

Log evidence among competing models for the 2-leaks case (300m and 1,200m)

Figure 5

Leak localization using Bayesian analysis, the solid lines and circles represent the actual and estimated leaks, respectively.

Figure 6

RMSE(m) of each leak location for $x_L = 300\text{m}$ and $1,200\text{m}$

Figure 7

(a and b) Log evidence among competing models for the 2-leaks case; (c and d) leak localization using Bayesian analysis; and (e and f) RMSE(m) of each leak location

Figure 8

Localization of double leaks using MUSIC-Like by plotting the spatial power spectrum. The actual leak positions are 400m 1,000m and 1,400m.

Figure 9

Log evidence among competing models for the 3-leaks case (400m, 1,000m and 1,400m)

Figure 10

Log evidence among competing models for the 3-leaks case (1,000m, 1,040m and 1,600m)