Multi-stage parameter-constraining inverse transient analysis for pipeline condition assessment

Chi Zhang, Aaron C. Zecchin, Martin F. Lambert, Jinzhe Gong and Angus R. Simpson

ABSTRACT

Fault detection in water distribution systems is of critical importance for water authorities to maintain pipeline assets effectively. This paper develops an improved inverse transient analysis (ITA) method for the condition assessment of water transmission pipelines. For long transmission pipelines ITA approaches involve models using hundreds of discretized pipe reaches (therefore hundreds of model parameters). As such, these methods struggle to accurately and uniquely determine the many parameter values, despite achieving a very good fit between the model predictions and measured pressure responses. In order to improve the parameter estimation accuracy of ITA applied to these high dimensional problems, a multi-stage parameter-constraining ITA approach for pipeline condition assessment is proposed. The proposed algorithm involves the staged constraining of the parameter search-space to focus the inverse analysis on pipeline sections that have a higher likelihood of being in an anomalous state. The proposed method is verified by numerical simulations, where the results confirm that the parameters estimated by the proposed method are more accurate than the conventional ITA. The proposed method is also verified by a field case study. Results show that anomalies detected by the proposed methods are generally consistent with anomalies detected by ultrasonic measurement of pipe wall thickness.

Key words | condition assessment, identifiability, inverse transient analysis, water pipelines

INTRODUCTION

Model calibration is the process of determining the model parameters to obtain a model representation of the system of interest that satisfies a certain criterion, i.e. a goodness of fit metric between the system measurements and the simulated model (Savic et al. 2009). inverse transient analysis (ITA) (Liggett & Chen 1994), one of the popular transient-based methods (Meniconi et al. 2013; Gong et al. 2014; Massari et al. 2014; Capponi et al. 2017; Duan 2017), is a process of pipe system parameter estimation using measured transient pressure data and a hydraulic transient model. The inverse analysis is achieved by adjusting pipeline system parameters (such as leak locations and sizes, friction factors, water demands or hydraulic transient wave speeds) so that the calibrated model's predicted response matches the measured response (typically the temporal measurements of pressure). In the case of hydraulic transient wave speeds, the calibrated parameter is closely related to the pipeline condition (Stephens et al. 2013), as the wave speed is related to the wall thickness and internal diameter of the pipe, which change as the pipe wall corrodes. As such, a well calibrated model provides valuable information about the pipeline condition, which serves as guidance for water authorities to assess infrastructure condition, and maintain assets strategically.

Despite the fact that ITA has been widely studied, the scale of the problems that have been considered (i.e. the
number of system parameters) is relatively small. ITA was first proposed by Liggett & Chen (1994) for leakage and friction factor calibration. Within their numerical study, 15 parameters, including 11 friction factors and four leak sizes, were calibrated for a small 11-pipe network. Jung & Karney (2008) considered a numerical case study, where the wave speed, diameter and friction factor of 11 pipes, four possible leakages and one water demand were calibrated simultaneously (38 parameters overall), on the same network as Liggett & Chen (1994). The average error of the parameter estimates was, in general, greater than 30% and the authors claimed the error associated with estimated wave speeds tends to compromise the calibration of other parameters.

Shamloo & Haghighi (2010) also conducted ITA on the same network to detect potential leaks. The absolute errors of the 17 parameter estimates (six leak sizes and 11 friction factors) were all below 5% by optimizing the transient generation and the choice of measurement site.

Concerning the laboratory verification of ITA, Vitkovsky et al. (2001, 2007) calibrated the leak size for a laboratory system consisting of a 37.2 m long copper pipe with an internal diameter of 22.1 mm. A total of 15 leak candidate locations were considered where the measured pressure trace contained multiple periods of the transient wave in the pipe system. Covas & Ramos (2010) applied ITA to a 272 m high-density polyethylene pipe with an internal diameter of 50.6 mm and a 1.3 km polyethylene pipe with an internal diameter of 70 mm. Multiple models were considered, where the number of calibrated parameters was less than 30 for each model. Soares et al. (2011) applied ITA to a system composed of looped PVC pipes with a total length of 203.20 m. The potential leak locations in the inverse model was 7. It is worth pointing out that both Covas & Ramos (2010) and Soares et al. (2011) adopted a strategy to minimize the number of leak candidates, starting with a set of leak candidates sparsely distributed throughout the system, and gradually reducing the set of candidates to those around the potential leak locations obtained in the earlier steps.

ITA in the frequency domain has also attracted attention. Kim (2008) developed an address-oriented impedance matrix method for the generic calibration of several parameters, such as the location and quantity of leakage, the friction factor, and the wave speed simultaneously. The method was numerically verified by a 12-pipe network. Kim (2008) also adopted a similar multi-stage strategy to Covas & Ramos (2010), starting the calibration with the initial search-space, which was the accumulated length of all of the pipeline elements, and then restraining the search-space to a candidate pipeline element. Zecchin et al. (2013) developed a frequency-domain maximum likelihood estimation method to identify parameters of fluid line networks, in which different types of parameters including wave speed, diameter, and length of pipe were calibrated simultaneously. The proposed method was applied on an 11-pipe network using turbulent steady friction and turbulent unsteady friction pipe models, which resulted in parameter estimation problems of dimension 33 and 55, respectively. Compared with time domain ITA, there are fewer laboratory or field verification works for frequency domain applications of ITA. Laboratory verification of the frequency domain ITA is found in the research by Kim et al. (2014), where location and discharge of a single leak, together with other parameters such as wave speed, were calibrated. Capponi et al. (2017) also successfully located and sized one leak, one blockage and one branch with high accuracy in a controlled laboratory environment.

As outlined above, the complexity of the inverse problem (i.e. the dimensionality of the parameter estimation space) is limited in most previous research. An exception is the work by Stephens et al. (2013), where ITA was applied to a mild steel cement-lined transmission mains in rural South Australia. In this work, the wave speeds of 390 reaches, along a 6 km section of pipe, were calibrated to infer internal pipe wall condition. The pattern of estimated wave speeds by ITA were generally consistent with ultrasonic measurements taken along the pipe, although this was not always the case. However, within this application, Stephens et al. (2013) claimed that it was almost certain the optimal solution was not found, due to a limitation on the number of objective function evaluations used by the optimizer (given the high computational cost of overall millions evaluations) and model error. Additionally, the authors also believed that, given the high dimensionality of the problem, it was very likely that non-unique solutions existed, meaning that there were very many local optimum solutions with similar objective function values. However, this issue was not explored in detail in their work.
Non-uniqueness of solutions for ITA applications has been observed in other work, and is a known problem for inverse problems more generally (Yeh 1986). Within the study by Jung & Karney (2008), three different ITA scenarios were analyzed on an 11-pipe network, where each had a different number of measurement sites. The authors concluded that fewer measurement sites led to a better calibration error and faster convergence to the calibrated parameter values, but to a less accurate calibration result. Kapelan et al. (2004) pointed out that some poorly estimated parameters were due to an inadequate quantity of observed information. That is, Kapelan et al. (2004) outlined that the parameter calibration problems could not be resolved by simply using an alternative search algorithm. To improve the ill-posed natural of the inverse problem, a framework for the effective incorporation of prior information into ITA for pipe networks was developed. However, these two approaches to enhance the identifiability are not always feasible in field tests, where access points to install pressure transducers are limited, and prior information is not always available. Instead of increasing the number of measurement sites or incorporating prior knowledge of a system, another strategy to enhance parameter identifiability of ITA is to reduce the complexity of the inverse problem by limiting the number of decisions (Kim 2008; Covas & Ramos 2010) or limiting the range of the search-space.

The interval for a parameter’s search-space is typically the minimum and maximum feasible values of the parameter to be calibrated, and they are typically the same for every reach. However, in the case of wave speed calibration in ITA, typically sections of pipe in a normal state (e.g. sections in a generally good condition) have wave speed values within a small band about a nominal value, and it is only anomalous sections (e.g. sections with severe deterioration) that possess significant uncertainty and variability in the wave speed values they can take. Therefore, if the search-space can be properly adjusted for each reach to account for the reduced uncertainty for normal sections of pipe, better parameter estimates can be expected within the limited search-space.

To enhance parameter identifiability of ITA on a complex transmission main, a multi-stage parameter-constraining ITA is proposed in this paper. This paper is organized as follows: to highlight lack of identifiability achieved by the conventional ITA on a complex transmission mains system, a numerical example with a highly variable pipe wall condition is considered (represented by hundreds of reaches, each with different wave speeds). It is confirmed within this example that the conventional ITA is not able to uniquely identify the correct wave speed values, despite achieving parameter estimations that provide a very good fit of the predicted and measured pressure responses. To tackle this problem, a multi-stage parameter-constraining ITA is proposed. The proposed algorithm works by iteratively generating sets of parameter estimates and coincidentally modifying the search-space interval for the subsequent estimates. Based on a statistical analysis of the set of estimates, each reach is classified as normal (parameter values close to the nominal values) or anomalous (parameter values significantly different from the nominal values). Different search-space intervals are allocated to each reach based on the outcome of the classification. Parameter estimation within the properly narrowed search-space will lead to concentration of parameter estimates and elimination of outliers, thus identifiability of parameters by ITA is enhanced. This algorithm is followed by a numerical verification and discussion about the sensitivity to the relevant parameters. Finally, the method is verified by a field case study.

BACKGROUND: LACK OF IDENTIFIABILITY WITHIN ITA APPLICATIONS

The problem of lack of identifiability for applications involving the conventional ITA method are explored within this section with a numerical example application on a reservoir-pipeline-valve (RPV) system. Within this section, the numerical example illustrates that when applying the conventional ITA to a relatively complex (yet realistic) inverse problem, the identifiability of the wave speed parameters is relatively poor. This is evidenced by the fact that independent inverse runs (using a stochastic or evolutionary optimization approach with different random number sequences) yielded a large variability in the optimal parameter estimates, despite the high similarity in the estimates’ associated calibration error. The data of the system are described in the next sub-section, followed by a
discussion, where it is demonstrated that the conventional ITA suffers from the problem of lack of identifiability.

**Numerical example**

A typical RPV system (Figure 1) is used to simulate a transmission main throughout this paper. The wave speed distribution along the pipe is depicted in Figure 2. Wave speeds of 115 normal reaches are in the range of 950–1,000 m/s. A total of 13 reaches (nine sections in Figure 2), whose wave speeds are either less than 950 m/s or greater than 1,000 m/s, are used to represent different wall thickness pipe sections due to different pipe classes or wall condition. A transient wave is generated by instantaneously shutting the side discharge valve at node G when \( t = 0.2 \) s. The pressure responses at locations M1, M2 and G are obtained by numerical simulation using the method of characteristics (MOC) in which the time step \( \Delta t = 0.01 \) s.

The length of each reach was determined by \( \Delta x_i = a_i \Delta t \) to satisfy Courant condition. This avoids interpolation errors in the forward model when synthetically generating the ‘measured’ pressure traces. Only a duration of 1.2 s after the transient generation is considered in the ITA to avoid reflections from the boundaries of the RPV system.

The ITA used in this paper is the efficient version proposed by Zhang et al. (2018), in which the head based method of characteristics (HBMOC) is employed as the transient simulator. Particle swarm optimization (PSO) (Poli et al. 2007) was employed as the optimization algorithm – to account for the stochastic nature of PSO, the ITA is run multiple times (with different starting random number seeds) to provide a statistical characterization of the algorithm’s performance. The objective function is defined as the squared difference between the measured responses and the predicted responses by the calibrated model summed over all measurement stations. A time step of 0.01 s was used in the ITA. Considering the 1.2 s pressure trace used in ITA and the location of measurement sites, an inverse model with 128 reaches is required, to accommodate all reaches contributing to the 1.2 s pressure trace. As such, the number of decision variables is 128, and for PSO, the number of particles is 150 with 10,000 iterations (resulting in 1,500,000 evaluations overall), which were selected so as to allow for convergence of the algorithm’s search. The search-space was set to be [800 m/s, 1,050 m/s], and the decision variables were treated as continuous within this range by the PSO.

**Example results and discussion**

Figure 3 presents a series of boxplots for the wave speeds estimated in each reach by ten independent runs. It can be seen that the interquartile range (IQR) of the wave speeds for each reach are generally large, which means that the wave speed estimates of the independent runs are spread diversely throughout the wide search-space (i.e. starting from different random number seeds, the PSO ended up with significantly different solutions). This high variability amongst the parameter estimates limits a clear interpretation of the results, and the ability to use these estimates for the detection of the anomalous sections.

To understand why the conventional ITA parameter estimates in Figure 3 yield such high variance, the envelopes of ten predicted pressure traces and the match between measured pressure trace and predicted pressure trace obtained by the best solution from the ten runs (ranked in terms of objective function) are given in Figure 4. Despite the existence of outliers in the estimates (O1–O4 in Figure 3),

![Figure 1](https://iwaponline.com/jh/article-pdf/20/2/281/198352/jh0200281.pdf)  
*Figure 1* | Numerical experiment pipeline configuration.
Figure 2 | True wave speeds of the numerical model.

Figure 3 | Boxplot of wave speeds estimated by ten independent runs, and the top two best estimated wave speed set (ranked in terms of objective function). O1–O8 are the distinct outliers existing in the top two best estimated wave speeds.
the best wave speed estimates solution still provides an excellent match between its predicted pressure traces and the measured pressure trace. The closely wrapped envelopes of predicted pressure traces indicate that reasonable match was achieved by every set of wave speed estimates. The fact that such an excellent model fit can be achieved for such vastly different parameter estimates means that the objective function is not sufficiently informative to allow for the discrimination between different parameter sets. These findings support the observations from Kapelan et al. (2004) concerning estimation issues arising from data information limitations for certain ITA applications.

**PROPOSED METHOD: MULTI-STAGE PARAMETER-CONSTRAINING ITA (ITA\textsubscript{MP})**

**Overview**

In the previous example, it has been illustrated that, for the conventional ITA, independent parameter estimates are widely spread throughout the search-space. Despite the reasonable match between the modelled and measured pressure traces, distinct outliers exist in the estimates, suggesting that multiple local optimums exist in the relatively large search-space. This limitation of the conventional ITA to correctly identify parameter estimates for large-scale problems has motivated the iterative parameter-constraining approach of the proposed method.

The motivation for the proposed method is outlined in the following. Transmission mains are typically installed with a similar design and are deteriorated in a similar way, unless the cumulative effect of soil or groundwater corrosion and external loaded forces increase degradation in some sections. Besides these severely deteriorated sections, undocumented class changes with different wave speeds and pipe sections replaced by plastic pipe section with lower wave speeds are also anomalous sections in transmission mains. Compared with the majority sections which are in similar conditions, these anomalous sections are localized to a limited number of short discrete sections along the pipeline. The wave speed parameter values for

![Figure 4](https://iwaponline.com/jh/article-pdf/20/2/281/198352/jh0200281.pdf)

*Figure 4* | Measured pressure trace, predicted pressure trace obtained by the best solution from the ten runs (ranked in terms of objective function) and the envelopes of ten predicted pressure traces.
the normal sections are confined to within a small interval about a nominal value, as there is much less uncertainty about the values that the normal sections can take. In contrast, the anomalous sections contain wave speed values that lie within a much broader interval, as there is significant uncertainty about the values that the anomalous sections can take. Therefore, it is expected that the effectiveness and accuracy of the ITA parameter estimation process would be increased if adaptive search-space parameter intervals were used to allow for a narrowing of the interval for sections that were recognized as normal, and maintaining a broader interval for sections that were identified as anomalous. Based on this idea, a multi-stage parameter-constraining ITA (ITAMP) method is proposed. In this approach, the search-space is iteratively limited, guiding the solutions closer to the true parameter values.

The ITAMP algorithm

The flowchart of the ITAMP algorithm is given in Figure 5. The algorithm comprises multiple iterative stages, and each stage involves the sub-stage of new solution generation, reach classification, search-space updating for the next stage and the termination criteria check. Details of the four sub-stages are presented in the following sub-sections.

Sub-stage 1: new solution matrix generation

The objective function used in the proposed ITA approach is defined by Equation (1):

$$OF(a) = \frac{1}{N_{MS}N_{TS}} \sum_{i=1}^{N_{MS}} \sum_{j=1}^{N_{TS}} \left( H_{ij}^m - H_{ij}(a) \right)^2$$  \hspace{1cm} (1)

where $N_{MS} = $ number of measurement stations (including the generator station if the transient pressure variation is also recorded there); $N_{TS} = $ number of time steps; $H_{ij}^m = $ measured pressure response at $i$th measurement station and $j$th time step; $H_{ij} = $ predicted pressure response at $i$th measurement station and $j$th time step from the calibrated model; $a = [a_1, ..., a_N]^T$ is the vector of unknown wave-speeds, and $a_k = $ wave speed of $k$th reach, where the search for the wave speeds is typically bounded as $a_k \in [a_{min,k}, a_{max,k}]$. The search-space in Stage 1 $[a_{min,1}, a_{max,1}]$ is determined by the maximum and minimum feasible parameter values. The search-space in Stage $n$ $[a_{min,n}, a_{max,n}]$ is determined in the Stage $n-1$ and is explained in the following section ‘search-space updating for the next stage’.

All the solutions from $M$ different independent ITA runs are saved in a solution matrix $A_n$, which is defined as:

$$A_n = \begin{bmatrix} a_{1,n}^T \\ \vdots \\ a_{M,n}^T \end{bmatrix} = \begin{bmatrix} a_{1,1,n} & \cdots & a_{1,N,n} \\ \vdots & \ddots & \vdots \\ a_{M,1,n} & \cdots & a_{M,N,n} \end{bmatrix}$$

where $a_{i,j,n} = $ $i$th wave speed estimation of the $j$th reach; $N = $ number of reaches; $M = $ number of independent ITA runs; $n = $ stage number.

Sub-stage 2: section classification

The solution matrix $A_n$ provides a data set of wave speed estimates for each reach. A fundamental assumption of the
proposed process is that the previous applications of ITA have been effective in their optimization strategy, and that the majority of estimates are, at least, within the neighborhood of the global optimum wave speed solution $a^\star$. Given this assumption, the classification of whether a reach is considered normal or anomalous is based on analyzing the data set for each reach to determine if the estimates are significantly different from the nominal properties of the entire set of wave speed estimates. That is, given the set of wave speed estimates $a_{k,n}$ for reach $k$, $\{a_{1,k,n}, \ldots, a_{M,k,n}\}$, the classification process aims to determine the likelihood that the statistical properties of this set are consistent with the statistical properties of the wave speed estimates for a normal section of pipe as characterized by the data set $A_{k,n}$ (the data set comprised of all elements of $A_k$ excluding column $k$, represented as the subscript $/k$). If the statistical properties of the $k$th reach are consistent with the properties of a normal section of pipe, then the reach is classified as normal; if not, then it is classified as anomalous.

Given the fact that the properties of the estimates of a normal section of pipe are unknown a priori, the data set $A_{k,n}$ comprised of the wave speed estimates for the rest of the pipeline are used to characterize a normal section. This approach to the characterization of the estimates for a normal section is justified on the basis that a fraction of reaches will be in an anomalous state (sections with severely deteriorated conditions, undocumented or replaced pipe sections with different material), and the majority reaches will be in the normal state (generally good condition or similarly deteriorated condition). This means that within this larger data set (all elements of $A_{k,n}$) the influence of estimates for the anomalous sections on the statistics of this set will be small compared to the influence of the normal sections. This is particularly the case when order statistics are used to characterize the set, as these are insensitive to outliers in a data set.

An appropriate statistical test to undertake the comparison of the $k$th reach data set $a_{k,n}$ with $A_{k,n}$ is the Mann–Whitney U test (Hollander et al. 2013). This test provides a non-parametric approach to compare two data sets to determine the likelihood that a random sample of one data set will be less than (or greater than) a random sample of the other data set. The null hypothesis of this test is that both sets are distributed about the same median value, meaning it is equally likely that a sample from one data set will be less than (or greater than) a sample from the other. A rejection of the null hypothesis means it is more likely that both sets are not distributed about the same median value. In the classification context of this paper, a reach is classified as normal if the null hypothesis holds, and is classified as anomalous if the null hypothesis is rejected. The acceptance or rejection of the null hypothesis is dependent on the comparison of the $p$-value for $k$th reach, $p_{h,n}$, returned by Mann–Whitney U, and the specified significance level $\alpha_n$. The null hypothesis is accepted if $p_{h,n} \geq \alpha_n$, and is rejected if $p_{h,n} < \alpha_n$.

The significance level $\alpha_n$ is a fundamental parameter of the proposed method. A physical interpretation of the significance level can be considered as follows: under the assumption that $m$ % of the pipe reaches will be classified as anomalous, the associated significance level is given by:

$$\alpha_n = m - \text{th percentile of } P_n. \quad (2)$$

where $P_n = \{p_{1,n}, \ldots, p_{M,n}\}$. The advantage of considering $m$ as the user specified parameter is its interpretation as providing an upper bound on the percentage of the pipe that is assumed to be in an anomalous state. It is therefore recommended that $m$ should be set greater than the assumed percentage of anomalous sections of the pipeline under investigation. The appropriate choice of $m$ will be investigated in the section on numerical verification. The outcome of this stage are the two sets: $C_{0,n}$ = the set of reaches that are classified as normal; and $C_{A,n}$ = the set of reaches classified as anomalous.

**Sub-stage 5: search-space updating for the next stage**

Different search-space intervals are allocated for different reaches in the next stage, depending on the set $C_{0,n}$ or $C_{A,n}$ that the reach belongs to. The detection of an anomalous section is the focus of the parameter estimation; as a result, a reach classified as anomalous ($a_{h,n}, k \in C_{A,n}$) will retain the original wide wave speed search interval $[a_{\min,1}, a_{\max,1}]$ so that the estimation strategy is still able to search within the full search-space.
A reach which is classified as normal \((a_{k,n}, k \in C_{0,n})\) will be allocated a narrower wave speed search interval \([a_{\text{min},n+1}, a_{\text{max},n+1}]\) in the next stage. The new search-space bounds \(a_{\text{min},n+1}, a_{\text{max},n+1}\) for a reach classified as normal are determined by percentile values of \(A_{n,0}\):

\[
a_{\text{min},n+1} = p - \text{th percentile of } A_{0,n} \tag{3}
\]

\[
a_{\text{max},n+1} = q - \text{th percentile of } A_{0,n} \tag{4}
\]

where \(q > p\); \(A_{0,n} = \{a_{i,n}, i \in C_{0,n}, j = 1, \ldots, M\}\).

\(A_{0,n}\) is the collection of multiple estimates for all reaches which are classified as normal, so the statistics of \(A_{0,n}\) can be used to represent the parametric range that the wave speed values of the normal reaches are within. Considering the likely existence of outliers in the estimates in \(A_{0,n}\), \(p\) and \(q\) percentile of \(A_{i,n}\) are used to determine the new search-space bounds. Compared with the original search-space \([a_{\text{min}}, a_{\text{max}}]\), the new search-space \([a_{\text{min},n+1}, a_{\text{max},n+1}]\) is a narrower interval. The choice of percentile rank \(p\) and \(q\) will be discussed in the section on numerical verification.

Sub-stage 4: termination criteria

The algorithm is terminated from iterating in Stage 2 when the \(M\) multiple solutions generated within the updated search-space at Stage \(n\) all have a larger objective function than the previous best objective function from iteration Stage \(n-1\). The algorithm aims to minimize the objective function, so if it fails to decrease the objective function, and thus fails to improve the match between the calibrated models and the measured data, there is sufficient reason to believe that a failure to find a better solution is attributed to an improper updated search-space, so the algorithm is terminated.

### NUMERICAL STUDY

A numerical experiment was conducted to verify the proposed ITAMP. The numerical model is the same as in Figure 1. The number of reaches is \(N = 128\); and the number of independent ITA runs for different starting random number seeds is \(M = 10\), where PSO was used as the optimization algorithm.

Nine cases were considered given different values of \(m, p\) and \(q\) (see Table 1). Case 3 \((m = 15, p = 5\) and \(q = 95\)) is discussed in detail in the section ‘Results and discussion’. The other cases are discussed in the section ‘Choice of \(m, p\) and \(q\) values’.

### Results and discussion

#### Detailed results for Case 3

For Case 3, the objective functions of all independent ITA runs are given in Figure 6. From this figure it is seen that the ITAMP terminated after Stage 3, making the total number of independent ITA runs equal to 30. The ten ITA runs in Stage 1 ended up with a diverse set of solutions due to the random nature of the optimization algorithm and the relatively large search-space, which is consistent with the diverse wave speed estimates in Figure 3. A significant reduction in the objective function value can be seen in Stage 2, after the first search-space update. Solutions in Stage 2 show more consistency in their objective function values within narrowed search space.

Figure 7 presents the boxplot of the wave speed estimates in Stage 2. Compared with the wave speed estimates in Stage 1 in Figure 3 (the result of Stage 1 is the result of the conventional ITA, since the conventional ITA is terminated after multiple independent runs is Stage 1), the IQR

| \(m\) | \(5\) | \(10\) | \(15\) | \(20\) | \(15\) | \(15\) | \(15\) | \(15\) | \(15\) |
| \(p\) | \(5\) | \(5\) | \(5\) | \(5\) | \(0\) | \(1\) | \(2\) | \(10\) | \(15\) |
| \(q\) | \(95\) | \(95\) | \(95\) | \(95\) | \(100\) | \(98\) | \(90\) | \(85\) |

*In Case 5, the search-space bound \(a_{\text{min}}\) and \(a_{\text{max}}\) are determined by the minimum and maximum of \(A_{n,0}\). For simplicity, the minimum and maximum of \(A_{n,0}\) are noted as 0th or 100th percentile of \(A_{n,0}\), which technically do not exist.*

---

"Downloaded from https://iwaponline.com/jh/article-pdf/20/2/281/198352/jh0200281.pdf by guest"
Figure 6 | Objective function of each independent run in all stages (ten runs at each stage).

Figure 7 | Boxplots of wave speed estimates in Stage 2.
of wave speed of each reach become significantly less. For the proposed approach, the wave speed estimates become increasingly concentrated around the true wave speeds of the model, making the interpretation of the results and detection of anomalous sections easier.

From Figure 7, it also can be seen that the outliers existing in the Stage 1 estimates (for example, O1–O8 in Figure 3) were not present in the estimates of Stage 2. Elimination of the outliers was due to the section classification, and consequent search-space updating, in ITAMP. Interestingly, these outliers from the single run in Stage 1 were identified as normal sections in Stage 2, because the estimates of the other independent runs yielded statistics consistent with the normal value. As a result, in the next stage, the search-space was properly narrowed and a better solution was found.

The enhanced identifiability of wave speed estimates also resulted in an improvement in the match of the measured and predicted pressure traces, for which the traces at the generator location are given in Figure 8. All independent wave speed estimates in Stage 2 provided an almost indistinguishable match between measured and predicted pressure traces, indicated by shrunk envelopes of predicted pressure traces in Stage 2, compared with those in Stage 1 in Figure 8.

Sensitivity to the percentile rank parameters m, p and q

The sensitivity of the ITAMP results to the percentile rank parameters m, p and q is discussed in this section. Three parameters m, p and q are percentile ranks in Equations (2)–(4) to determine the significance level in the Mann–Whitney U test and search-space bounds for next stage ITA runs in Stage 2.

As explained in the section ‘numerical experiment’, among 128 reaches of the inverse model, 13 reaches (or a total of nine sections) are anomalous, meaning that the proportion of anomalous reaches is 10.2%. It has been explained that the percentile rank m represents the proportion of reaches which will be identified as anomalous reaches within ITAMP. Case 1 (m = 5, p = 5, q = 95) is used as an example to illustrate the impact of a smaller m on the results. In Case 1, the ITAMP algorithm terminated after Stage 2. Figure 9 represents the single best wave speeds of Stage 2.

![Figure 8](https://iwaponline.com/jh/article-pdf/20/2/281/198352/jh0200281.pdf)
It can be seen in Figure 9 that the anomalous sections D1, D4, D5, D6, and D9 were successfully identified. However, anomalous sections D2, D3, D7, which are closer to the normal wave speeds (compared with those of D1, D4, D5, D6, and D9), were identified as normal sections because a smaller $m$ was used and only 5% of reaches were allowed to be identified as anomalous reaches. The wave speeds of these three sections reach their maximum or minimum values (the boundaries of the search-space for normal sections).

With regard to Cases 2–4 where the $m$% values are all greater than 10.2% (the true proportion of anomalous reaches), the boxplots of estimates in Cases 2–4 combined (that is 30 estimates for each reach) are given in Figure 10. Comparison between the boxplots in Case 3 alone (Figure 7) and Cases 2–4 combined demonstrates that as long as $m$% is greater than the proportion of anomalous sections, the choice of the percentile rank $m$ has little sensitivity to the wave speed calibration. In conclusion, it is recommended that $m$ is set to be greater than the assumed proportion of anomalous sections. In field applications, $m$ can be chosen based on the prior information, if available, or by a trial and error strategy.

A comparison of Cases 5–9 with Case 3 is made to explore the impact of variations in the percentile rank parameters $p$ and $q$. Figure 11 shows the search-space interval for normal sections of different stages for the different cases. In Case 5 ($p = 0$, $q = 100$), the search-space bound $a_{n,\text{min}}$ and $a_{n,\text{max}}$ for the normal sections are determined as the minimum and maximum of normal section matrix $A_{n,0}$. These Case 5 bounds are almost the same as the original search-space, and as a result, ITA Mp terminated after Stage 2 as no further improvement was achievable. This proves that limiting the search-space is the driving force behind the improvement of the wave speed estimation in ITA Mp.

As the percentile rank $p$ becomes greater and $q$ becomes smaller, the updated search-space becomes narrower, and ITA Mp was observed to terminate earlier. This is because for the increasingly narrow search-space, more normal reaches have wave speed values that violate the search-space bounds, yielding a reduced calibration.
Figure 10 | Boxplots of wave speed estimates in Cases 2–4.

Figure 11 | Search-space for identified normal section when different percentile ranks \( p \) and \( q \) were used.
performance. This side effect of ITA\textsubscript{MP} is that some normal reaches are not allowed to take their true values, if their wave speeds are out of the narrowed search-space range. However, from the perspective of the detection of anomalous sections, this side effect can be considered unimportant as the accurate calibration of the normal sections does not affect the anomalous section detection. Overall, the results from Cases 5–9 demonstrate that the estimated wave speeds, especially those of anomalous sections, are not sensitive to different values for the percentile ranks $p$ and $q$.

FIELD CASE STUDY

Preliminaries

A field case study on the Morgan–Whyalla transmission pipeline (MTP) in South Australia was conducted to verify the multi-stage parameter-constraining ITA (the reader is referred to Stephens \textit{et al.} (2015) and Gong \textit{et al.} (2015) for more details on the field case study). The pipeline is made of mild steel with cement mortar lining (MSCL), is in good condition with intact cement mortar lining, and has an

![Figure 12](https://iwaponline.com/jh/article-pdf/20/2/281/198352/jh0200281.pdf)

**Figure 12** | Configuration of the section of MTP (Note that CH refers to chainage, and sections A, B, C and D are four known thicker-walled sections).

![Figure 13](https://iwaponline.com/jh/article-pdf/20/2/281/198352/jh0200281.pdf)

**Figure 13** | Objective function of each independent run in all stages.
internal diameter of 727.5 mm. The configuration of the pipeline under investigation (from chainage 14,900 to 18,900 m) is given in Figure 12. For the series of field experiments, the transient waves were generated by shutting scour valve SC24 abruptly. Three transducers were installed at air valves AVFP43 and AVFP44, and scour valve SC24 to record the transient pressure response.

The measured pressure traces at AVFR43, SC24 and AVFR44, in which the underlying long-period pressure oscillations were removed by a low-pass filter, are given in Figures 16–18 (where they are compared with predicted pressure trace of the calibrated model). A time step of 0.0135 s and a duration of 4.4 s of the measured pressure traces were used in ITAMP to cover the 4 km pipe section of interest. This setup of the inverse problem resulted in the number of reaches in the inverse model being \( N = 383 \).

The number of independent ITA runs of different random number seeds in each sub-stage was set to be \( M = 10 \). In each independent ITA run, for the adopted optimization algorithm PSO, the number of particles and the number of maximum iterations were set to be 400 and 3,000, respectively. The percentile ranks were chosen as \( m = 20 \), \( p = 5 \) and \( q = 95 \).

Results

Objective function values of all independent ITA runs are given in Figure 13. Similar to Figure 6 (the objective function of the numerical study), a significant reduction is found when the search-space was updated for the first time. The ten objective function values became concentrated gradually up to Stage 4, when multiple ITA runs failed to find a better solution within the third updated search-space.

Figure 14(a) and 14(b) represent boxplots of the wave speed estimates in Stage 1 and Stage 3. Extensive ultrasonic wall thickness measurements were manually undertaken using an ultrasonic thickness measurement instrument at 5 m intervals along the MTP between chainages 14,900 and 18,900 m (Stephens et al. 2015). Measurements were taken at eight points around the circumference of the pipe at each location. All ultrasonic wall thickness measurements are given in Figure 14(c).

Figure 14 | (a) Boxplots of wave speed estimates in Stage 1; (b) boxplots of wave speed estimates in Stage 3; (c) boxplots of wall thickness ultrasonic measurements.
Four known thicker-walled sections of pipeline A, B, C and D, confirmed by the ultrasonic measurements, are captured by wave speed estimates in both Stage 1 and Stage 3. Other unknown anomalous sections (for example, section F) are also identified by ultrasonic measurements and wave speed estimates in both Stage 1 and 3. However, the IQR of Stage 3 is significantly less than that of Stage 1. Less variation in wave speed estimates enable less false positives identification and more accurate abnormal section detection. Section E, which is depicted in Figure 15, is used to demonstrate the benefits the ITA\textsubscript{MP} brings.

According to the ultrasonic measurements, a small fraction of the section from chainage 16,159.2 to 16,329.2 m (Xc in Figure 15(c)) is very likely to be in the anomalous state. It is identified by ITA results in both Stage 1 and Stage 2; they are Xa in Figure 15(a) and Xb in Figure 15(b), respectively. However, several sections in the dashed box in Figure 15(a) are likely to be interpreted as the anomalous sections, although they should not be. After ITA\textsubscript{MP} is applied, wave speed estimates of those sections are close to their nominal values, thus they are identified as normal sections, which is consistent with ultrasonic measurements. The true anomalous section Xb is highlighted as a result. To summarize, ITA\textsubscript{MP} provided a clear interpretation of the parameter estimates, and improved the ability to use these estimates for the detection of the anomalous sections.

However, a few true anomalous sections (for example, section G in Figure 14) were missed by ITA\textsubscript{MP}. This is likely due to the coarse spatial resolution (roughly 12 m, the 4 km pipe section was discretized into more than 300 reaches) used in the ITA\textsubscript{MP}. The ultrasonic measurement is taken at 5 m intervals, so the length of several anomalous sections was less than the spatial resolution used in ITA\textsubscript{MP}. As a result, those anomalous sections were missed by ITA\textsubscript{MP}.

The comparison between the measured pressure trace and the predicted pressure trace obtained by the best estimated wave speeds (ranked in terms of objective
Figure 16 | Comparison of measured pressure trace and predicted pressure trace obtained by best estimated wave speeds of Stage 3 at ACFP43.

Figure 17 | Comparison of measured pressure trace and predicted pressure trace obtained by best estimated wave speeds of Stage 3 at SC24.
function value) in Stage 3 is also examined. Figures 16–18 show the comparison between the measured pressure trace and the predicted pressure trace obtained by the best estimated wave speeds from Stage 3. The high frequency pressure fluctuations at SC24 are due to high frequency pressure oscillations, and resultant cavitation, within the pipework of the generator. The inverse model is not able to describe this phenomenon, so they did not appear in the predicted pressure traces in Figure 17. The pressures at ACFP43 and ACFP44 also exhibit significant oscillations in the first 0.5 s period after the transient has arrived. The oscillations were likely due to the oscillations generated at SC24, so the pressures of the first 0.5 s have been excluded from the objective function calculations. It can be seen that a reasonable match between the measured and predicted pressure trace is achieved for the time period 0.5 s after the wave front.

Overall, given the difficulty of the field problem that ITA_MP attempted, the anomalous sections identified by Stage 3 are generally consistent with the sections identified by ultrasonic measurements. The identifiability is improved compared with the conventional ITA (Stage 1 in the ITA_MP), which is indicated by less variation in the wave speed estimates and a decreased objective function (meaning a better match between the predicted and measured pressure traces).

CONCLUSIONS

Parameter identifiability is a problem commonly encountered in large-scale ITA applications for determining pipeline condition. To tackle this problem, multi-stage parameter-constraining ITA (ITA_MP), a new method for pipeline condition assessment, has been developed and is presented in this paper. ITA_MP iteratively generates sets of parameter estimates and modifies the search-space interval for the generation of subsequent estimates based on identifying a reach as either in a normal or anomalous state. The parameter estimates for the normal reaches become increasingly concentrated and
outliers in the parameter estimates are eliminated, thus identifiability of parameter values is enhanced.

This method has been verified by a numerical study in this paper. The results show that the wave speeds estimated by the conventional ITA (equivalent to Stage 1 of ITAMP) possess a number of anomalous wave speed estimates, despite the methods seemingly being a reasonable match to the numerical data. After Stage 2, the estimated wave speeds are much closer to the true wave speeds, and a large proportion of anomalous estimates are eliminated.

A field case study on a section of the Morgan transmission pipeline in South Australia has been conducted to verify the capability of ITAMP to assess pipeline condition assessment by identifying anomalous sections. Wave speed estimates by ITAMP have less variance and higher identifiability than the conventional ITA. The results show a reasonable consistency between anomalous sections detection by ITAMP and available ultrasonic measurements, demonstrating that ITAMP can be used as a tool for pipeline condition assessment in the field.

ACKNOWLEDGEMENTS

The research presented in this paper has been supported by the Australia Research Council through the Discovery Project Grant DP140100994 and the Linkage Project Grant LP130100567. The first author thanks the Chinese Scholarship Council and the University of Adelaide for providing a joint postgraduate scholarship.

REFERENCES


of Hydraulics; Proceedings, Institution of Engineers, Australia, 373.


First received 18 November 2017; accepted in revised form 22 December 2017. Available online 1 February 2018