Reliability Engineering and System Safety 162 (2017) 40–50

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Automated matching of pipeline corrosion features from in-line inspection data

Markus R. Dann[a,*], Christoph Dann[b]

[a] Department of Civil Engineering, University of Calgary, Calgary, Canada
[b] Machine Learning Department, Carnegie Mellon University, Pittsburgh, USA

ARTICLE INFO

Keywords:
Pipeline
Corrosion
In-line inspection
Feature matching
Integrity assessment

ABSTRACT

The integrity assessment of corroded pipelines is often based on in-line inspection (ILI) results. Before determining the corrosion growth for the integrity assessment, the detected corrosion features from two or more ILIs need to be matched with respect to their location in the pipeline. The objective of this paper is to introduce a framework for automated feature matching. The input for the framework is the locations of all detected corrosion features and girth welds from each ILI. Using a multi-step approach, the size of several ILIs with a possibly large number of features is reduced to a set of independent smaller problems to match efficiently the corrosion features. The results include the matched features for the subsequent corrosion growth analysis and the identification of outliers that cannot be matched. The applied probabilistic matching assigns to each feature pair a probability of being a match to reflect the inherent uncertainty in the matching process. The proposed framework replaces manual matching, which can be time intensive and prone to errors, particularly for internal corrosion with high feature densities. It reliably matches features in pipelines and supports the integrity and risk assessment of pipeline systems.

1. Introduction

Pipelines are large infrastructure systems that efficiently transport hydrocarbon products over short and long distances from production to end users. Most pipelines consist of carbon steel, which makes corrosion one of the most common integrity threats of these systems. Corrosion is a time dependent process that gradually reduces the wall thickness of pipelines. It can lead to leak and rupture failures with significant consequences for society, the environment, and the economy.

The integrity of corroded pipelines has to be managed over the entire lifetime of the pipeline to prevent service interruptions and failures. In-line inspections (ILIs) are commonly used as a non-destructive means of structural health monitoring to detect and size corrosion features in pipelines [1]. ILIs are usually performed every few years. To determine feature-specific corrosion growth and time to failure, the features from at least two ILIs have to be matched with respect to their location in the pipeline.

In principle, feature matching can be performed using either the raw signal information from the ILI tools or the individual corrosion features that are already identified and sized as the outcome of the raw signal processing. The first option has the advantage of relying on the raw signals, which may contain more information than the reported locations of the corrosion features and, therefore, can increase the chance of obtaining more accurate matching results. However, the focus of this paper is on the second option. This option may be slightly affected by the already sized features, but it is adopted for the following three reasons:

1. Engineers, who typically perform the corrosion growth analysis and integrity assessment of pipelines, have neither access, knowledge, nor time to work with the raw signal data.
2. Changing the inspection technique from one ILI to another can make the direct comparison of the raw signals and the feature matching more difficult, if not impossible.
3. Feature matching based on the sized features can be performed independently of the ILI vendor. This possibility is helpful if different vendors have been used over time to inspect the pipeline.

Feature matching using the sized features is often performed manually. It is usually a time intensive process that compares the feature locations from one ILI to another and manually identifies the features that belong together; further, it is prone to error. The objective of this paper is to introduce a framework for the automated matching

* Corresponding author.
E-mail address: mrdann@ucalgary.ca (M.R. Dann).

http://dx.doi.org/10.1016/j.ress.2017.01.008
Received 1 March 2016; Received in revised form 10 January 2017; Accepted 13 January 2017
Available online 21 January 2017

Nomenclature

α	parameter to control the number of outliers
β	parameter to control transformations close to identity
$\boldsymbol{a}$	transformation matrix
$\boldsymbol{I}$	identity matrix
$\boldsymbol{p}$	correspondence matrix
$\boldsymbol{r}$	vector of outlier probabilities for the moving set
$\boldsymbol{s}$	vector of outlier probabilities for the reference set
$\boldsymbol{u}$	vector of center coordinates of the outlier clusters for the moving set
$\boldsymbol{v}$	vector of center coordinates of the outlier clusters for the reference set
$\boldsymbol{x'}_i$	vector of transformed homogeneous coordinates for feature i in the moving set
$\boldsymbol{x}_i$	vector of homogeneous coordinates for feature i in the moving set
$\boldsymbol{y}_k$	vector of homogeneous coordinates for feature k in the reference set
λ	annealing rate
$\widetilde{X}$	extended moving point set
E	objective function
i	index for the features in the moving set
k	index for the features in the reference set
L	number of annealing steps to reach the final temperature T_{final}
l	current annealing step
m	number of features in the reference set
n	number of features in the moving set
p_{ik}	probability of a match of the features i and k
q	number of ILIs available for feature matching
r_i	probability that feature i is an outlier
s_k	probability that feature k is an outlier
T	current annealing temperature
T_0	initial annealing temperature
T_{final}	final annealing temperature
X	moving point set
Y	reference (fixed) point set

of corrosion features for a more reliable and time efficient matching that replaces the manual process. The framework consists of five steps. In the first three steps, the overall problem of matching features from several ILIs is transformed into multiple instances of a 2D point matching problem. It is necessary to maintain a high computational efficiency even for numerous ILIs with many defects. The core step of the framework is the application of a probabilistic point matching model that identifies both matching corrosion features and outliers, which are features that can not be matched. The individual 2D point matching results from the previous step are compiled in the final step into a consistent solution for the original matching problem. The proposed framework helps pipeline operators to obtain the required matched features from ILI data for the integrity and risk assessment of corroded pipelines.

In essence, the framework mirrors manual feature matching, but it outperforms its manual counterpart in terms of processing time and accuracy. For example, the proposed framework eliminates human errors that can lead to mismatching features, particularly for high density corrosion. Instead of using one complex feature matching model, the proposed framework with its five steps is not only simpler, but also computationally more efficient. Additionally, the obtained matches can be easily verified using a visual representation of the results.

The remainder of this paper is organized as follows. The problem of matching pipeline corrosion features is described in detail in Section 2, and the challenges for the framework of automated feature matching are identified. Section 3 introduces feature matching as a duality between correspondence and transformation and briefly reviews existing matching models. The proposed framework for automated matching of pipeline corrosion features is given in Section 4. Section 5 provides a numerical example where the framework is successfully applied to identify matching features and outliers. The final conclusions are discussed in Section 6.

2. Matching of pipeline corrosion features

2.1. ILI measurement errors and thresholds

The integrity of corroded pipelines needs to be continuously managed to prevent unnecessary shutdowns and failures. A part of the analytical integrity assessment is the estimation of the feature-specific (local) corrosion growth, which determines the safe remaining lifetime of the individual corrosion features. Feature-specific results are preferred over global and segment-wide results since they better support decision making regarding local verification inspections and repairs.

Local corrosion growth is often determined from ILI data by analyzing the measured feature depth from at least two ILIs. An ILI is a non-destructive inspection where an ILI tool, which is commonly referred to as an intelligent pig, moves through the pipeline to inspect the pipe wall. The objective of an ILI is to obtain number, location, and size of the corrosion features in the pipeline [1]. Standard ILI measurement principles are based on magnetic flux leakage (MFL), ultrasonic testing (UT), and Eddy current (EC) testing. Each ILI uses its own coordinate system to reference the location of the detected corrosion features in a pipeline. Consequently, coordinate systems from different ILIs are not aligned [2], and the measured positions of the corrosion features from two and more ILIs cannot be meaningfully compared without adjustment. Therefore, before the local corrosion growth analysis is conducted, the detected corrosion features must be matched by aligning the coordinate systems and identifying the features that belong together.

A perfect match is a feature pair where a feature from one ILI can be assigned with certainty to a feature from another ILI. Ideally, all features in a pipeline can be perfectly matched. However, the following ILI measurement errors and thresholds [3] limit perfect feature matching for pipelines:

1. Detection, measurement and reporting thresholds: The sensors of the ILI tool can only detect and measure features with sizes larger than the detection threshold and measurement threshold, respectively. A common detection threshold for MFL-based tools is 10% of the wall thickness (wt) of the pipeline; further details on the detection threshold are provided in [3,4]. Additionally, a reporting threshold can be applied where only detected features with sizes above the reporting threshold are reported. It is a means to remove many false calls as false calls frequently occur on the features below the threshold.
2. Detection error: If the feature size is above the detection threshold, the ILI tool does not detect the existing corrosion feature with certainty. The probability of detection (POD) describes the detection capabilities of the ILI tool above the detection threshold.
3. False call errors can occur when the ILI tool reports a corrosion feature, but the pipeline wall is intact and the feature does not in fact exist. The probability of false calls (POFC) describes the likelihood that a reported feature is a false call.

4. Location error: The reported feature locations from two or more ILIs are not perfectly aligned due to (1) ILI-specific coordinate systems, (2) noise on the reported locations, and (3) corrosion growth of the features between ILIs.

The location error causes the coordinate systems of the ILIs and the feature locations to be misaligned, while all remaining errors and thresholds result in outliers. An outlier is a reported feature that does not have a counterpart at the other ILI because either the reported feature is a false call, the existing counterpart is not detected or reported, or the counterpart does not exist due to the growth of a new feature after the previous ILI.

2.2. Manual matching of corrosion features

In practice, corrosion feature matching is often performed as a manual and iterative process. First, reference points such as girth welds and anodes are used to determine an initial transformation for the coordinate system of one of the ILIs. Second, a few matching features are identified based on their distinct or isolated location in order to minimize matching error. Third, the locations of the few matched features are used to determine a refined transformation by aligning the coordinate system of first ILI and using the coordinate system of the second ILI as a reference. Fourth, the transformation is applied to the coordinates of all features from the first ILI to identify more matches. The procedure continues by alternating between refining the transformation and identifying additional matches until a satisfactory number of matched features is obtained for the subsequent corrosion growth analysis. Manual and reliable feature matching is possible if the feature density is low in a pipeline. High densities of corrosion features, often encountered in upstream pipelines, lead to a time intensive matching process. In addition, the matching process is prone to errors and feature mismatch can occur due to the close proximity of other features.

2.3. Why is feature matching important for the reliability assessment of pipelines?

The integrity and reliability of corroded pipelines is assessed by structural reliability methods [5–7]. A corrosion feature can fail either as small leak, large leak, or rupture. A small leak is pressure independent and occurs if the depth of the feature exceeds the wall thickness. Large leaks and ruptures are pressure driven failure modes. Limit state models have been developed for all three failure modes [8,9]. They are used to estimate the time dependent pipeline reliability that can be compared with target reliability levels to determine the remaining lifetime of the pipeline.

The reliability analysis requires the size of the corrosion feature as input in all three limit state models. Corrosion is a time dependent process and, therefore, the present and future feature size needs to be estimated. Stochastic corrosion growth models are applied to infer the corrosion growth from the ILI results. They account for spatial and temporal uncertainty of the corrosion growth, measurement error, and epistemic model uncertainty. Basic models use uncertain corrosion growth rates [10,11]. More advanced models rely on stochastic processes for the corrosion growth such as the gamma process [12–14], the inverse Gaussian process [15], and others [16,17] to better capture the temporal uncertainty of corrosion growth [18]. To the best knowledge of the authors, all available stochastic models implicitly assume perfectly matched features to infer the feature specific corrosion growth from the inspection results. None of the stochastic corrosion growth models considers the uncertainty of feature matching. The objective of this paper is to develop a model to match corrosion features to support the existing approach of reliability-based integrity assessment of corroded pipelines using stochastic corrosion growth models. The feature matching model should improve predictions when combined with current stochastic corrosion growth models but also opens the avenue for development of new stochastic corrosion growth models that explicitly take matching uncertainty into account.

2.4. Example: impact of feature mismatch

A small example of two adjacent corrosion features without measurement error is considered in Fig. 1 to demonstrate the ramifications of a feature mismatch. The depths of each feature at the time of the two inspections, at times zero and four years, are provided. The two solid lines that go through the ILI-measured depths indicate the correct corrosion growth path for the two features assuming they are correctly matched and have linear past and future corrosion growth. A leak-based repair criterion is often the 80th percent level of the wall thickness [19], which is reached at 2 years after the most recent inspection for the second feature. If the features are cross-mismatched, the dashed line, which connects the depths of the first feature at the first inspection with the depth of the second feature at the second inspection, shows the incorrect corrosion growth path for the second feature. In this case, repair would be necessary in 4 years after the second inspection rather than 2 years according to the correct corrosion growth. The implications of a feature mismatch can be significant for the remaining lifetime analysis and subsequent decision making regarding repair and maintenance actions. Mismatches can lead to high-risk pipeline operation and even to pipeline failure. One of the objectives of the proposed framework for automated feature matching is to reduce matching errors and the resulting consequences.

2.5. Challenges and requirements for automated feature matching

Based on the previous discussion, the following challenges for a framework for automated matching of pipeline features are identified:

1. Pipelines can have more than two ILIs where each ILI can result in a massive data set. It is necessary to process such cases efficiently.
2. Large numbers of outliers are possible as many reported corrosion features exist in only one ILI and do not have a counterpart in another ILI. This effect can occur due to the following three events:
   (a) An existing corrosion feature is not detected by the ILI tool due to detection error and threshold.
   (b) The ILI tool reports a false call.
   (c) The growth of new corrosion features occurs between inspections.

   It is necessary to identify such outliers and remove them from or adjust accordingly the matching process. The outliers do not have to be classified according to the three events that cause them.
3. Two or more corrosion features can grow and merge into one feature. It should be possible to match multiple features from previous ILIs to one feature from a later ILI and vice versa.
4. The reported feature locations are subject to measurement error.

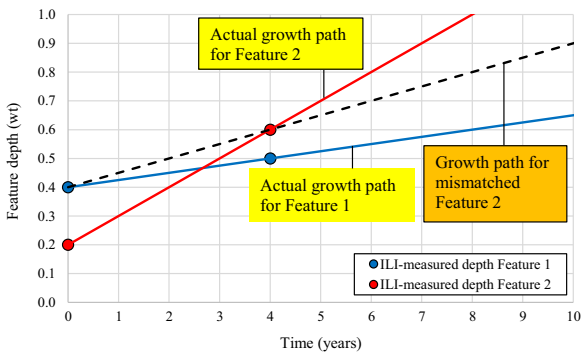

Fig. 1. Corrosion growth paths for correctly and mismatched features.

Consequently, perfect feature matching is limited and the measurement error needs to be accounted for in the feature matching model.

The main requirements for the framework can be summarized as follows: (1) able to quickly compute a solution even for large matching problems, (2) able to process multiple ILI problems, (3) has a generic matching model that handles outliers well and can be easily adjusted to different problems, (4) visualizes the results to easily verify them, and (5) quantifies the uncertainty inherent in the matching process. The proposed framework in Section 4 satisfies all the requirements.

3. Point matching: the coupled problem of correspondence and transformation

The approach to the matching problem is to treat the corrosion features in the proposed framework as points without dimensions. The points are located at the center of the features. The locations of the points and the corresponding point patterns provide enough information for matching the features. The objective of this section is to examine the 2D point matching problem in preparation for the framework in Section 4.

Two point sets $X = \{\boldsymbol{x}_i \text{ in } \mathbb{R}^2: i = 1, \ldots, n\}$ and $Y = \{\boldsymbol{y}_k \text{ in } \mathbb{R}^2: k = 1, \ldots, m\}$ are considered. An example is shown in Fig. 2. The vectors $\boldsymbol{x}_i = (x_{i1}, x_{i2})$ and $\boldsymbol{y}_k = (y_{k1}, y_{k2})$ describe the locations of the points. The coordinates of the points are not aligned and, therefore, one point set needs to be matched to the other point set. For example, point set X is matched to set Y.

The correspondence relationship between the point sets X and Y is of primary interest. The relationship is defined by the $n \times m$ correspondence matrix $\boldsymbol{p}$. Each row and column of the matrix represents a feature from X and Y, respectively. Element p_{ik} in $\boldsymbol{p}$ describes a score or degree of match for the feature pair (i, k). If a binary correspondence is applied, the elements p_{ik} take either 0 or 1 for no match or a perfect match between the points i and k, respectively. This case is also referred to as "hard-assign", and an example is provided in Table 1 (left).

In principle, point matching consists of two tasks:

- Determine the correspondence given the coordinate alignment: Identify the pairs of points in the two point sets that belong together. The results are reported in a correspondence matrix that contains all possible feature pair combinations from the two point sets X and Y. In the case of a binary correspondence assignment, a value one in the matrix indicates a matched feature pair, while zero is no match.
- Determine the transformation given the correspondence: The coordinate system of set X needs to be moved to improve the alignment with the coordinate system of set Y. The transformation mathematically describes the movement of the one coordinate system and can only be determined with the knowledge of (at least some) matched feature pairs. Perfect alignment, which means the locations of all features match perfectly, is usually not possible using standard (e.g. affine) transformations due to the location errors of the features. More advanced (e.g. non-affine) transformations increase the alignment and can lead to perfect alignment of all features, but this improvement is not necessary for practical applications such as pipeline feature matching.

Performing one task, if the information from the other task is already available, is relatively straight forward. However, a reasonable transformation cannot be determined without a good correspondence and vice versa. Hence, feature matching is a coupled problem where the transformation and correspondence have to be solved simultaneously. This duality of correspondence and transformation is also experienced by manually matching pipeline corrosion features as explained in Section 2. An iterative procedure is typically applied to obtain a solution for the matching problem. A suitable framework for matching corrosion features in pipelines needs to be able to solve this coupled problem of determining the transformation and the correspondence.

Research has been focused on matching models to simplify the analysis of the coupled problem and solve for either the transformation when the correspondence is given or the correspondence when the transformation is given. An extensive literature review on models for point matching is given by Chui and Rangarajan [20]; a short summary of this work is provided below. Approaches for determining the transformation include the methods of moments [21], the Hough Transform [22], the Hausdorff distance [23], the alignment method [24], geometric hashing [25], and tree searches [26]. Methods for determining the correspondence typically belong to one of the following three groups [20]: dense feature-based methods [27–29], methods for sparsely distributed point sets [30–32], and methods for weighted graph matching [33–35].

Methods that determine both the correspondence and the transformation are of increased practical importance for the matching of corrosion features in pipeline. The iterative closest point (ICP) algorithm [36] is an iterative method where the correspondence is determined based on the nearest-neighbor relationship, which assigns a match to the features that are the closest to each other. The transformation is then refined before the next iteration steps begin. The ICP algorithm uses binary (1 to 1) correspondence variables and, as a result, performs poorly due to the presence of outliers [20]. It reflects basic manual feature matching, and both methods have similar performance issues.

For non-perfect or uncertain matching, the binary (0–1) assignment in the correspondence matrix $\boldsymbol{p}$ is replaced, and the value of p_{ik} is interpreted as the probability of a match between the points i and k [37]. In this case, p_{ik} can take any real value in the closed interval [0,1] and this type of probabilistic correspondence definition is referred to as "soft-assign" [20]. Examples for a 3×3 correspondence matrix with soft-assign is also given in Table 1 (right). Soft-assign has been used in probabilistic approaches [38,39] and joint linear assignment-least square optimization problems [20,40] to match features. Pose estimation as proposed by [22,40] refers to problems where the transformation of 3D objects are estimated from 2D images.

4. Framework for automated matching of corrosion features in pipelines

The input for the framework is the ILI results. In the general case, q ILIs ($q \geq 2$) are available. The results for each ILI are summarized in table format, the so-called pipe tally [4]. The pipe tally contains the location coordinates and size of all detected corrosion features (Fig. 3). The number of features varies from one ILI to another due to growth of new defects and measurement errors (Section 2). The locations of the

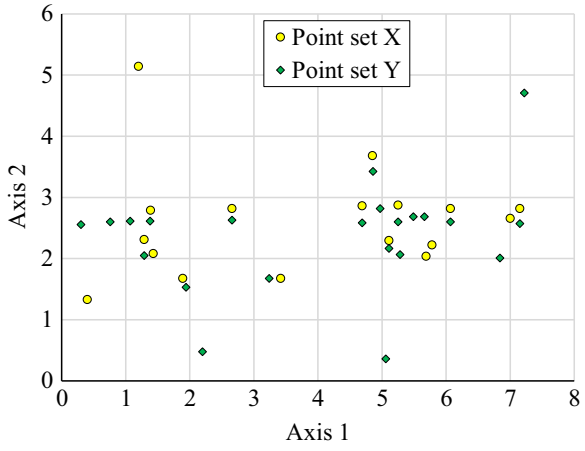

Fig. 2. Example of two point sets in a planar system before matching.

Table 1
Examples for 3×3 correspondence matrix using hard-assign (left) and soft-assign (right).

p_{ik}	k=1	2	3
i=1	1	0	0
2	0	0	0
3	0	1	0

p_{ik}	k=1	2	3
i=1	0.8	0.2	0
2	0	0.1	0
3	0	0.7	0.2

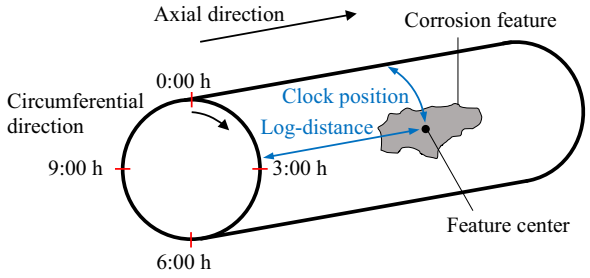

Fig. 3. Location of a corrosion feature in a pipeline.

corrosion features are given in axial and circumferential directions of the pipeline using the so-called log-distance[1] and clock position, respectively. The objective of the framework is to match the corrosion features with respect to their location in the pipeline using log-distance and clock position.

The locations obtained from the ILI are the coordinates of the center of the features as shown in Fig. 3. The features are treated in the subsequent matching model as points, which are represented by the center of the feature. The feature's length, width, and depth are ignored at the current stage, because they are subject to sizing errors, and it is assumed that the center points of the features contain enough information to successfully match the features. The extension of the proposed approach to include feature size information is left as future work.

The proposed framework for the automated matching of corrosion features consists of the following five steps:

1. Reduce multi-ILI matching to 2-ILI matching problems (Section 4.1)
2. Split the pipeline into sections to reduce the problem size (Section 4.2)
3. Map the feature locations to a 2D plane (Section 4.3)
4. Match the corrosion features (Section 4.4)
5. Transform the results to the original matching problem (Section 4.5)

The first step of the framework addresses the feature matching problem for more than two ILIs and how it can be reduced to a set of independent pairwise-ILI problems. The second step aims to reduce the problem size of matching features for an entire pipeline by splitting the pipelines into sections that can be analyzed independently. The features are then mapped for each section on a planar surface in preparation for matching in the fourth step. The mapping needs to take the tubular shape of the pipeline into account. The core step of the framework consists of a point matching model to match the corrosion features and identify the outliers. The results from the set of smaller matching problems are then transformed to provide the final solution for the original matching problem in the fifth step. The implementation

[1] The word log in log-distance is not related to the logarithm in mathematics.

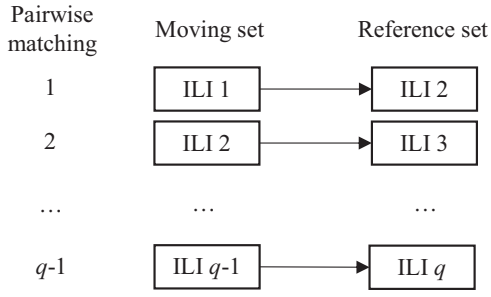

Fig. 4. Pairwise analyses of multi-ILI matching problems.

of the point matching model from step four is explained in Section 4.6.

While the individual steps seem straight forward, the five-step process is a powerful tool to support lifetime analyses and risk assessments of pipelines. It is fully automatable, requires minimal user interaction, and can easily be controlled by only three parameters:

1. α to control the number of outliers
2. β to control the transformation
3. T_{final} to control the correspondence

4.1. Step 1: reduce multi-ILI matching to 2-ILI matching problems

Standard matching problems are formulated for two sets of features where a moving set is matched to a reference set. The coordinates of the moving set are adjusted by a coordinate transformation, while the coordinate system and the locations of the features of the reference set remain fixed. It is possible to develop a model that can process all q ILIs at once, but it can lead to a large problem size and, consequently, can require a significant amount of computing time. Instead of having one complex feature matching model for pipelines, the problem is reduced to series of $q-1$ independent matching problems as shown in Fig. 4. Consecutive ILI pairs are considered where the feature set from the former ILI is the moving set and the subsequent ILI results belong to the reference set. For example, assume q=3 ILIs (ILI 1, ILI 2, ILI 3) are available for feature matching. ILI 1 is then matched to ILI 2 in the first matching problem and then ILI 2 is matched to ILI 3 in the second problem. The advantage of this sequential matching procedure is to track the growth of the corrosion features from the first to the most recent ILI. The results of a new ILI can simply be added by analyzing only one additional feature matching problem that consists of the new ILI and its direct precursor. The remainder of the framework focuses on a single matching problem where a moving set needs to be matched to a reference set.

4.2. Step 2: split the pipeline into sections

A model can be applied to an entire pipeline to match all corrosion features at once. However, this approach can create two problems. First, the coordinate transformation needs to be flexible enough to provide appropriate transformations over the entire length of the pipeline. An affine transformation is not always suitable this case. Second, the size of the problem can be an issue due to the possibly large numbers of features that need to be matched. The runtime of the point matching model in Step 4 (Section 4.4) is (at least) quadratic in the number of features. Matching thousands of features from two ILIs, which is a reasonable task for upstream pipelines, can be extremely limited if personal computers with standard computing power are used. To avoid the two possible issues, the following matching procedure is proposed:

1. ILIs provide the locations of girth welds and anodes, which can be considered as fixed reference points along the pipeline. These reference points are matched first to divide the pipeline into a set

of shorter segments.
2. The proposed matching model in Section 4.4 can then be applied independently to each segment. It reduces the amount of features that are processed simultaneously.

In our approach, the pipeline is segmented as shown in Fig. 5. A pipeline usually consists of a series of joints, each with a length of approximately 12 m, which are welded together at both ends. Each section for the analysis consists of one joint. Instead of having the boundaries of each section at the girth welds, the sections overlap (e.g. 0.3 m) to avoid features not being covered by either of the two adjacent sections. The number of corrosion features in a joint is sufficiently small to achieve an efficient matching process.

It was empirically verified that an affine transformation for a segment is sufficient based on a maximum segment length of approximately 13 m. Empirically, shift and scaling typically lead to satisfactory results; more complex non-affine transformations are not necessary.

4.3. Step 3: map the feature locations to a 2D plane

Pipelines are 3D longitudinal systems, and one would expect to use models to match pipeline corrosion features in 3D. However, corrosion can only occur in the pipe wall, a 2D manifold, as shown in Fig. 3. The wall can be considered as a thin 2D steel sheet that was bent in the shape of a tube and welded together along the axial direction to get a pipe joint. Instead of using the tubular shape of the pipeline for feature matching, the features are mapped and matched on a 2D planar surface as shown in Fig. 6. The exact feature shape is usually not available and the assumed shape for the integrity assessment is often a rectangle (Fig. 6) with the size of the maximum feature length and width detected by the ILI tool. It is assumed that the locations obtained from the ILI are the coordinates of the center of the features as shown in Fig. 3, which are approximately the center of the feature rectangle (Fig. 6). The planar surface is essentially the unrolled pipeline joint and allows the application of existing 2D matching models. However, simple point matching with this method does not lead to the desired results. One simple approach is to "double unroll" the pipeline.

The double unrolling assumes the proposed affine coordinate transformation is applicable to planar surfaces as shown in Fig. 6. Features located close to the two boundaries at 0:00 h and 12:00 h are not considered to be a match as they appear to be far apart on the planar surface. However, the circumferential positions 0:00 h and 12:00 h in the actual pipeline are identical due to the pipeline's circular cross section (Fig. 3). Consequently, features located close to the two boundaries should be considered as a possible match. The formal way of including this condition in the subsequent feature matching analysis, which is primarily based on the distance between the feature pairs, is to determine the two possible distances of a feature pair in a circular cross section and select the minimum of the two distances for further processing. This formal approach can limit the application of closed form solutions for determining the optimal correspondence and transformation. However, if only one Euclidean distance exists for a feature pair, the application of closed form solutions is allowed. This solution to the limits of the formal approach can contribute to obtaining a fast solution. Therefore, one possibility to reflect the pipeline's circular cross section in the planar surface is to expand the moving set by adding the two halves of the features along the 0:00 h and 12:00 h positions as shown in Fig. 7. In essence, this expansion

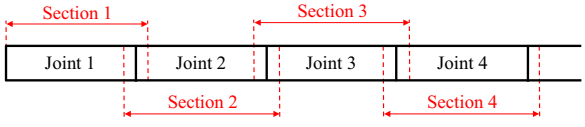

Fig. 5. Recommended segmentation of the pipeline for efficient feature matching analysis.

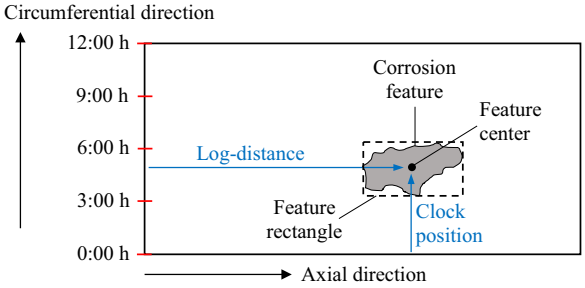

Fig. 6. Location of a corrosion feature on a 2D plane.

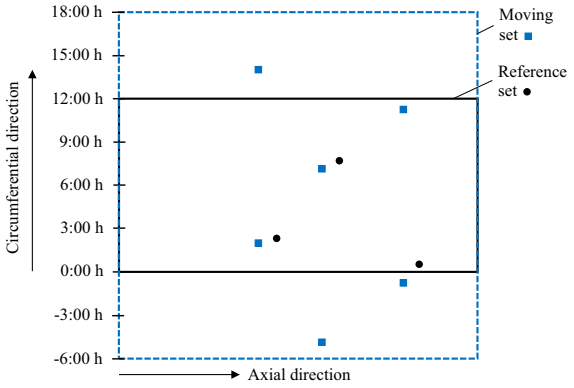

Fig. 7. Mapping of the reference set and twice the moving set on a planar surface to account for the tubular shape of the pipeline.

means unrolling the tubular pipeline twice in a circumferential direction, first from −6:00 h to 6:00 h and second from 06:00 h to 18:00 h. A feature from the moving set that is located just before 12:00 h is then also located just before 0:00 h (Fig. 7). They can now be matched to the reference features after 0:00 h and before the 12:00 h. This simpler approach returns an increased number of outliers because approximately half of the features from the moving set are not matched. The point matching model in Section 4.4 is designed for this case where large outlier ratios can be present even in both point sets. Overall, the double unfolding of the moving set is a very efficient approach to transfer the circular cross section of the pipeline to the planar surface in the 2D matching problem.

The detected corrosion features from two ILIs are considered as two point sets where $X = \{\boldsymbol{x}_i \text{ in } \mathbb{R}^2: i = 1, \ldots, n\}$ is the point set from the first ILI and $Y = \{\boldsymbol{y}_k \text{ in } \mathbb{R}^2: k = 1, \ldots, m\}$ is the point set from the second ILI. Point set X is extended to set $\widetilde{X} = \{\boldsymbol{x}_i \text{ in } \mathbb{R}^2: i = 1, \ldots, 2n\}$ to account for the double unrolling as described above (Fig. 7). Set $\widetilde{X}$ is treated as the moving set and matched to the reference set Y. Homogeneous coordinates are applied in vector format where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, 1)$ and $\boldsymbol{y}_k = (y_{k1}, y_{k2}, 1)$. The coordinates describe the feature locations on the surface. The first and second elements in the coordinate vectors refer to the axial and circumferential positions of the corrosion features, respectively.

Corrosion in pipelines is divided into external and internal corrosion. The feature matching model proposed below does not allow the simultaneous processing of external and internal corrosion. This limitation is considered to be minor because the model is tailored toward high density internal corrosion where features are in close proximity to each other. External corrosion usually occurs at isolated locations due to coating failure resulting in reduced manual matching effort for those features. Nevertheless, the matching model can still be applied independently to internal and external corrosion.

4.4. Step 4: match the corrosion features

Fig. 7 is the basis for matching corrosion features in pipelines as points on a planar surface. The extended moving set $\widetilde{X}$ needs to be matched to the reference set Y, because the measured features are not perfectly aligned. A coordinate transformation is necessary for the moving set to achieve improved alignment between $\widetilde{X}$ and Y. Due to Step 2 where an entire pipeline is split into short segments, it was empirically found that affine transformation for $\widetilde{X}$ is sufficient. However, the proposed model can be extended to non-affine transformations. The affine transformation from $\boldsymbol{x}_i$ to $\boldsymbol{x}'_i$ consists of translation, scaling, rotation, and shear; it is described by the following equations:

$$\boldsymbol{x}'_i = \boldsymbol{a}\boldsymbol{x}_i \quad \forall\, i = 1, \ldots, 2n \tag{1}$$

$$\boldsymbol{a} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where $\boldsymbol{a}$ is the feature-independent 3×3 unknown affine transformation matrix. The elements a_{13} and a_{23} account for the translation in axial and circumferential direction, respectively. The remaining elements a_{11}, a_{12}, a_{21}, and a_{22} introduce scaling, rotation, and shear in the transformation. The last row in $\boldsymbol{a}$ consists of 0 and 1 to retain homogeneous coordinates.

The correspondence relationship between the point sets $\widetilde{X}$ and Y is defined by the $2n \times m$ correspondence matrix $\boldsymbol{p}$. The element p_{ik} in $\boldsymbol{p}$ describes the probability of a match for the feature pair (i, k) using soft-assign (Table 1). If a feature from either set can not be matched, it is identified as an outlier. The vectors $\boldsymbol{r}$ and $\boldsymbol{s}$ contain the probabilities of each feature being an outlier, where $\boldsymbol{r}$ is a $(2n \times 1)$ vector for the outliers in $\widetilde{X}$ and $\boldsymbol{s}$ is a $(1 \times m)$ vector for the outliers in Y. Soft-assign is also applied for the outliers; hence, the elements of $\boldsymbol{r}$ and $\boldsymbol{s}$ can also take any real value in the closed interval [0,1].

The correspondence matrix $\boldsymbol{p}$ and the outlier vectors $\boldsymbol{r}$ and $\boldsymbol{s}$ are subject to the following constraints due to soft-assign and their interpretation as probabilities:

$$0 \le p_{ik} \le 1 \quad \forall\, i = 1, \ldots, 2n;\ k = 1, \ldots, m \tag{3}$$

$$0 \le r_i \le 1 \quad \forall\, i = 1, \ldots, 2n \tag{4}$$

$$0 \le s_k \le 1 \quad \forall\, k = 1, \ldots, m \tag{5}$$

$$\sum_{k=1}^{m} p_{ik} + r_i = 1 \quad \forall\, i = 1, \ldots, 2n \tag{6}$$

$$\sum_{i=1}^{n} p_{ik} + s_k = 1 \quad \forall\, k = 1, \ldots, m \tag{7}$$

Eqs. (3)–(5) result directly from the interpretations of p_{ik}, r_i, and s_k as a probabilities as mentioned above. The conditions in Eqs. (6) and (7) ensure that each feature can only be assigned once.

The unknown variables of the feature matching problem are the correspondence matrix $\boldsymbol{p}$, the vector $\boldsymbol{r}$ for the outlier correspondence of $\widetilde{X}$, the vector $\boldsymbol{s}$ for the outlier correspondence of Y, and the matrix $\boldsymbol{a}$ in (2) for the affine coordinate transformation of $\widetilde{X}$. All unknowns are determined by optimizing an objective function presented below.

The objective function is based on the Euclidean distance between the features. A feature pair that does not belong together has a large distance, while the distance for a true pair should be (close to) zero. Therefore, feature matching can be considered similar to a weighted least squares optimization problem with the following objective function:

$$E(\boldsymbol{a}, \boldsymbol{p} | \widetilde{X}, Y) = \sum_{i=1}^{2n} \sum_{k=1}^{m} p_{ik} \parallel \boldsymbol{y}_k - \boldsymbol{a}\boldsymbol{x}_i \parallel^2 \tag{8}$$

The term $\parallel \boldsymbol{y}_k - \boldsymbol{a}\boldsymbol{x}_i \parallel^2$ represents the square of the Euclidean distance between the features i and k where $\boldsymbol{a}\boldsymbol{x}_i$ in (1) and (8) is the transformed coordinate for feature i. The probability p_{ik} acts as a weight factor for the square distance. The double summation in (8) covers all possible $2\,nm$ feature combinations. If the features i and k are a perfect match, the distance between the two features should be as small as possible, and the probability p_{ik} should be as large as possible. The motivation behind (8) is that E is smaller if the features are more accurately matched. A trivial solution for minimizing the objective function is $\boldsymbol{p} = 0$, and all features in the two sets are declared as outliers; the following extension of the objective function prevents this undesired trivial solution:

$$E(\boldsymbol{a}, \boldsymbol{p}, \alpha | \widetilde{X}, Y) = \sum_{i=1}^{2n} \sum_{k=1}^{m} p_{ik} \parallel \boldsymbol{y}_k - \boldsymbol{a}\boldsymbol{x}_i \parallel^2 - \alpha \sum_{i=1}^{2n} \sum_{k=1}^{m} p_{ik} \tag{9}$$

The additional term on the right-hand side in (9) describes the expected number of matched features by the summation over the entire correspondence matrix $\boldsymbol{p}$. The parameter α is an unknown multiplier. The objective of the second term is to avoid the situation where all features in $\widetilde{X}$ and Y are declared as outliers and to match as many features as possible in $\widetilde{X}$ and Y.

As noted in [20], the affine transformation in (2) sometimes leads to a reflection transformation, which is not reasonable for pipelines. To avoid the reflection problem, a third and final reflection penalty term is added to the objective function:

$$E(\boldsymbol{a}, \boldsymbol{p}, \alpha, \beta | \widetilde{X}, Y) = \sum_{i=1}^{2n} \sum_{k=1}^{m} p_{ik} \parallel \boldsymbol{y}_k - \boldsymbol{a}\boldsymbol{x}_i \parallel^2 - \alpha \sum_{i=1}^{2n} \sum_{k=1}^{m} p_{ik} + \beta \mathrm{tr}[(\boldsymbol{a} - \boldsymbol{I})^{\mathrm{T}}(\boldsymbol{a} - \boldsymbol{I})] \tag{10}$$

The third term in (10) consists of the parameter β, the unknown transformation matrix $\boldsymbol{a}$, and the 3×3 identity matrix $\boldsymbol{I}$. This term stabilizes the optimization by penalizing transformations that vary significantly to the identity transformation. Transformations close to identity are reasonable for pipelines because the misalignment of the coordinates for $\widetilde{X}$ and Y is usually small. The objective of feature matching is to determine the unknown transformation matrix $\boldsymbol{a}$ and correspondence matrix $\boldsymbol{p}$ by minimizing the objective function E in (10) subject to the constraints in (3)–(7). The sole purpose of the parameters α and β is to control the optimal solution by avoiding many outliers and fostering transformations close to identity. If $\boldsymbol{p}$, α, and β are given in (10), the remaining unknown is the transformation matrix $\boldsymbol{a}$. In this case, a closed form solution for $\boldsymbol{a}$ is derived and provided in the Appendix; see [20,37] for further details. The advantage of an available closed form solution for $\boldsymbol{a}$ is shown in Section 4.6, which outlines and discusses the implementation of the model in a feature matching algorithm.

4.5. Step 5: transform the results to the original matching problem

After Step 4 of the framework is executed, it is necessary to check the obtained results at the overlap of the sections (Fig. 5) and transform the section-based results using the double unrolling (Fig. 7) back to the original matching problem. Both items are briefly discussed below including an outlook on how the final results of Step 5 can be used in the subsequent corrosion growth analysis.

The proposed segmentation of the pipeline in Fig. 7 leads to each girth weld having two results, one from each of the two adjacent sections. Ideally, the two results at a girth weld are identical or at least similar, but this is not guaranteed. Therefore, it is necessary to make a decision at each girth weld whether to keep the results from either one of the two sections. This decision can be achieved with a brief analysis at each girth weld. The objective function in (8), which is solely based on the Euclidean distance between the features, is determined only for the features in the overlapping zone around the girth weld using the transformation and correspondence from either section. Comparing the two values for the objective function, the section that has a lower

value provides a better match between the two point sets and this section should be used for the final results.

The extension of the moving point set from X to $\widetilde{X}$ in Step 3 (Section 4.3), to account for the tubular shape of the pipeline in the 2D matching analysis (Fig. 7), affects the final results. The correspondence matrix $\boldsymbol{p}$ contains each feature of X twice. Hence, the results need to be filtered to identify unambiguously the features in X that are matched to Y and the corresponding probabilities of having a match. The probabilities in the two rows of $\boldsymbol{p}$ that belong to a given feature need to be compared. The following three cases can occur:

1. Each row indicates that the feature cannot be matched; all probabilities in $\boldsymbol{p}$ are close to zero. The feature is a true outlier that cannot be assigned to any feature in Y.
2. One of the two rows identifies the feature as an outlier, while at least one probability in the second row is larger than zero. In this case the results of the second row are decisive and the feature is matched to Y with the stated probabilities. The other row containing close to zero probabilities can be ignored.
3. In the extreme case, both rows contain non-zero probabilities.

The first and second cases are frequently encountered when matching pipeline corrosion features. The third case is very rare and can only occur if the feature is located very close to the circumferential position of 0:00 h and 12:00 h and at least two features from Y are in the direct vicinity. The two rows of X should be compared to identify the result with the largest single probability of a match. This result should then be used in the original matching problem. The other row is removed from $\boldsymbol{p}$. This rule applies to all three cases, and it sequentially decreases the size of the initial $2n \times m$ correspondence matrix to a reduced $n \times m$ matrix by going over all features in X. It is necessary to normalize the reduced matrix, including the outlier vectors, to obtain the final correspondence matrix that satisfies the constraints of soft-assign. The proposed reduction of $\boldsymbol{p}$ supports the full automation of the framework.

The final result of the matching analysis is a reduced $n \times m$ correspondence matrix for each of the $q - 1$ matching problems. The matching probabilities can be directly incorporated into the corrosion growth analysis. However, such an advanced analysis has not been developed to the knowledge of the authors. If a corrosion growth analysis based on perfect feature matching is applied, the matching results can be divided into the following three groups:

1. Feature pairs that are identified with a high probability, for example 0.9, can be considered as perfect 1-1 matches.
2. Several features from the first ILI are matched to one feature of the second ILI and vice versa. The matching probabilities are below 0.9.
3. Outliers with probability of 0.9 and above.

The perfect matches from the first group correspond to the feature pairs that would have been identified by manual matching. They are of primary interest for the corrosion growth analysis where the growth is inferred either feature-specific or for the entire population of matches from the ILI-measured corrosion growth. The second group of features in the above list require further analysis, which is outside the scope of this paper, to better understand their relationship. For example, it is possible that several individual pits from the first ILI grow together to one feature by the time of the second ILI. A refined matching analysis is necessary where the size of the features is considered. The identified outliers in the third group are only detected in one of the two ILIs and do not have a counterpart in the other ILI. The features from the second and third group are usually removed from the corrosion growth analysis.

4.6. Implementation of the feature matching model in Step 4

The objective function in (10) for determining the transformation $\boldsymbol{a}$ and correspondence $\boldsymbol{p}$ is non-convex and, therefore, hard to optimize. However, deterministic annealing, which is also used in [20] and [37], has been shown to provide good solutions for such optimization problems. It is adopted in this paper to determine the optimal solution for $\boldsymbol{a}$ and $\boldsymbol{p}$ as outlined below.

The idea and purpose of deterministic annealing is similar to physical or metallurgy annealing where heat is used to alter the physical properties of a material. The variable to control the annealing is the temperature T. Deterministic annealing uses the temperature variable T to control the probabilities in the correspondence matrix $\boldsymbol{p}$. The temperature is modeled as $T = \lambda^l T_0$ where T_0 is the initial temperature at the beginning of the annealing process, $\lambda < 1$ is the annealing rate, and l is the current step of the annealing process. High temperatures lead to uniform probability assignment without a preference for matched feature pairs or outliers; as the temperature lowers, the fuzziness in $\boldsymbol{p}$ decreases. The correspondence matrix $\boldsymbol{p}$ becomes binary as the temperature approaches zero.

The elements of the matrix $\boldsymbol{p}$ and the vectors $\boldsymbol{r}$ and $\boldsymbol{s}$ are initially determined as follows [37]:

$$p_{ik} = \frac{1}{T}\exp\left[\frac{\alpha}{T} - \frac{(\boldsymbol{y}_k - \boldsymbol{a}\boldsymbol{x}_i)^{\mathrm{T}}(\boldsymbol{y}_k - \boldsymbol{a}\boldsymbol{x}_i)}{T}\right] \quad \forall\, i = 1, \ldots, 2n;\ k = 1, \ldots, m \tag{11}$$

$$r_i = \frac{1}{T_0}\exp\left[-\frac{(\boldsymbol{x}_i - \boldsymbol{a}\boldsymbol{x}_i)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{a}\boldsymbol{x}_i)}{T_0}\right] \quad \forall\, i = 1, \ldots, 2n \tag{12}$$

$$s_k = \frac{1}{T_0}\exp\left[-\frac{(\boldsymbol{y}_k - \boldsymbol{v})^{\mathrm{T}}(\boldsymbol{y}_k - \boldsymbol{v})}{T_0}\right] \quad \forall\, k = 1, \ldots, m \tag{13}$$

where T is the current temperature of the annealing process, T_0 is the initial (large) temperature at the beginning of the annealing process (l=0), and α is the multiplier parameter in (9) and (10) to control the number of outliers. Eq. (11) shows the interdependence of correspondence and transformation. The vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are the center of mass for the two point sets $\widetilde{X}$ and Y [37], respectively. They are referred to as outlier clusters [20], which represent artificial points for matching the outliers.

$$\boldsymbol{u}(\boldsymbol{x}) = \frac{1}{2n}\sum_{i=1}^{2n}\boldsymbol{x}_i \tag{14}$$

$$\boldsymbol{v}(\boldsymbol{y}) = \frac{1}{m}\sum_{k=1}^{m}\boldsymbol{y}_k \tag{15}$$

The algorithm shown in Fig. 8 is used to determine the unknown correspondence and transformation of the matching problem to match the corrosion features from two ILIs.

At the beginning of the matching analysis, the governing parameters and variables need to be initialized. For example, α controls the outlier assignment and can be considered as the percentage of outliers in the two point sets. It stays fixed throughout the entire analysis and, consequently, the final results in terms of transformation $\boldsymbol{a}$ and correspondence $\boldsymbol{p}$ depend on α. Initial tests with ILI data indicate that values between [0;0.1] for α seem reasonable as the transformation needs to be close to identity. The parameter β penalizes the deviance from the identity transformation. Values for $\beta \geq 200T$ are suggested based on experience to avoid unreasonable scaling transformations caused by the extended point set $\widetilde{X}$. λ is the annealing rate with values in the interval [0.90;0.99].

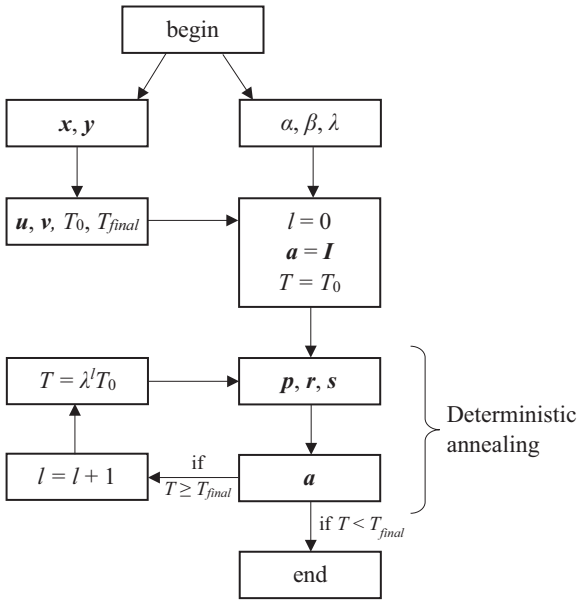

Fig. 8. Block diagram of the model for matching corrosion features as points.

In the next step, both outlier centers $\boldsymbol{u}$ and $\boldsymbol{v}$ are determined as given in (13) and (16). As recommended in [20], the initial temperature T_0 at the beginning of the annealing process is set to the largest square distance of all point pairs in $\widetilde{X}$ and Y. The final temperature T_{final} is the average of the square distance between the nearest neighbors within the moving point set.

The initial transformation $\boldsymbol{a} = \boldsymbol{I}$ is set to identity where all elements on the main diagonal are unity, and the remaining elements are zero. The annealing process begins with determining the current temperature T, which is equal to T_0 in the initial step (l=0) of the process. Afterwards, the temperature has to be reduced according to $T = \lambda^l T_0$ in each step $l > 0$ of the annealing process.

The alternating updates of $\boldsymbol{a}$ and $\boldsymbol{p}$ begin with a draft estimate of the correspondence matrix $\boldsymbol{p}$ and the outlier vectors $\boldsymbol{r}$ and $\boldsymbol{s}$ based on Eqs. (11)–(13). The $\boldsymbol{p}$, $\boldsymbol{r}$, and $\boldsymbol{s}$ are iteratively normalized to achieve a proper soft-assign using (6) and (7). The optimal transformation matrix $\boldsymbol{a}$ is then determined using (22) from the Appendix, which minimizes the conditional objective function $E(\boldsymbol{a}|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{p}, \alpha, \beta)$ obtained from (10).

The alternate updating is repeated by adjusting the current temperature T as mentioned above until the final temperature T_{final} is reached. More details on how to set up and initialize the matching algorithm can be found in [20,37].

In essence, the primary setup of the feature matching model rests on the idea by Chui and Rangarajan [20]. The extension by Yang [37] is adopted in Eqs. (8)–(10), and a solution is derived in the Appendix to allow simultaneous identification and proper processing of outliers in both sets $\widetilde{X}$ and Y. The feature matching problem for an entire pipeline can be reduced to independently match features for smaller sections of the pipeline; the non-affine portion of the transformation as proposed by [20,37] is not necessary and is removed from the proposed model.

5. Numerical example

The objective of this section is to demonstrate the proposed point matching model from Section 4. Due to the lack of publicly available ILI data, a hypothetical situation is considered where a pipeline is subject to internal corrosion and has been inspected twice (q=2) by in-line inspections. For demonstration purposes and better visualization only, a short 1 m section of the pipeline is analyzed. The pipeline diameter is 10″, which is a pipeline circumference of approximately 0.8 m. The center points of the reported features are shown in Fig. 9 for the first (ILI 1) and second (ILI 2) inspections. The numbers of reported features are 31 and 37 for ILI 1 and ILI 2, respectively.

The developed point matching model in Section 4.4 is applied to match the features from ILI 1 to ILI 2. The initial values for the parameters α and β in (10) are set to 10^{-4} and 200T, respectively. The initial temperature T_0 and final temperature T_{final} of the annealing process are set based on Section 4.5 to 2.4056 and 0.0004, respectively. The annealing rate λ is set to 0.96, which leads to L=216 steps in the annealing process to go from T_0 to T_{final}. The proposed algorithm in Section 4.5 is used to obtain the unknown transformation and correspondence. The entire analysis was executed within 3.1 s on a standard laptop.

The optimal transformation matrix $\boldsymbol{a}_{opt}$ at $T_{final} = 0.0004$ for $\alpha = 10^{-4}$ and β=200T is obtained using (22):

$$\boldsymbol{a}_{opt} = \begin{bmatrix} 1.0199 & 0.0132 & 0.0018 \\ 0.0216 & 1.0231 & 0.0027 \\ 0 & 0 & 1 \end{bmatrix} \tag{16}$$

The optimal $\boldsymbol{a}_{opt}$ shows that the transformation is close to identity as expected. The transformation consists of translation, which is described by $a_{13} = 0.0018$ and $a_{23} = 0.0027$, scaling, shear, and rotation, which are represented by the parameters a_{11}, a_{12}, a_{21}, and a_{22}. Fig. 10 shows the transformed feature locations based on (1) from ILI 1 and the reported locations from ILI 2. Comparing Figs. 9 and 10, the location transformation of $\widetilde{X}$ is noticeable. Most features from ILI 1 are matched well to the results from ILI 2.

Table 2 shows the correspondence matrix $\boldsymbol{p}$ and outlier correspondences $\boldsymbol{r}$ and $\boldsymbol{s}$ for the first eight features in both X and Y at the final temperature. The features are counted in the order of their axial direction; they are all located from 0.0 to 0.2 m in the axial direction of Fig. 10. Features i=3 and i=7 in X and k=1 in Y are identified as outliers that cannot be matched because $r_3 = r_7 = 1$ and $s_1 = 1$. The correspondence matrix $\boldsymbol{p}$ is mainly non-zero around the main diagonal; the feature pairs (i=5, k=6), (6, 7), and (8, 8) are matched with a probability above 0.95. The pairs (1, 3) and (2, 2), which are in close proximity to each other, still have matching probabilities of $p_{13} = 0.94$ and $p_{22} = 0.94$, respectively. Feature i=4 is not clearly matched to either feature k=4 or k=5.

Of particular interest for the feature-specific corrosion growth analysis and the integrity assessment of pipelines is the correspondence $\boldsymbol{p}$, which is used to identify the matched feature pairs. The transformation $\boldsymbol{a}$ in (16) is simply a by-product of the feature matching analysis, but it is very useful for visualizing the results. The subsequent corrosion growth analysis would rely on the feature pairs that are matched with a high probability (e.g. above 0.9) and the identified outliers would be removed from this analysis.

Fig. 11 shows the performance of the deterministic annealing by displaying the objective function E in (10) for $\alpha = 10^{-4}$ and $\beta = 200T$ as a function of the temperature T. Both axes in the figure are in log-scale and the horizontal axis is in reverse order. As shown, at the beginning of the annealing process the objective function slowly starts to decrease. At a temperature around 0.0026, a significant drop in E is

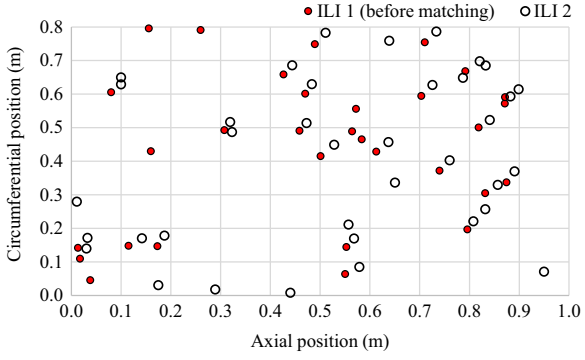

Fig. 9. Reported feature positions before matching.

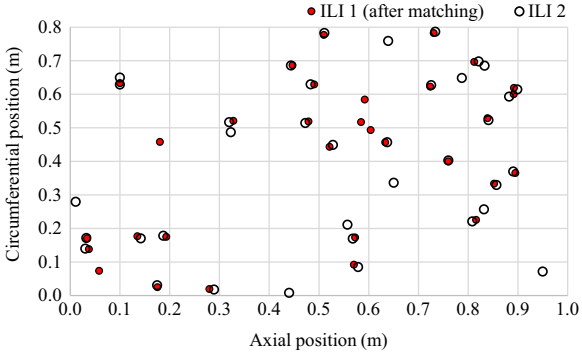

Fig. 10. Feature positions after matching ($\alpha = 10^{-4}$, $\beta = 200$ T, $T_{final} = 0.0004$).

Table 2
Correspondence matrix $\boldsymbol{p}$ and outlier correspondences $\boldsymbol{r}$ and s for the first 8 features from X (index i) and Y (index k) at final temperature $T_{final} = 0.0004$.

p_{ik}	k=1	2	3	4	5	6	7	8	r_i
i=1	0	0.06	0.94	0	0	0	0	0	0
2	0	0.94	0.06	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	1
4	0	0	0	0.59	0.41	0	0	0	0.01
5	0	0	0	0	0	0.99	0	0	0.01
6	0	0	0	0	0	0	0.99	0	0.01
7	0	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	0.99	0.01
s_k	1	0	0	0.41	0.59	0.01	0.01	0.01	

observed, presumably due to the identification of most matching feature pairs and a considerable improvement in the transformation. Afterwards, the objective function decreases at a higher rate than before the drop until the final temperature is reached.

Overall, the model identified 18 matching feature pairs in Fig. 10 with a probability of at least 0.9. Five and six outliers are found in X and Y with a probability of at least 0.9.

6. Conclusions

This paper introduces a comprehensive framework for the automated matching of corrosion features, which are obtained from ILI results. The framework replaces manual feature matching, which can be time intensive and prone to errors. It consists of five steps to reduce the complexity of a multi-ILI 3D matching problem to set of simpler 2D problems to match the features using (existing) 2D probabilistic point matching models, and to transfer the results back to the original matching problem.

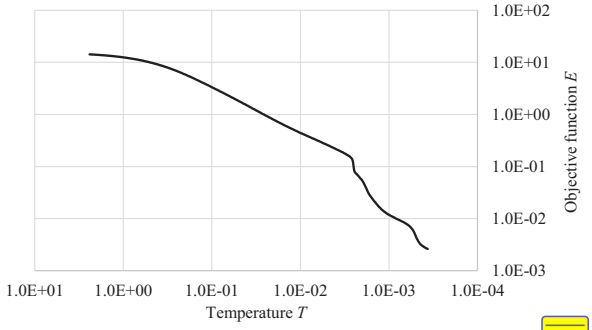

Fig. 11. Objective function as a function of temperature ($\alpha = 10^{-4}$, $\beta = 200$ T).

Reducing a multi-ILI 3D matching problem to a set of 2D matching problems makes the computation of the solution fast even for complex matching tasks. This breakdown is a clear advantage of the framework compared to manual matching. The framework can be adapted to a broad set of matching problems, and the 2D point matching model is governed by only three parameters: α to control the number of outliers, β to control the transformation close to identity, and T_{final} to control the probabilistic correspondence assignment. The results of the matching analysis can be visualized using the transformation obtained from the matching analysis for better verification of the final solution.

The matching uncertainty, which is inherent in any matching problem, is quantified in the correspondence matrix. The uncertainty can be directly included in the corrosion growth analysis for a consistent propagation. Detection and false call uncertainty still apply to the obtained matching results. For example, a separate analysis needs to determine whether the detected outliers are false calls or the missing counterparts have not been detected. The likelihood is high that the perfect matches are correctly identified, but some pairs may be false calls.

The next step is to relax the point assumption for the features and include the measured feature size in the matching analysis. This next step will help to improve our understanding of the features that are not perfectly matched. Additionally, detection and identification uncertainties can be incorporated to enhance the matching model.

Acknowledgments

The first author acknowledges the financial support from NSERC (RGPIN-2015-04135) and the University of Calgary (Grant no. 10008004).

Appendix

The objective function $E(\boldsymbol{a}, \boldsymbol{p}|\widetilde{X}, Y, \alpha, \beta)$ in (10) for matching points on a planar surface is a function of the transformation $\boldsymbol{a}$ and the correspondence $\boldsymbol{p}$ given the two point sets $\widetilde{X} = \{\boldsymbol{x}_i \text{ in } \mathbb{R}^2: i = 1, \ldots, 2n\}$ and $Y = \{\boldsymbol{y}_k \text{ in } \mathbb{R}^2: k = 1, \ldots, m\}$ and the two multipliers α and β to control the matching.

$$E(\boldsymbol{a}, \boldsymbol{p}|\widetilde{X}, Y, \alpha, \beta) = \sum_{i=1}^{2n}\sum_{k=1}^{m} p_{ik} \| \boldsymbol{y}_k - \boldsymbol{a}\boldsymbol{x}_i \|^2 - \alpha \sum_{i=1}^{2n}\sum_{k=1}^{m} p_{ik} + \beta \mathrm{tr}[(\boldsymbol{a} - \boldsymbol{I})^{\mathrm{T}}(\boldsymbol{a} - \boldsymbol{I})] \tag{17}$$

The deterministic annealing requires the conditional objective function $E(\boldsymbol{a}|\widetilde{X}, Y, \boldsymbol{p}, \alpha, \beta)$ to determine the optimal transformation:

$$E(\boldsymbol{a}|\widetilde{X}, Y, \boldsymbol{p}, \alpha, \beta) \propto \sum_{i=1}^{2n}\sum_{k=1}^{m} p_{ik} \| \boldsymbol{y}_k - \boldsymbol{a}\boldsymbol{x}_i \|^2 + \beta \mathrm{tr}[(\boldsymbol{a} - \boldsymbol{I})^{\mathrm{T}}(\boldsymbol{a} - \boldsymbol{I})] \tag{18}$$

According to Theorem 3.1 by Yang [37], Eq. (18) can be written as the following:

$$E(\boldsymbol{a}|\widetilde{X},\ Y,\ \boldsymbol{p},\ \alpha,\ \beta) \propto \sum_{i=1}^{2n} \frac{\| z_i - \boldsymbol{a}\boldsymbol{x}_i \|^2}{\delta_i^2} + \beta \mathrm{tr}[(\boldsymbol{a} - \boldsymbol{I})^{\mathrm{T}}(\boldsymbol{a} - \boldsymbol{I})] \tag{19}$$

where

$$z_i = \frac{\sum_{k=1}^{m} p_{ik}\boldsymbol{y_k}}{\sum_{k=1}^{m} p_{ik}} \quad \forall\ i = 1,\ \ldots,\ 2n \tag{20}$$

and

$$\delta_i^2 = \frac{1}{\sum_{k=1}^{m} p_{ik}} \quad \forall\ i = 1,\ \ldots,\ 2n \tag{21}$$

The optimal transformation $\boldsymbol{a_{opt}}$ conditional on $\boldsymbol{p}$ is determined by setting the first derivative of $E(\boldsymbol{a}|\widetilde{X},\ Y,\ \boldsymbol{p},\ \alpha,\ \beta)$ in (19) with respect to $\boldsymbol{a}$ to zero.

$$\boldsymbol{a_{opt}} = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}^{-1}\boldsymbol{x} + \beta m\boldsymbol{I})^{-1}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}^{-1}\boldsymbol{z} + \beta m\boldsymbol{I}) \tag{22}$$

where $\boldsymbol{w} = \mathrm{diag}(\delta_1^2,\ \delta_2^2,\ \ldots,\ \delta_n^2)$ and z is the stacked version of z_i.

References

[1] Walker J. In-line inspection of pipelines: advanced technologies for economic and safe operation of oil and gas pipelines. Munich, Germany: Verlag Moderne Industrie; 2010.
[2] Reber K, Beller M, Barbian AO. Run comparison: using in-line inspection data for the assessment of pipelines. In: Proceedings of the 2006 pipeline technology conference; 2006.
[3] API, ANSI/API Standard 1163: in-line inspection systems qualification standard. American Petroleum Institute; 2010.
[4] POF. Specifications and requirements for intelligent pig inspection of pipelines. Available online at ⟨www.pipelineoperators.org⟩; 2009.
[5] Madsen HO, Krenk S, Lind NC. Methods of structural safety. Mineola, US: Dover Publications; 2006.
[6] Melchers RE. Structural reliability analysis and prediction. Chichester, UK: John Wiley & Sons; 1999.
[7] Zhou J, Rothwell B, Nessim M, Zhou W. Reliability-based design and assessment standards for onshore natural gas transmission pipelines. J Press Vessel Technol 2009;131. http://dx.doi.org/10.1115/1.2902281, 031702-1-6.
[8] DNV. Recommended practice DNV-RP-F101: corroded pipelines. Det Norske Veritas AS; 2010.
[9] CSA. Oil and gas pipeline systems. National Standard of Canada; 2015.
[10] Pandey MD. Probabilistic models for condition assessment of oil and gas pipelines. NDTE Int 1998;31(5):349–58. http://dx.doi.org/10.1016/S0963-8695(98)00003-6.
[11] Bai Y, Bai Q. Subsea pipeline integrity and risk management. Waltham, US: Gulf Professional Publishing; 2014.
[12] van Noortwijk JM. A survey of the application of gamma processes in maintenance. Reliab Eng Syst Saf 2009;94(1):2–21. http://dx.doi.org/10.1016/j.ress.2007.03.019.
[13] Zhang S, Zhou W, Al-Amin M, Kariyawasam S, Wang H. Time-dependent corrosion growth modeling using multiple in-line inspection data. J Press Vessel Technol 2014;136. http://dx.doi.org/10.1115/1.4026798, 041202-1-7.
[14] Qin H, Zhou W, Zhang S. Bayesian inferences of generation and growth of corrosion defects on energy pipelines based on imperfect inspection data. Reliab Eng Syst Saf 2015;144:334–42. http://dx.doi.org/10.1016/j.ress.2015.08.007.
[15] Zhang S, Zhou W, Qin H. Inverse gaussian process-based corrosion growth model for energy pipelines considering the sizing error in inspection data. Corros Sci 2013;73:309–20. http://dx.doi.org/10.1016/j.corsci.2013.04.020.
[16] Bazan FAV, Beck AT. Stochastic process corrosion growth models for pipeline reliability. Corros Sci 2013;74:50–8. http://dx.doi.org/10.1016/j.corsci.2013.04.011.
[17] Zhang S, Zhou W. Bayesian dynamic linear model for growth of corrosion defects on energy pipelines. Reliab Eng Syst Saf 2014;128:24–31. http://dx.doi.org/10.1016/j.ress.2014.04.001.
[18] Pandey MD, Yuan XX, van Noortwijk JM. The influence of temporal uncertainty of deterioration on life-cycle management of structures. Struct Infrastruct Eng 2009;5(2):145–56. http://dx.doi.org/10.1080/15732470601012154.
[19] ASME. Manual for determining the remaining strength of corroded pipelines: supplement to ASME B31 code for pressure piping (ASME B31G). The American Society of Mechanical Engineers; 2012.
[20] Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. Comput Vis Image Underst 2003;89:114–41. http://dx.doi.org/10.1016/S1077-3142(03)00009-2.
[21] Hibbard LS, Hawkins RA. Objective image alignment for three-dimensional reconstruction of digital autoradiograms. J Neurosci Methods 1988;26(1):55–74. http://dx.doi.org/10.1016/0165-0270(88)90129-X.
[22] Stockman G. Object recognition and localization via pose clustering. Comput Vis Graph Image Process 1987;40(3):361–87. http://dx.doi.org/10.1016/S0734-189X(87)80147-0.
[23] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the hausdorff distance. IEEE Trans Pattern Anal Mach Intell 1993;15(9):850–63. http://dx.doi.org/10.1109/34.232073.
[24] Ullman S. Comparing images using the hausdorff distance. Cognition 1989;32(3):193–254. http://dx.doi.org/10.1016/0010-0277(89)90036-X.
[25] Lamdan Y, Schwartz JT, Wolfson HJ. Object recognition by affine invariant matching. In: Proceedings of the computer science conference on computer vision and pattern recognition; 1988. p. 335–344. http://dx.doi.org/10.1109/CVPR.1988.196257.
[26] Grimson WEL, Lozano-Perez T. Localizing overlapping parts by searching the interpretation tree. IEEE Trans Pattern Anal Mach Intell 1987;PAMI-9(4):469–82. http://dx.doi.org/10.1109/TPAMI.1987.4767935.
[27] Feldmar J, Ayache N. Rigid, affine and locally affine registration of free-form surfaces. Int J Comput Vis 1996;18(2):99–119. http://dx.doi.org/10.1007/BF00054998.
[28] Metaxas D, Koh E, Badler NI. Multi-level shape representation using global deformations and locally adaptive finite elements. Int J Comput Vis 1997;25(1):49–61. http://dx.doi.org/10.1023/A:1007929702347.
[29] Szeliski R, Lavallee S. Matching 3-d anatomical surfaces with non-rigid deformations using octree-splines. Int J Comput Vis 1996;18(2):177–86. http://dx.doi.org/10.1007/BF00055001.
[30] Sclaroff S, Pentland AP. Modal matching for correspondence and recognition. IEEE Trans Pattern Anal Mach Intell 1995;17(6):545–61. http://dx.doi.org/10.1109/34.387502.
[31] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models-their training and application. Comput Vis Image Underst 1995;61(1):38–59. http://dx.doi.org/10.1006/cviu.1995.1004.
[32] Shapiro LS, Brady JM. Feature-based correspondence: an eigenvector approach. Image Vis Comput 1992;10(5):283–8. http://dx.doi.org/10.1016/0262-8856(92)90043-3.
[33] Amit Y, Kong A. Graphical templates for model registration. IEEE Trans Pattern Anal Mach Intell 1996;18(3):225–36. http://dx.doi.org/10.1109/34.485529.
[34] Lohmann G, von Cramon DY. Automatic labelling of the human cortical surface using sulcal basins. Med Image Anal 2000;4(3):179–88. http://dx.doi.org/10.1016/S1361-8415(00)00024-4.
[35] Shapiro LG, Haralick RM. Structural descriptions and inexact matching. IEEE Trans Pattern Anal Mach Intell 1981;PAMI-3(5):504–19. http://dx.doi.org/10.1109/TPAMI.1981.4767144.
[36] Besl PJ, McKay ND. A method for registration of 3-d shapes. IEEE Trans Pattern Anal Mach Intell 1992;14(2):239–56. http://dx.doi.org/10.1109/34.121791.
[37] Yang J. The thin plate spline robust point matching (tps-rpm) algorithm: a revisit. Pattern Recognit Lett 2011;32(7):910–8. http://dx.doi.org/10.1016/j.patrec.2011.01.015.
[38] Cross ADJ, Hancock ER. Graph matching with a dual-step em algorithm. IEEE Trans Pattern Anal Mach Intell 1998;20(11):1236–53. http://dx.doi.org/10.1109/34.730557.
[39] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 2002;24(4):509–22. http://dx.doi.org/10.1109/34.993558.
[40] Gold S, Rangarajan A, Lu C-P, Pappu S, Mjolsness E. New algorithms for 2d and 3d point matching: pose estimation and correspondence. Pattern Recognit 1998;31(8):1019–31. http://dx.doi.org/10.1016/S0031-3203(98)80010-1.