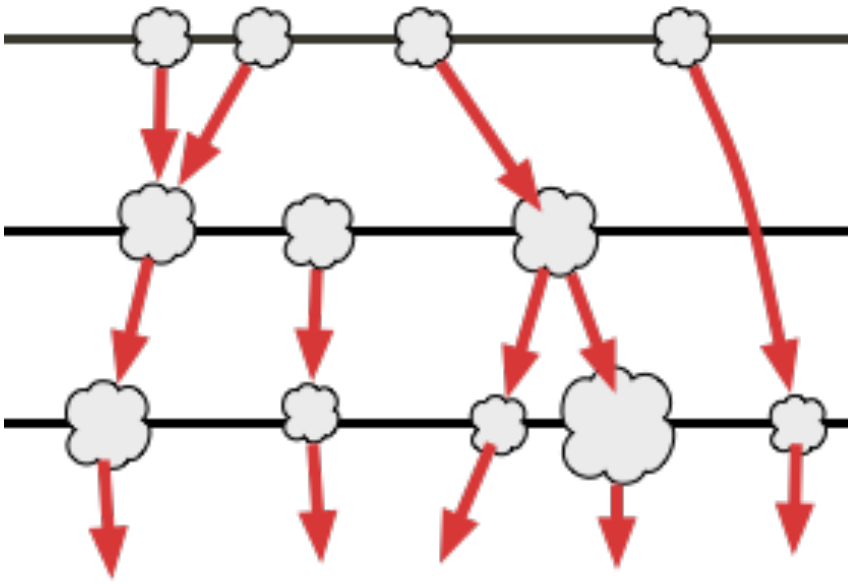
COLORADO SCHOOL OF MINES

COLORADOSCHOOLOFMINES

EARTH • ENERGY • ENVIRONMENT



Final Project Report

Robust Anomaly Matching for ICIPs: Reducing Pipeline Assessment Uncertainty Through 4-Dimensional Anomaly Detection and Characterization

Prepared for: Department of Transportation

Prepared by: Colorado School of Mines; Craig Champlin and John P H Steele

December 13, 2016

CAP Award: DTPH5614HCAP06

EXECUTIVE SUMMARY

Objective

Current models for corrosion growth within a pipeline assume a single rate along a pipeline's length. These rates tend to be too conservative for most of the pipe, and yet in some cases fail to account for regions of excessive growth. Most anomalies grow at slower than assumed rates, resulting in pipes being replaced sooner than necessary. At the same time isolated regions of unusually aggressive corrosion can fail when corrosion growth exceeds calculated reinspection intervals. The project supports the direct measurement of local corrosion growth rates by monitoring the rate of change of individual anomalies as seen by internal pipeline inspections.

Goals

First, most pipeline operators rely on commercially available high resolution pipeline inspections by smart pigs. The vendors who operate these tools deliver inspection reports whereby individual anomalies have been called out as sized boxes with a depth. Matching these called out anomalies across inspections is a tedious process. The goal is to define the problem and develop an algorithm based on this analysis.

Second, there is an emerging class of cleaning and inspection tools, ICIPs, which deliver low resolution scans of the pipe interior during frequent cleaning operations. In the interest of rapid turn-around times, these tools return raw signals to the operator. Matching anomalies with these tools is a question of matching raw signals then identifying changes which may have occurred since the last inspection. The goal is to develop an algorithm which can align two raw inline inspection sensor signals in the face of irregular velocities.

Solution

The team intended to develop two algorithms for matching anomalies across coincident internal pipeline inspections. The first algorithm would match boxed anomalies. The second algorithm would match raw signals. The goal for each algorithm is slightly different. The boxed algorithm is intended to do a complete mapping of individual called-out anomalies from one inspection to the next. The raw signal algorithm velocity corrects and aligns raw inspections signals.

Report Content

This report presents an overview of the challenges encountered in doing anomaly alignment over multiple pig-run scans. In addition, it provides insights into the applicability of robotic SLAM techniques to pipeline anomaly alignment.

INTRODUCTION

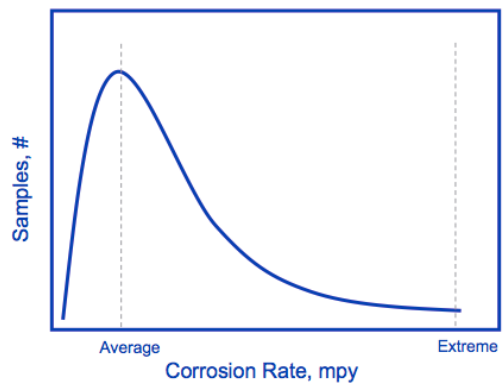
The PHMSA CAAP award funded a student at the Colorado School of Mines to develop computer algorithms for automatically matching anomalies across consecutive inline inspections of oil and gas pipelines. These algorithms are related in that they both intended to align pipeline inspections for measuring local corrosion growth rates.

This introduction section discusses the use of aligned anomalies for determining local corrosion growth rates. Next, the report will define the distinct, but related, problems each of the two algorithms would solve. The first algorithm, called BACO, matches boxed-and-called-out anomalies from vendor-supplied inspection spreadsheets. The second algorithm, VVFM (variable velocity feature matching), algorithm would alignment points for raw signal matching in support of the emerging class of integrated cleaning and inspecting pigs – ICIPs, which are discusses in a later section.

MOTIVATION

Real Corrosion Growth Rates

It is well known that corrosion growth rates differ along a pipeline's length. Many people in the pipeline industry are familiar with corrosion rate distribution curves as seen in Figure 1 **Corrosion Rate Distribution** (Moghissi, 2008). This curve matches a statistical distribution pattern know as an extreme value distribution. Examining this figure which represents the distribution of corrosion rates as measured at discrete points along a pipe's length, we see the x-axis representing real corrosion rates. The origin represents no corrosion. As we move to the right along this axis, we see faster and faster local corrosion growth rates. The y-axis represents population - the number of samples experiencing a specific local corrosion rate.



Corrosion Rate Distribution: Extreme value distribution of measured corrosion rates.

Figure 1 Corrosion Rate Distribution

Most of the corrosion rates seen along a pipeline fall into the fat part of the curve on the left side. On this graph, there is a vertical dotted line representing the average measured corrosion rate. This average is slightly to the right of the peak of the distribution. The right side of the curve represents the extreme value portion. There is a non-zero probability that any arbitrarily large corrosion rate exists within the population as the distribution approaches but never quite reaches zero on the right side of the curve.

When one examines an individual corrosion pit, the question that is asked *it is growing*. If it is growing, how quickly? Where does this data point lie on the corrosion rate distribution curve? How long until the pipe must be replaced? From the distribution curve, it is most probable that any given anomaly is growing very slowly, but one cannot be sure. Without local corrosion measurements, it is impossible to know how fast any specific anomaly is growing.

The standard approach is to use an assumed rate. If a slow rate is selected, there is a finite chance that somewhere the pipe is corroding faster than this rate. If a faster rate is selected, one penalizes the large portion of the population that is corroding slowly. Because it is possible that any arbitrarily large corrosion rate can exist in the pipeline, regardless of what rate is selected there is a non-zero probability that somewhere along the pipe there exists a faster rate. There is some chance that some point on the pipe will corrode faster than expected. Thus, while assumed rates tend to be too conservative, there is still the risk of missing extreme values. (McNealy et al. 2012)

The only real solution to this conundrum is to use locally measured corrosion rates along the pipeline's length. By doing this one can tailor the reinspection intervals to data-supported values. Most of the pipe will be maintained using slow corrosion growth rates. This allows us to confidently leave pipe alone for longer periods of time. Some small percentage of pipe will require more frequent attention in response to the measured fast corrosion rates. Thus, one can operate safely by maintaining these sections more aggressively. Knowing real corrosion growth rates provides for the optimally use of maintenance budgets.

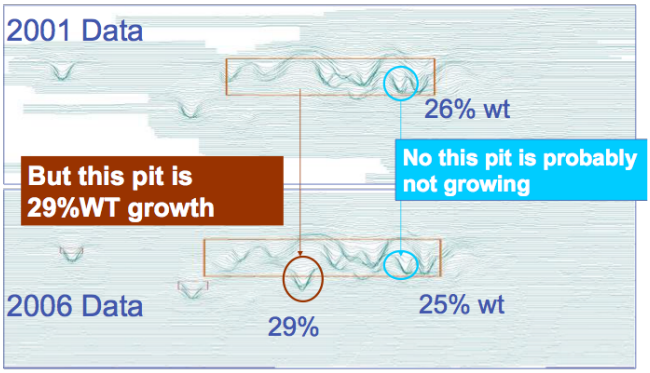
It would appear that it's relatively straight forward to obtain real corrosion growth rates along the length of a pipeline. It is a simple matter of tracking changes in anomaly size over time. One can watch how anomalies found in inline inspections change over time. If the pipes are being inspected at regular intervals of every several years, it should be relatively easy to compare the size of anomalies from one inspection to the next and report the change in size as a function of the time between inspections as the anomaly's growth rate. Indeed, many operators and vendors have noticed the theoretical merit of this approach. (PHMSA waar 2008),

In practice, however, this approach can be complicated and time consuming. One vendor, who provides the match up corrosion anomalies service, describes the process as being labor intensive, e.g., taking months to align anomalies between two inspections (Yeagley and Madden 2014). Also, when there are many anomalies in the same area it can be hard to figure out which anomaly goes where. Finally, differences between tools and even between inspections can obscure correct anomaly matches. In short, the corrosion anomaly matching process is more complicated than it may appear at first glance. In addition, because of the relatively long intervals between high resolution pipeline inspections, (on the order of multiple years), it can

be the case that too few spatiotemporal data points are available to provide an accurate representation of local corrosion rate.

Raw Signals vs Boxed Anomalies

Pipeline inspection vendors usually return inspection results as a spreadsheet of anomalies. These inspection reports "call out" each anomaly as a box with a depth at locations along the pipe. Creating a box from the raw signal requires that the vendor does some signal interpretation. Figure 2, **Inspection Signals,** shows the raw signal from two inspections, one taken in 2001 and one taken in 2006. Typical returned boxes are overlaid on top of these raw signals. This gives us some idea of what the raw signal underlying "boxed and called out" anomalies might look like.



Inspection Signals: Signal and box matching across two inspections. The boxed anomaly is reported as 3% WT growth. (Kevin Spencer 2008))

Figure 2 Inspection Signals

In the raw signal of this figure we have the pig's direction of travel from left to right along the pipeline's length. The radial direction is up and down along the y-axis. Each blue scan line is the sequence of values recorded by a single sensor as the pig travels along the pipe in the x direction. When the pipe wall material is good and all sensors are operating correctly, we see a featureless, uniform baseline signal level recorded as a straight line. When there is a detectible feature, sensor outputs change from this baseline value. When a sensor goes offline, we see a bare spot in the data. There are several bare spots in the 2001 data.

To create boxed anomalies an inspection vendor runs an algorithm which processes the raw signal. This algorithm looks for contiguous regions where raw sensor readings differ from the baseline levels by some threshold value. When such a region is detected this algorithm draws a bounding box around it and establishes its depth by looking for the deepest point within the box. We see two such red bounding boxes in the figure, one in the 2001 data and one in the 2006 data. These two boxes represent the same boxed-and-called-out anomaly as observed by two coincident inspections - inspections taken at different times, over the same segment of pipe. These bounding boxes represent an "anomaly matching" across two inspections.

Most inline inspection vendors return only the boxes shown in this figure, not the signal. We can see that although these boxes are slightly different sizes, they are at roughly the same position both axially and radially. The reported depth of the 2001 box is 26% of the wall thickness. The reported depth of the 2006

box is 29% of the wall thickness. If we had been given only the boxes from these two inspections and did not have the benefit of seeing the raw signal underlying them, we might assume that this boxed anomaly had grown by a modest 4% WT over the 5 years between these inspections. By looking at the raw signal, however, we see that the 29% WT pit has sprung up virtually instantaneously between the two inspections. Instead of an anomaly which is growing at a modest rate, we have an extremely rapid corrosion growth area. Instead of a region of no concern, as suggested by the boxed anomalies, we see by the raw signal that we have a region of serious concern. Clearly in this situation we would make better maintenance decisions from the raw signal than we would make from the reported boxes. It is desirable to use raw signals.

While we can see the benefit of matching raw inspection signals, at the time of this writing these raw signals are rarely available from inspection vendors. For several reasons inspection vendors almost exclusively return spreadsheets of boxed anomalies. This limitation is of no great concern, however, when it comes to equating anomalies from one inspection to the next. We will perform matching of boxed anomalies. A raw signal is essentially a boxed anomaly where we know what is in the box. If we can match boxes, we can draw boxes around a raw signal and match them too. Because there is less information in vendor-supplied boxes, matching empty boxes is the harder problem.

Anomaly Matching Problem

The anomaly matching problem maps zero or more anomalies in a first inspection to zero or more anomalies in a second inspection across the same physical joint of pipe. Ultimately, we will use these matched anomalies to extract a corrosion growth signal. On its face the anomaly matching problem is quite simple. If there is an anomaly in the same position in two different inspections, then it is the same anomaly.

Things are never simple. Three elements conspire to make anomaly matching difficult:

1) Anomaly density - anomaly identity gets confusing in a crowd of other anomalies
2) Sensing differences - different sensors and sensing conditions return different signals
3) Positional uncertainty - we estimate tool location, these estimates have errors

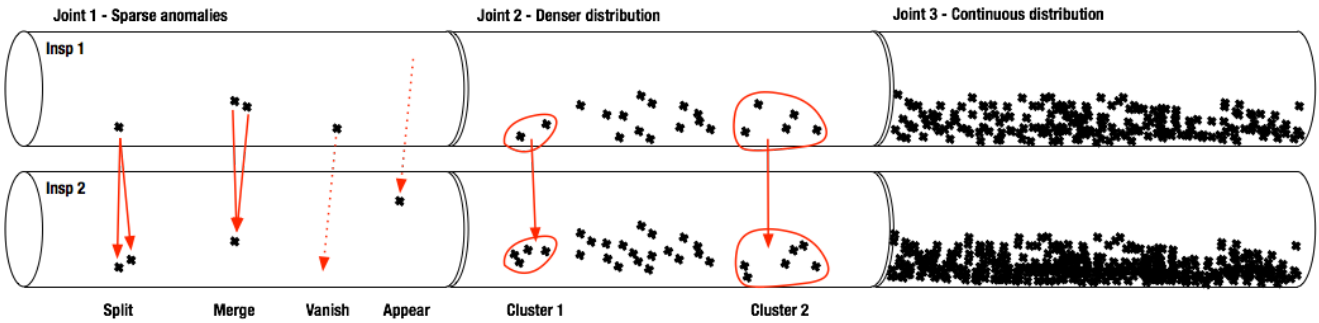
Each of these elements conspire to obscure the identity of individual anomalies from one inspection to the next. It is these ambiguities of anomaly identity which bog down manual attempts to match anomalies and cause anomaly matching to be an inexact and laborious process. In fact, one commercial anomaly matching service strives to simply match the deepest anomaly per joint (Yeagley and Madden 2014). This deepest measured anomaly is not likely to be the same anomaly from one inspection to the next. How valid is such a corrosion rate?

This section of the report will establish the challenges inherent in the anomaly matching problem. We will describe each of these three challenges in detail. We will present a morphology of anomaly matches. We will introduce a vocabulary which we can use to discuss the challenges faced when tracing an anomaly's identity across multiple pipeline inspections. We will only describe the structure of these differences without discerning the cause of these differences. These concepts apply whether the bounding box is a simple shape, as a boxed anomaly, or is a complex shape, as a raw signal.

Anomaly Density

Consider Figure 3, **Anomaly Matching Problem**. This figure presents several anomaly-matching examples. In it we see two pipeline inspections separated in time. Inspection 1 is the three joints at the top of the figure and Inspection 2 is the three joints in the lower half of the figure. We require that Inspection 1 occurs in time before inspection 2. Each inspection traverses from left to right through three joints separated by two girth welds. We have correctly aligned each joint using known features like above-ground markers, valves, and weld counts. In this figure, the small black crosses indicate anomaly locations but not their geometry.

Each successive joint in the example pipe has increasing anomaly density. These three joints represent slices across an anomaly density continuum from no anomalies on one extreme to completely covered in anomalies on the other extreme. We see that Joint1 has very few anomalies, Joint 2 has an intermediate anomaly density, and Joint 3 has almost continuous corrosion along the bottom of the pipe.



Anomaly Matching Problem: Two inspections across three joints showing impact of anomaly density on the anomaly matching problem. Joint 1 has low density. Anomalies may easily be matched by inspection. Joint 2 has a moderate density. The best anomaly matches are not always apparent. Joint 3 is heavily corroded. Anomaly matching here is more complex.

Figure 3 Anomaly Matching Problem

Before we consider the impact of anomaly density let us use Joint1 to help us establish a vocabulary to describe the morphology of anomaly matches across repeated inspections. Referring to the figure, we begin

by moving from left to right along this joint. The first set of diverging red arrows show what was reported as a single anomaly in the first inspection reported as two anomalies in the second inspection. When this happens, we will say that the single anomaly has "split" into two anomalies even though this splitting is more of a reporting inconsistency than a reflection of what is happening inside the pipe.

Next, we see what was reported as two anomalies in the first inspection converge into one anomaly in the second inspection. It appears as if these two anomalies "merge" in the second inspection. This merging could be due to anomalies physically growing such that the boundary between them blurs. This merging is just as likely to be the result of reporting differences between the two inspections. In either case, what was reported as multiple anomalies in the older inspection, Inspection 1, appear to be a single anomaly in the more recent inspection, Inspection 2.

The next two sets of arrows on Joint1 show disappearing and appearing anomalies. First, we see an anomaly which was reported in Inspection 1 which seems to "vanish" in Inspection 2. Next, we see an anomaly which seems to suddenly "appear" between the first and second inspection. Although we will use this terminology in the context of matching, these anomalies haven't really disappeared or appeared. The most reasonable explanation is that we are seeing differences due to both sensing and reporting between these two inspections. We will discuss this in more detail in "Sensing differences" subsection below.

Now that we have some vocabulary to talk about anomaly matching, let us return the attention to the impact of anomaly density on forming matches. Observe how the nature of the anomaly matching problem changes with increasing anomaly density. In Joint 1 it is easy to match anomalies by simple observation. Anomalies which have similar position in the joint are probably the same. Even if anomalies merge or split or shift their position slightly we can easily identify analogous regions by where they are in the pipe and where they are in relation to neighboring anomalies.

Contrast the ease with which we can match up anomalies in Joint 1 with how difficult it is to see matches in Joint 3. In Joint 3 we see a near continuous distribution of anomalies. Pipe joints with this high level of corrosion may confound the matching process. It may be a misuse of time to try to match every anomaly. At these high densities, it may be more meaningful to match some identified subset, such as deepest 10% of anomalies.

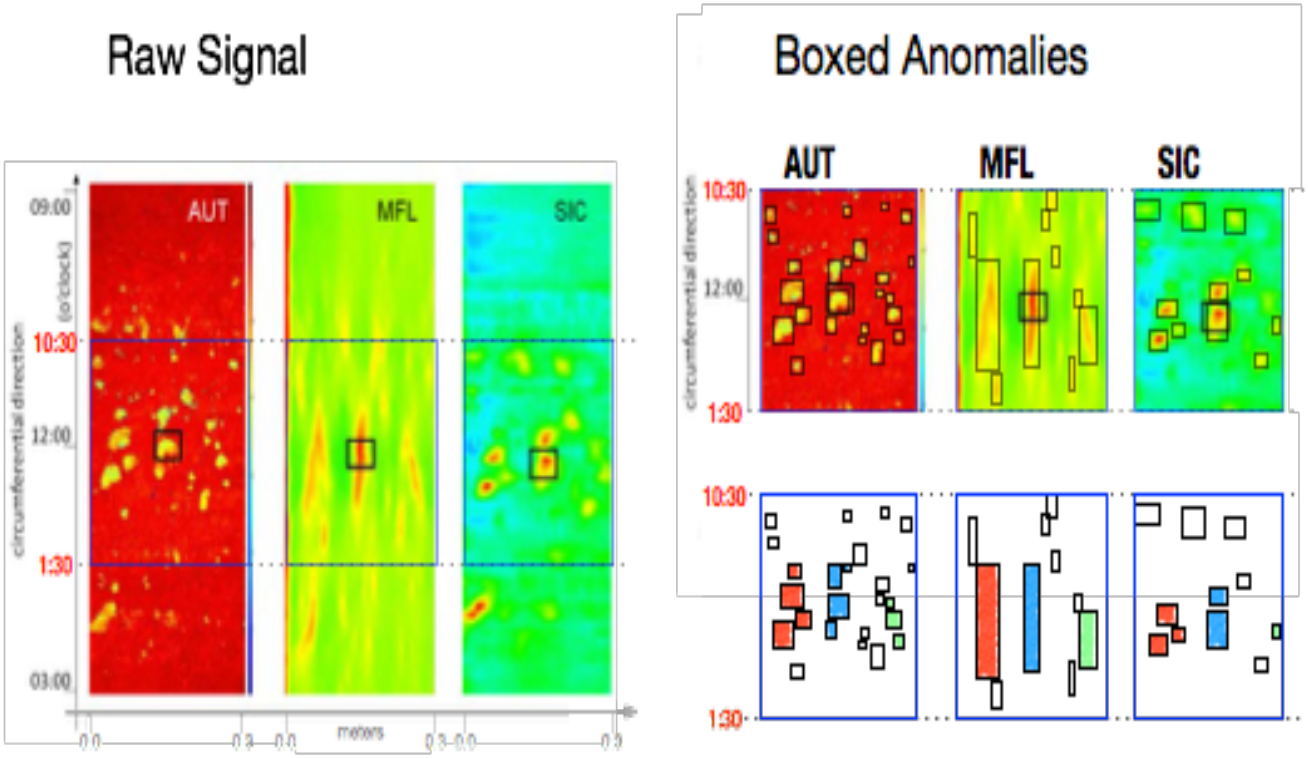
Finally explore the non-trivial case of Joint 2 where anomaly density is too great to match by observation, as we did with Joint 1, but anomaly density is not so dense that individual anomalies are essentially indistinguishable as was the case with Joint 3. Here the matching process becomes more complicated. We have circled two clusters of anomalies, Cluster 1 and Cluster 2. Not only is it less than obvious where to draw these clusters, within each cluster it is not obvious which anomalies are equivalent between the first and second inspections.

The focus of the anomaly matching problem is on pipes like Joint 2. These intermediate anomaly densities present the greatest challenge since mappings from one inspection to the next are not obvious. In Cluster 1 we see the 2 anomalies reported in inspection 1 reported as 4 anomalies in inspection 2. How do we match these? Do we pick the single best match across the two inspections? Inspection 1 shows an anomaly at the top of Cluster 2 which appears to vanish in Inspection 2. Is there no direct equivalent of this anomaly in inspection 2? Is it reported as multiple anomalies in inspection 2?

In this report thinking we will focus on pipes that look like Joint 2. We are concerned with nontrivial amounts of corrosion over regions of pipe where corrosion is not continuous. Establishing exact and unambiguous matches in this density regime is non-trivial and would benefit from the logically consistent repeatability afforded by assigning matches with a computer algorithm.

Sensing Differences

While the anomaly density of an inspection is a function of the number of anomalies in a joint, it is also a function of the sensors used in the inspection. For example, when inspection vendors report results they establish a baseline signal level. Signals below this chosen level are reported as being anomaly free, while signals above this level are reported as anomalous. Increasing or decreasing this level will cause anomalies to seem to disappear and appear as threshold levels are increased and decreased. Anomalies will seem to change shape as shallow regions of the anomaly go in and out of scope. Thus one parameter can have a profound impact on the nature of reported anomalies. Other parameters such as sensor calibration, sensor size, scanning speed, and different processing algorithms can all affect the reported results.



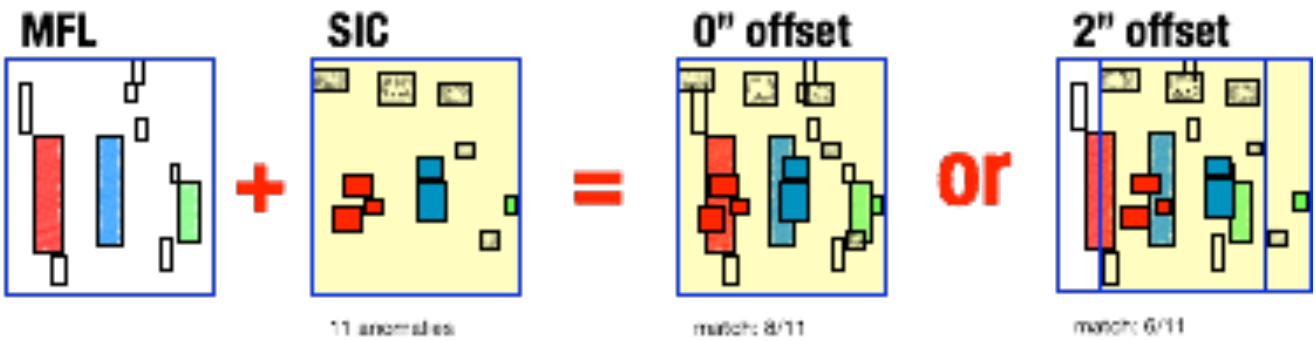
Signal Differences: [Left] The raw signal of three sensors showing the in the top 60º over the same 0.3 meter section. (From: (Huyse 2010)) **[Right]** The raw signal in the top 30º (between 10:30 and 1:30) of the same 0.3 m long section of pipe. For the top set we have overlaid bounding boxes onto each anomaly. For the lower set, we have removed the raw

Figure 4 Signal Differences using AUT, MFL, & SIC

Now we will look at an example of differences due to different sensor technologies. Consider Figure 4 **Signal Differences**. On the left side of this figure are the raw signal of three different sensors over the same 0.3 meter by 180º (from 9:00 to 3:00) section of pipe. The sensors from left to right are: AUT (automated ultrasonic testing), MFL (magnetic flux loss), and SIC (shallow internal corrosion - an eddy current technology by Rosen). Scans have been carefully controlled so that sensor positions are well known and are directly comparable. The black box shows the same physical location on all three scans.

The color values in these scans represent signal intensity which correlates to anomaly depth. Because of scaling differences, these intensity/depth values cannot be compared directly from sensor to sensor. In contrast, anomaly shape and position can be compared because the anomaly boundary is independent of the

signal values. Anomaly boundary is a threshold difference in signal intensity from the background signal in regions of unblemished pipe.



Positional Uncertainty: Overlaying the SIC scan onto the MFL scan. We have shaded the SIC boxes so that they can be more easily seen. The matching scores are the number of SIC boxes matched to MFL boxes over the total number of SIC boxes.

Figure 5 Comparison of positional uncertainty

The right side of this figure shows a zoomed in view of the original scans. Here we see the top 90º (from 1:30 to 10:30). The top portion continues to show signal intensity as color. A bounding box has been drawn around each anomaly using these intensity values. In the lower half of the right-side figure, we have removed the signal intensity information so that we can focus exclusively on the box patterns produced by these three sensor types. Examining the resulting box patterns, we see that AUT, which produces the highest resolution signal, produced 24 boxes. The next highest resolution signals is from the SIC sensor which produced 11 anomaly boxes. Finally, the lower resolution MFL scan in the center produced 10 boxes.

Superficially these patterns look quite dissimilar. If we did not know in advance that these represent the same section of pipe and know how these patterns align, it would not be obvious how these patterns match up. Matching these patterns by hand would be a time-consuming process which may not be repeatable from one trial to the next. Look closely at the box patterns generated by the MFL sensors in the lower-center image. We have assigned a color to the three largest anomalies - from left to right we have assigned the colors red, blue and green.

The single red anomaly in the center-left of the MFL scan corresponds to the left-most set of four anomalies also colored red in the AUT scan and the left-most set of three anomalies in the SIC scan. The center-most anomaly which is colored blue in the MFL scan corresponds to the center group of four anomalies in the AUT scan and the center group of two anomalies in the SIC scan. Finally, the right-most green anomaly in the MFL scan corresponds to the three anomalies which have been colored green on the AUT scan and the single

green anomaly on the center-right of the SIC scan. What the MFL sensor shows as three anomalies, the AUT sensors show as 11 anomalies and the SIC sensors show as 6 anomalies.

This example shows that even when the sensor location is well known it can still be difficult to match individual anomalies across successive inspections. More generally, differences between individual sensors conspire to make anomaly matching difficult.

Positional Uncertainty

No matter if the internal inspection is taken by a pig or a crawler, the position of the sensor in an operating pipeline is an estimate. While there are occasional checkpoints inside the pipe - valves, aboveground markers, et al - as soon as we move away from these known points the location of the sensors becomes uncertain. Location is estimated by fusing the input from one or more distance sensors - odometer wheels, inertial measurement units (IMUs), etc. While one might assume that combining the input from multiple odometer wheels or from an IMU would give a pretty good estimate of position, in practice wheels slip and IMU drift results in some error that accumulates over long distances inside pipelines. According to a study done in 2010, commonly used sensor fusion techniques for estimating pig position predict location with about 0.4% error (Santana 2010). This sounds pretty good until you consider that this translates to a positional error of about 2.2 inches over a typical 40 foot pipe joint[1] which suddenly seems pretty big when compared to a 1 inch corrosion pit.

Consider an example of the impact of this uncertainty on the real corrosion signal we looked at earlier. How much impact will uncertainty of 2.2 inches per joint have on the ability to match anomalies? Consider the two boxed signals from the previous section as shown in Figure 5 Comparison of positional uncertaintyFigure 5 **Positional Uncertainty**. The sample in this figure is 0.3 meters long, about 12 inches. This figure shows the MFL and the SIC signals. The SIC signal is shaded with light yellow and filled in with the non-colored boxes so one can see its contribution when it's overlayed onto the MFL signal. After the equals sign in the figure are two cases, 0" offset and 2" offset.

For the first case, labeled "0" offset", the two signals are directly overlaid. This represents the case where there is no positional uncertainty. The 6 red, green, and blue anomalies in the SIC scan match the 3 red, green, and blue anomalies from the MFL scan. In addition to these anomalies, also there are two non-colored anomalies from the SIC scan which match the MFL scan. This gives a total of 8 matched anomalies from the SIC scan out of 11 anomalies total resulting is a match score of 72%.

[1] A pipe joint is defined as a continuous section of pipe between circumferential girth welds. While there is no set standard joint length, 40 foot joints are commonly used because this length can be conveniently transported on a flatbed truck.

Now consider the second case labeled as the 2” offset. This case represents the linear uncertainty near the end of a joint. Here the SIC signal is shifted by about +2”. However, there are still matches. The red SIC boxes match the blue MFL boxes, giving 3 matches. The blue SIC boxes match the green MFL box and an empty MFL box, giving us 2 more matches. Finally, the top-center uncolored SIC box has an MFL match. This gives us a total of 6 of the SIC boxes which have found matches for a matching score of 54%.

The point of this example is that a reasonable metric of match quality can be fooled by a coincidental positional error. While in this example the 54% offset match is not as good as the 74% no-offset match, it is still a pretty good match. It is conceivable that, given a different pattern of anomalies, we could see more matches with the offset case than with the non-offset case. Since matches are only correct for the no-offset case, positional uncertainty can complicate the matching process with incorrect answers.

Since the goal for anomaly matching is to establish local corrosion rates, we need better than simply consistent matching. We need a consistently correct anomaly matching. We should measure correctness of matching against externally established anomaly identities. We can gauge the effectiveness of an algorithm by comparing its matching results against a laboriously established ground truth dataset.

In addition to linear uncertainty, the pig also spins while it travels introducing rotational uncertainty.

The overlaying of the SIC scan from this figure onto the MFL scan. On the 0” offset overlay image we see that we are able to match 8 of the 11 SIC anomalies in this region of the scan with MFL features.

Imagine if instead there is error in the x-axis location. The MFL scan is 0.05 meters, about 2 inches, to the left of where we think it is. Consider only the matching between the MFL scan and the SIC scan. Now the blue MFL box is closer to the red SIC box. The green MFL box matches to the blue SIC box. The red MFL box is unmatched and the green SIC box just looks like noise. A match scoring scheme shows that this match is as good as the correct one. This scenario is what actually happens.

Consider what would happen if the signals from the figure **Signal Differences** were off by about 2 inches. Recall that in this figure we compare boxed anomalies across three different sensor types.

Since there is uncertainty in the location of the pig, anomalies are shifted both linearly and rotationally. This makes the matching much more difficult. In this example anomaly location has been carefully controlled, in a free-swimming pig anomaly location is less certain. So, in addition to differences in reported anomaly shape if any of these signals were shifted forward or backward by 0.05 meter the patterns would overlay the wrong vertical anomaly cluster. These spatiotemporal considerations apply whether matching boxed raw signals or vendor-supplied boxed-anomalies. To match raw signals, we must ultimately box individual features. This

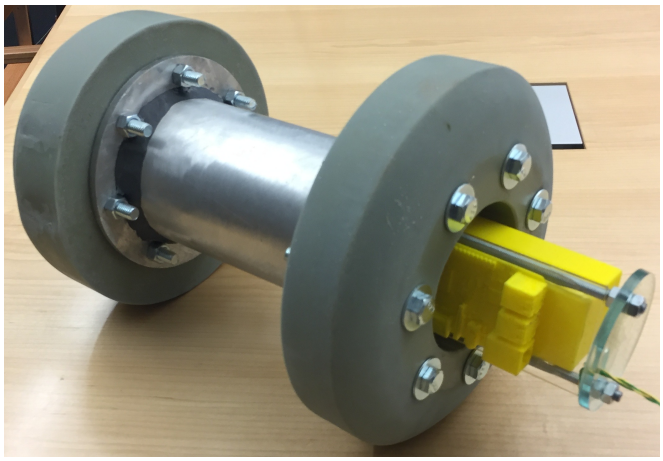
makes the anomaly matching problem identical for both boxed anomalies and raw signals, just different in the size and number of the features that need to be matched.

Matching across sensor types is even more challenging because one set of sensors sees two boxed anomalies - one deep and one shallow, a second vendor may see them as a single deep anomaly. A third vendor may completely filter out the anomaly and report nothing. The goal is to map one or more anomalies in one survey to zero or more anomalies in a second survey.

What will not change is the relative location of groups of anomalies as well as the relative location of anomalies within the group.

Integrated Cleaning and Inspecting Pigs (ICIPs)

In 2010 PHMSA recognized the need to provide more frequent measurements along the pipeline. They awarded work to Electricore, Inc. to investigate the feasibility of an integrated internal inspection and cleaning tool, ICIP (PHMSA integrated pipeline cleaning and inspection tool 2010). This work, which was rated "very effective" in its peer review, has generated some industry interest. Although PHMSA canceled the project, both PRCI and Chevron expressed interest in continuing the work because it "showed considerable near-term promise and benefits to the pipeline industry" (J. O. Brien et al., 2012). Currently there are several commercial ICIP offerings being developed by independent vendors.



CSM ICIP: The 8" ICIP developed by the Colorado School of Mines senior design team. This ICIP has sensors embedded in the trailing disk on the left. The yellow electronics package is exposed for reference.

Integrated Cleaning and Inspection Pigs (ICIPs) are cleaning pigs which have been instrumented with inexpensive, ruggedized sensors. The ICIP was developed by PHMSA to offer operators a low-cost and self-reliant alternative to complex high-resolution tools. These low-cost, low-resolution devices are designed to be run frequently by operators as part of their normal cleaning operations or as part of a routine maintenance program (PHMSA ICIP 2012). While using these devices does not preclude the need for occasional high-resolution inspections, they do allow an operator to maintain near-real-time eyes on the conditions of the pipe in the intervening years between traditional inspections.

One potential benefit of these tools is that frequent pipeline inspections allow an operator to observe the impact of operational changes on corrosion rates. An operator can measure how changing corrosion inhibitor or how adjusting water separation techniques impact local corrosion rates. Instead of a handful of corrosion coupons which may or may not be representative of conditions in the rest of the pipe, ICIPs allow operators more refined monitoring of pipeline conditions. To support this use-case an operator needs to be able to rapidly process the data collected by an ICIP run so that it can be combined with previous inspection. This needs to occur before the next scheduled routine ICIP inspection. A reasonable target would be to completely process an ICIP inspection in a week with minimum human intervention.

As a usage scenario, imagine that an operator has just completed an ICIP inspection. A technician extracts the tool from the pipeline and scrapes away cleaning debris. She brings her laptop computer within range and pairs her computer to the tool via Bluetooth radio. Once paired, the ICIP streams its data to the laptop where it is received by data processing software. As it receives this data, the software matches the anomalies seen by this new inspection against previously recognized ones. Within a short time, the entire run has been processed and the technician has been alerted to any new anomalies or new areas where the pipeline may be deteriorating at an unexpectedly rapid rate. The output of this scenario is a database of matched anomalies, a 4-dimensional profile of anomaly dimensions as they evolve through time.

The work here assumes that the process of creating boxed and called out anomalies is too computational intensive to be performed onboard an ICIP. To achieve the simplicity and compatibility envisioned for these tools we endeavor to match raw sensor readings as raw signals. The intent is to align raw input signals across multiple runs and data sets from different types of tools.

Application of SLAM to Pipeline Inspection

Over the last twenty years, there has been a concerted effort on the part of the mobile robotics research community to develop SLAM (simultaneous localization and mapping) capabilities for mobile robots traveling through unknown environments. The idea is that the robot, to be successful, can localize itself relative to obstacles and boundaries of the environment in which it is traveling, e.g., indoor, outdoor, or underground. In addition, it needs to build a map of its environment both for its own future use as well as for documentation of its activities. This project was first envisioned as an effort to leverage the developments of the robotics community to develop a similar capability for a pipeline pig where the features or landmarks that are used for navigation the features, e.g., girth welds, valves, and other prominent features of the pipeline as well as the anomalies that are detected as the pig travels through the pipe are used localize the robot. The idea is that the anomalies are in fixed locations (albeit there can be more features, i.e., anomalies that are observed in succeeding runs as was discussed earlier in this report), and thus they can be used to localize the pig, but the fact they are localized means their characteristics, e.g., wall loss, can be tracked from run to run.

From the work on SLAM, it has been shown that robots can localize themselves both when they have odometry and when they don't. We envision ICIPs that would have odometry would be developed first since this simplifies the algorithm, and would allow the use of the Extended Kalman Filter approach to the localization, i.e., path tracking.

In some ways, the problem of localization of a pig is simpler that for a mobile robot, e.g. the path is one-dimensional (or two dimensional if you include the "roll" of the pig, i.e., it rotation, as it travels), and thus the computation of position is simplified.

The key capability of the SLAM algorithm for run to run anomaly alignment will be the ability to identify sensed features as being the "same" anomaly. In SLAM the analogous problem is the challenge of recognizing the same landmark 1) after some incremental movement and 2) to recognize that landmark after larger traversals such that an area is revisited and goal is to recognize that the robot has been there before. This is called loop closure and requires recomputing large segments of the path (map) to fix errors in the heading. For pipeline pig data, the issue will be to align anomalies with those of previous runs, and unlike the typical robot problem, the features are expected to have changed from the previous run, thus making the matching problem more difficult.

While several attempts were made to obtain real pipe pig data, none were available for this project. Experiments were done using pseudo-pipe (fiber construction tubes) with holes rather than surface flaws. A 2D laser scanner was used as the sensor. For this scenario, the localization was straight-forward since the number of anomalies was quite small.

Probabilistic SLAM[2]

$$P(\boldsymbol{x}_k, \boldsymbol{m} | \boldsymbol{Z}_{0:k}, \boldsymbol{U}_{0:k}, \boldsymbol{x}_0)$$

Joint posterior density of the landmark locations (aka anomalies), given the recorder observations (measurements), control inputs, and initial pose. A recursive solution is desired since it will allow update at each time-step.

Starting with the estimate for the distribution at the previous time step (k-1),$P(\mathbf{x_{k-1}}, \mathbf{m} | \mathbf{Z_{0:k-1}}, \mathbf{U_{0:k-1}})$ using the control input, i.e., the motion of the pig in the interval from (k-i to k) and the observation, i.e., measurement update, the new joint posterior is computed using

[2] much of the material for this section is derived from the SLAM tutorial by Durrant-Whyte & Bailey

Bayes' theorem. To do the computation, both a state transition model and a observation model are defined and include the effect of the control input and the observation.

The observation (measurement) model expresses the probability of making an observation (e.g., specific measurement), z_k, when the pig is in a known location and the anomaly (aka landmark) location is known, i.e., the anomaly was detected and mapped in a previous inspection.

$$P(\boldsymbol{z}_k, |\boldsymbol{x}_k, \boldsymbol{m})$$

The motion model is described in terms of a probability distribution of the state transitions expressed as

$$P(\boldsymbol{x}_k, |\boldsymbol{x}_{k-1}, \boldsymbol{u}_k)$$

It is important to know that the state transition model is modeled as a Markov process for which the next state is only a function of the previous state and the input at that time step.

The probabilistic SLAM algorithm is then a two-step recursive process constructed using a prediction (called a motion or time update) followed by a correction (called a measurement or observation update). These are expressed as:

Time-update:

$$P(\boldsymbol{x}_k, \boldsymbol{m}|\boldsymbol{Z}_{0:k}, \boldsymbol{U}_{0:k}, \boldsymbol{x}_0) = \int P(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{u}_k) \times P(\boldsymbol{x}_{k-1}, \boldsymbol{m}|\boldsymbol{Z}_{0:k-1}, \boldsymbol{U}_{0:k-1}, \boldsymbol{x}_0) d\boldsymbol{x}_{k-a}$$

Measurement-update:

$$P(\boldsymbol{x}_k, \boldsymbol{m}|\boldsymbol{Z}_{0:k}, \boldsymbol{U}_{0:k}, \boldsymbol{x}_0) = \frac{P(\boldsymbol{z}_k|\boldsymbol{x}_k, \boldsymbol{m}) P(\boldsymbol{x}_k, \boldsymbol{m}|\boldsymbol{Z}_{0:k-1}, \boldsymbol{U}_{0:k}, \boldsymbol{x}_0)}{P(\boldsymbol{z}_k|\boldsymbol{Z}_{0:k-1}, \boldsymbol{U}_{0:k})}$$

For a map building activity, one can compute the conditional density, i.e.,

$$P(\boldsymbol{m}|\boldsymbol{X}_{0:k}, \boldsymbol{Z}_{0:k}, \boldsymbol{U}_{0:k})$$

which expresses the probability of a landmark, m (i.e., anomaly), being at a specific location given a certain pose of the robot (pig), a given set of sensor readings, and a given set of inputs (motions) of the robot (pig).

Application of the Kalman Filter to Localization of Pipeline Anomalies

Applying the Kalman filter to anomaly alignment has two phases, *a priori* and *a posteriori* state estimates, often referred to as the prediction and measurement updates (Siewart et al. 2011).

Prediction Update: The prediction update estimates the position of the pig based on motion input and would most directly involve the use of encoded movement sensor input, e.g., an instrumented wheel in contact with the pipe wall. The pig position at time k, $\hat{x}_k$, is based on the position at the previous time step and movement as indicated by the encoder measurement, u_k

$$\hat{x}_k = f(x_{k-1}, u_k)$$

Note we recognize that the pose will include both linear and angular displacement and will require these two inputs, thus x is in fact 2-vector. With knowledge of the plant (pig) and the error model on can compute the variance, P_{k-1}, associated with this estimate using an equation derived from an expression of the total probability theorem as applied to normal distributions:

$$\hat{P}_k = F_x P_{k-1} F_x^T + F_u Q_k F_u^T$$

where P_{k-1} is the covariance of the previous robot (pig) state, x_{k-1}, and Q_k is the covariance of the noise associated with the motion model. These two equations will be used to update the mean and covariance of the position distribution for the pig.

Measurement Update: This phase actually has four steps:

- Observation
- Measurement prediction
- Matching
- Estimation

Observation: This is the step where sensor measurements, z_k, are gathered. These measurements would consist of the wall assessment measure as dictated by the sensor used, e.g., MFL, AUT, piezo-sensor, collected at this time step.

Measurement prediction: The pig position and the map (for a previous inspection run) are used to generate multiple predicted feature observations, $\hat{z}_k^j$. The predicted observations represent the expected features (signal signatures) for that location in the pipeline. These features are expressed in the local coordinate frame of the robot, which will mapping their global pose into the local frame,

$$\hat{z}_k^j = h^j(\hat{x}_k, m^j)$$

where m^j represents the position of the feature in the map.

Matching: In this step, the measured features are compared with the predicted features. This involves an assignment of the observed features zj to the predicted features zi and specifically the computation of their difference, which is called the innovation. This is express as:

$$\nu_k^{ij} = (z_k^i - \hat{z}_k^j) = (z_k^i - h^i(\hat{x}_k, m^j))$$

The innovation covariance can then be found by applying the error propagation law to the system and measurement covariances:

$$P_{in}^{ij} = H^j \hat{P}_k H^{j^T} + R_k^i$$

where H^j is the Jacobian of h^j and R_k^i is the noise covariance of the measurement system, i.e., the sensors. The correspondence between the observation and the estimate are checked to see that the suggested correspondence makes sense. This is done by checking the "distance" between the two feature (anomaly) locations. The distance that is often used is the Mahalanobis distance (an expression of the distance between the features in terms of a normal distribution, i.e., a mean and variance. This observation (measurement) is assumed to be normally distributed with the observed value as the mean and the variance being that of the measurement. This check is called the validation gate, g, and is expressed as:

$$\nu_k^{ij} (P_{in}^{ij})^{-1} \nu_k^{ij} \leq g^2$$

Estimation: In this step, the estimate of the pig's position is made based upon the prediction, x_k, and the observations (sensed features) at time k. To do this the validated observations are stacked into a a vector, z_k, In a similar fashion a composite Jacobian is created by stacking the H^j, for each validated measurement. These are used to compute the composite innovation covariance.

$$x_t = \hat{x}_k + K_k \nu_k$$

$$P_k = \hat{P}_t - K_k P_{in} K_k^T$$

where

$$K_k = \hat{P}_k H_k^T (P_{in}^{ij})^{-1}$$

is the Kalman gain.

Summary and Conclusion

The problem of anomaly alignment and matching has been described in detail and the challenges of producing good matches have been identified. Descriptions of two approaches to the algorithms that could be used to address this problem were presented. Due to unavailability of data, no analysis has been

presented. The similarity of the anomaly matching to the problem of robot SLAM has been presented and some details of how the SLAM algorithms work have been presented.

REFERENCES

T. Bailey and H. Durrant-Whyte, 2006, "Simultaneous Localisation and Mapping (SLAM): Part II State of the Art," *IEEE* Robot*. Autom. Mag.*, vol. 13, no. 3, pp. 108–117.

M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, 2001, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Trans. Robot. Autom.*, vol. 17, no. 3, pp. 229–241.

H. Durrant-Whyte and T. Bailey, 2006, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–108.

M. Hassan, R. Rajkumar, D. Isa, and R. Arelhi, 2013, "Pipeline Defect Classification by Using Non-Destructive Testing and Improved Support Vector Machine Classification," *Int. J. Eng. Innov. Technol.*, vol. 2, no. 7, pp. 85–93.

Huyse L, van Roodselaar A, Onderdonk J, et al., 2010, "Improvements in the Accurate Estimation of Top of the Line Internal Corrosion of Subsea Pipelines on the Basis of In-Line Inspection Data". ASME. International Pipeline Conference, *2010 8th International Pipeline Conference, Volume 1* ():75-82. doi:10.1115/IPC2010-31038.

Richard McNealy, Ming Gao, Sergio Limon-Tapia, and Bill Deaton, 2008, "*Defect Assessment Using Effective Area Method from In-Line Inspections Data*". 7th International Pipeline Conference. Paper No IPC2008-64481. September 30, 2008. Calgary, AB.

Moghissi, O.C., 2008. "Corrosion Growth Rates Background: External Corrosion of Pipelines", PHMSA - Anomaly Assessment and Repair Workshop, October 22, 2008, https://primis.phmsa.dot.gov/meetings/MtgHome.mtg?mtg=55

J. O'Brien, I. Wood, and S. Drive, 2012, "Project WP # 474 : Integrated Internal Inspection and Cleaning Tool Technology for Pipelines,", PHMSA - DTPH56-10-T-000011. August 30, 2012, https://ntl.bts.gov/lib/46000/46300/46310/FilGet.pdf

D. D. S. Santana, N. Maruyama, and C. M. Furukawa, 2010, "Estimation of trajectories of pipeline PIGs using inertial measurements and non linear sensor fusion," *9th IEEE/IAS International Conference on Industry Applications, INDUSCON 2010*, Nov. 8-10 pp. 1–6.