Fusing More Frequent and Accurate Structural Damage Information from One Location to Assess Damage at another Location with Less Information

Roohollah Heidary[*], Katrina M. Groth, and Mohammad Modarres

Systems Risk and Reliability Analysis Lab, Center for Risk and Reliability, Department of Mechanical Engineering, University of Maryland, College Park, MD

Abstract:
This paper proposes a two-phase data fusion framework to be used within a prognostic and health management-based degradation model to estimate remaining useful life (RUL) of a segment of corroded oil and gas pipelines due to internal pitting. The existing degradation models for internal pitting corrosion are based on the assumption that operational conditions remain the same during the life of the pipeline. In contrast, this framework addresses the actual case where operational conditions change over time. The change in operational conditions is reflected in on-line inspection data of a specific active pit (pit *M*) and this framework is proposed to link this change to the growing behavior of other pits. In this way, dummy measurements of pit *i* are simulated based on on-line inspection data of pit *M* as well as the similarity between pit *i* and pit *M*. This similarity is defined as the average of the ratio of the estimated depth of pit *i* and pit *M* which is modified by a location factor. A hierarchical Bayesian-gamma process model and augmented particle filtering method are used respectively in the first and second phase of this framework to estimate the RUL of the pipeline.
Keywords: Data fusion, augmented particle filtering, hierarchical Bayesian model, gamma process, pitting corrosion.

1. INTRODUCTION

High cost of failure and maintenance in oil and gas pipelines has motivated us to consider developing a model to optimize maintenance policy (e.g., inspection frequency and method) and minimize costs without increasing the risk of failure. Among different failure mechanisms, corrosion is the primary and most severe failure mechanism of oil and gas pipelines, and pitting corrosion is of most concern because of its high rate of growth [1]. Historical failure data shows that 15% of all transmission pipeline incidents between 1994 and 2004 in the US [2] and 57.7% of oil and gas pipeline failures in Alberta, Canada between 1980 and 2005 [3] were due to internal corrosion. . Hence, a key step in developing an optimal maintenance policy for oil and gas pipelines is to develop a condition-based degradation model for internal pitting corrosion.

An appropriate degradation model should be able to incorporate data from multiple sources and should also account for all uncertainties, including epistemic uncertainty, variability in the temporal aspects, partial heterogeneity and inspection errors. Specifically for pitting corrosion, the degradation model should also consider time and depth dependency of pit growing rate [4]. This degradation model should also take into account any changes in the operational conditions. To the best knowledge of the authors, present internal pitting corrosion degradation models for oil and gas pipelines don't consider dynamic operational conditions of the pipelines. Therefore, the objective of this research is to introduce a framework for developing a condition-based degradation model to estimate RUL of a corroded pipeline due to internal pitting corrosion in which operational conditions steadily change over time.

Generally, degradation models can be categorized as population-based and PHM-based models. The first category is suitable when there is a huge number of similar components under similar operational conditions (e.g., electronic components). In contrast, the PHM approaches are more appropriate when each component has different degradation behavior, by taking advantage of having the degradation

[*] Heidary@umd.edu

measurement signals for each component [5]. Hence, in the case of pitting corrosion, a PHM-based model is more appropriate because each pit has different growing behavior.

Most PHM models are either physics-of-failure-based or data driven-based degradation models [6]. In the case of pitting corrosion, because of its highly stochastic nature and due to the variety of influencing parameters, data-driven approaches are more suitable, especially when the results of modeling are used for reliability assessment [7]. The available data driven-based pitting corrosion degradation models in the literature can be categorized as the random variable and stochastic process-based models, and the main difference between these two categories is that the former one does not consider the temporal variability of the pitting corrosion process. Since operational conditions and pits' growing behavior are changing over time, stochastic process-based models are more suitable to assess pitting corrosion process [4][8]. In the following, some leading pitting degradation models discussed in the literature are reviewed, followed by the motivation behind the proposed framework in this research and the differences with the available models are explained.

Ossai et al. [9] developed a random variable-based model to estimate maximum pit depth of internal pits in oil and gas pipelines. In this work, a multivariate regression analysis was conducted to correlate the maximum pit depth and eleven operational condition parameters (e.g., temperature, pH, flow rate, water cut). This model is developed by using pit depth and operational parameters measurement data for sixty X52 pipelines that were used for oil and gas gathering from a field in Nigeria over ten years. In this work, it has been assumed that operation conditions remain the same for the operation life of the pipeline. To the best knowledge of the authors, this model is the most comprehensive published model for internal pitting corrosion in oil and gas pipelines that correlate pit depth growth with eleven different affecting parameters. Despite the huge amount of data collection for developing this model, it is just applicable for those pipelines that have the production rates and physico-chemical parameters that fall within the range of those used to develop this model [9]. This model is not a condition-based one and it does not take into account the current depth of a pit in the prediction.

Caleyo et al. [10] developed a Markov process-based model to estimate pit depth of an externally corroded pipelines. In this work, a non-homogeneous linear pure birth Markov process (NLPBMP) is used to model external pit growing behavior in underground pipelines. In this model, based on some underlying assumptions [11], transition rates of NLPBMP are correlated to the parameters of a power function degradation model. That power function model has been developed by using a multivariate regression analysis to obtain a predictive model for pit growth as a function of soil properties. To develop this power function model, data acquisition has been done in 259 excavation sites over a three-year period for pipelines operating in southern Mexico to collect pit depth and soil properties data. This model also is not a condition-based one and it does not take into account the current depth of a pit in the prediction.

Maes et al. [12] developed a hierarchical Bayesian model to predict pipeline defect growth subject to in-line inspection uncertainty. In this model, different types of uncertainty (i.e., epistemic, local, temporal and measurement error) have been taken into account. This model takes into account the new inspection data and updates the degradation model's parameters accordingly.

The common underlying assumption in above models is that the corrosive environment of the pipeline (soil properties in external corrosion and operational conditions in internal corrosion) remains constant in the pipeline. In practice, operational conditions change over time (e.g., changing the chemistry of the products or changing the service of the pipeline) to react to market forces due to changes in the supply of and demand for products transported by hazardous liquid and gas pipelines [13]. To the best knowledge of the authors, there is no approach in the literature that estimate the RUL of a pipeline taking into account these changes. This research proposes a framework to develop a stochastic-process based degradation model that takes into account this issue.

2. PROBLEM DEFINITION

The integrity assessment of corroded pipelines is commonly performed based on in-line inspection (ILI) data. ILIs are usually utilized as a non-destructive tool for structural health monitoring to detect and size corrosion features [14]. ILI technologies use magnetic flux leakage or ultrasonic testing to assess damage. In many pipeline systems, ILI-based inspections are performed every three to ten years and at least two ILIs results have to be matched with respect to the location in the pipeline to determine pit-specific corrosion growth and time to failure [14]. In practice, it is not unusual that the operational conditions of the pipeline change over time. These changes can affect internal corrosion rate due to the change in total sulfur content, total acid number, or chloride salt concentrations. Operating temperature, water content, and size of the sediments may also change [13]. These operational changes are of the following types: flow reversal, product change (e.g. crude oil to refined products), or conversion to service (e.g. convert from natural gas to crude oil) [13]. Recently conversion of the gas system to transport hydrogen is also considered as a possible change in operational conditions of a pipeline [15].

When the operational conditions of a pipeline change, the pits' growing behavior changes as well. Conventionally, there are two ways to consider this change in behavior. First, monitoring the change in the operational variables and using an empirical (e.g. regression-based) pit depth growing model to correlate the changes with the maximum pit depth. The problem with this approach is that there is no benchmarked pit depth growing model available based on the new operational conditions. In addition, for all pits, distributions of the operational variables (e.g. temperature, pH, flow rate, etc.) remain constant, with the exception of some variables such as pressure that decreases along the pipeline. Therefore it is not straightforward to correlate operational variables with the maximum depth of each pit specifically. The second approach is to run more frequent pigging operations in order to trace changes in the operational conditions, which is not economically feasible because of the high cost of this operation. Accordingly, relying on a degradation model developed based on the earlier ILIs is inaccurate and increases the uncertainty of RUL estimation. This leads to unnecessary maintenances or unpredicted failures of the pipeline. A novel two-phase data fusion framework is proposed in this research to take into account changes in the operational conditions in the assessment of the pit depth.

3. PROPOSED FRAMEWORK

In order to consider the change in operational conditions in RUL estimation of a segment of a pipeline, a two-phase data fusion framework is introduced in this research. In this framework "more frequent-less uncertain" on-line inspection (OLI) data of one specific active pit (pit *M*) is fused with the "less frequent-more uncertain" ILIs data of other pits i (i = 1,…, m), in order to predict the depth of pit i with a higher confidence level. The ILIs data of m pits are gathered by the common practice pigging operations for integrity assessment of pipelines and the OLI data are gathered from on-line sensor (e.g. ultrasonic testing, magnetic flux leakage) at one specific location with at least one known active pit, *M*.

This situation is illustrated with an example which is depicted in Figure 1. In this example, it is assumed that a pipeline is in operation since time t_0 and all pits are initiated at this time. It is also assumed that there are 4 ILI datasets available at t_1, t_2, t_3 and t_4 for m pits along the pipeline. OLI data of a specific active pit, pit *M*, is also available since time t_1. Based on the PHM analysis up to time t_4, there is an estimation for RUL at t_5, but operational conditions change at time T ($t_4 < T < t_5$). This research proposes a framework to figure out how to update the RUL estimation at t_5, by taking into account the effects of changes in the operating conditions.

In the phase I of the proposed framework a hierarchical Bayesian model based on the gamma process (HB-GP) (Appendix. A.1) is used to fuse ILI data of m pits and estimating the depth of those pits. This model takes into account different levels of uncertainty (i.e., epistemic uncertainty (due to lack of knowledge about localized corrosion process), defect-specific uncertainties, temporal uncertainties, and local measurement errors). In this model hierarchical Bayesian analysis is used to update the hyper-parameters of the degradation model by using n ILI data sets of m pits and the gamma process is used to model the degradation process due to pitting corrosion.

Since 1975 that the gamma process was introduced in the area of reliability, it has been increasingly utilized to model stochastic degradation processes, because the temporal variability of stochastic degradation processes can be modeled properly by the gamma process [16]. Another advantage of using the gamma process is that it is appropriate to model monotonic and gradual degradation processes [16]. Noortwijk [16] compared the possibilities of modeling a stochastic degradation process (e.g., failure rate function, random-variable based models, Brownian motion with draft) and concluded that the gamma process is an appropriate candidate to model degradation processes such as corrosion, wear, and fatigue, that involves monotonically accumulating damage over time in a sequence of tiny increments.

In phase II, augmented particle filtering (Appendix. A.2) is used to fuse ILI and OLI data. The proposed framework is shown in Figure 2 and the detail of phase I and II are explained in Sections 3.1 and 3.2.

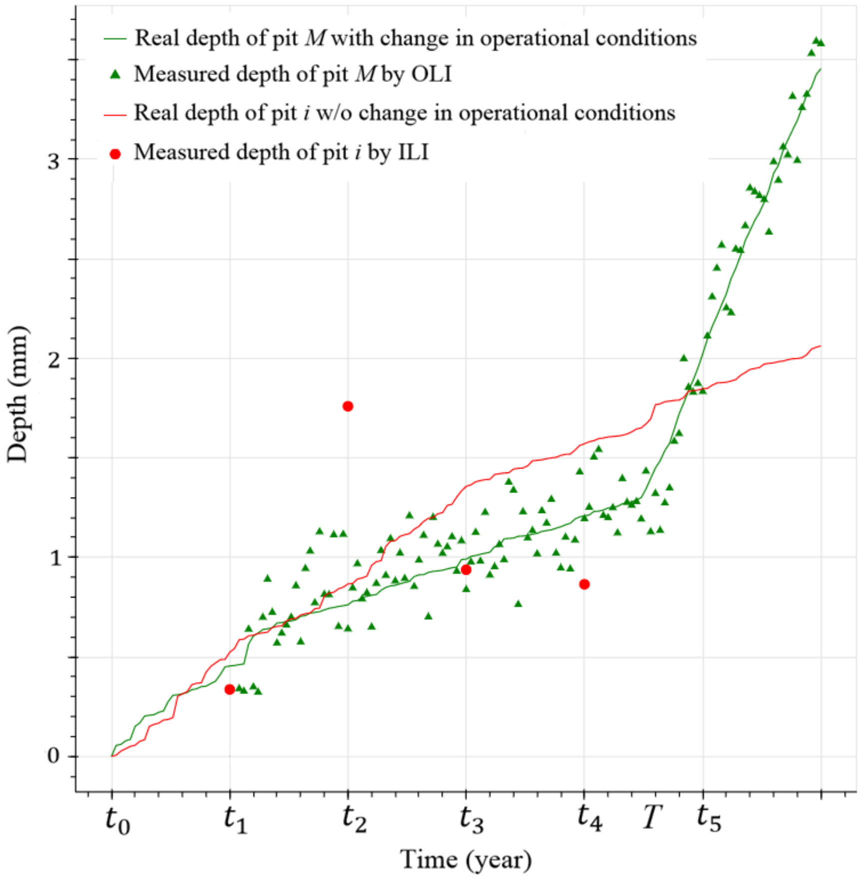

Figure 1. ILI data of pit i and OLI data of pit M

3.1. Phase I

In the first phase, *n* ILI datasets of *m* pits are fused to have an estimation of the maximum depth of those *m* pits at ILIs times. According to Figure 2, the inputs for this phase are ILIs of *m* pits, measurement error of ILIs tools and prior distributions for θ_1 and θ_2 (parameters of the power function given in Equation (4) in Appendix. A.1). Then, by applying HB-GP analysis (step 5), the depth of pit *i* (*i*=1,…,*m*) at t_1, t_2, t_3 and t_4 are estimated (step 6). The prior distributions of θ_1 and θ_2 are estimated from the OLI data of pit *M* (step 4). It is worth noting that in Bayesian analysis, selection of prior distributions changes the posterior distributions dramatically, especially when there is a scarce number of evidence data points (which is the case of ILI data). By installing an online sensor and monitoring the behavior of a single active pit (pit *M*), and applying a non-linear regression analysis on the OLI data of that pit (step 4), a proper estimation for prior distributions of θ_1 and θ_2 are obtained. Generalizing prior distributions of θ_1 and θ_2 of one active pit to the other active pits is valid based on this assumption that all pits are under similar operational conditions at each time (i.e., same probability density functions for temperature, pH, etc.) and therefore they have similar growing behaviors. These prior distributions are also used as the prior values in APF for pit *M* (step 7)

3.2. Phase II

In the second phase, the OLI data of a specific pit (pit *M*) from time *T* (the time of the last ILI) to time *T*+*t* is used to generate dummy measurements for pit *i* (*i*= 1, …, *m*), in order to estimate RUL of a pipeline segment which includes the *i*th pit (step 14). In this approach, after time *T*, when OLI data of pit *M* arrives, a dummy measurement for pit *i* is generated based on the similarity between pit *M* and pit *i* (step 12). This similarity is defined by using the maximum information that is available for pit *M* and pit *i* up to time *T*. This information for pit *i* are *n* ILI data points (in this example four data points), circumferential and longitudinal locations of that pit along the pipeline and for pit *M* are corresponding OLI data and its location parameters. The similarity index between these two pits are defined as the average of the ratio of the estimated depth of pit *i* and pit *M* at ILI times (step 8) which is modified by a location factor (step 9). This location factor is defined based on the circumferential location of a pit on the pipeline to consider the effect of the location on pit depth growing behavior (e.g. top of line). The dummy measurements for pit *i* are the OLI data of pit *M* at time *T*+*t* multiply by the similarity index between pit i and pit *M*. Those dummy measurements are used to estimate the depth of pit *i* at time *T*+*t* by APF analysis (step 13). In this step, θ_1 and θ_2 (Appendix. A, Equation (4)), h and U_k (Appendix. A, Equation (15)) are borrowed from APF analysis of pit *M*. Eventually, RUL for the pipeline segment with pit *i* is estimated in any point in time in the future when there is no ILI data for pit *i* (step 14).

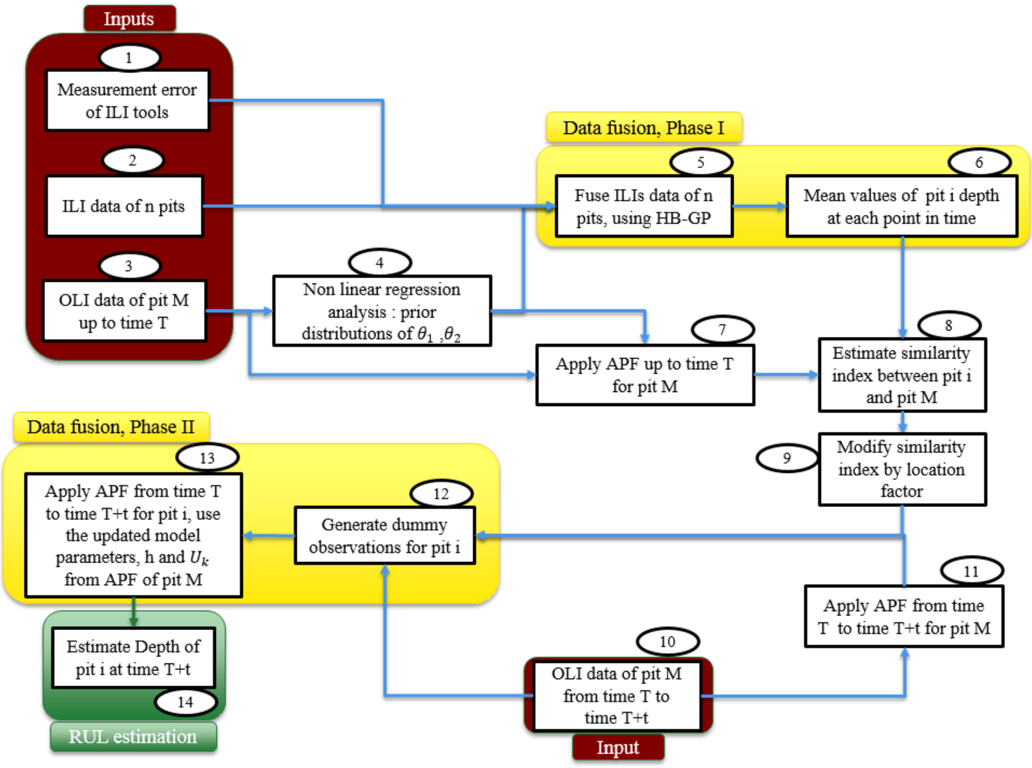

Figure 2. Flowchart of the proposed data fusion framework

4. CONCLUSION

In this research a two-phase data fusion framework is proposed to estimate the RUL of a corroding oil and gas pipeline due to internal pitting corrosion. The available pit depth prediction models in the literature are based on the assumption that operational conditions are steady during the operation life of the pipeline. This framework addresses the realistic case where such conditions change over time.

In this framework, more frequent and more accurate inspection data of one pit (pit *M*) is fused with less frequent and more uncertain inspection data of another pit (pit *i*, *i* =1,….,*m*) to have an estimation for the depth of pit *i* when the operational conditions change, and there is no evidence about change in pit *i* growing rate until the next ILI.

This framework is based on the idea of generating dummy measurements for pit *i* based on the OLI data of a specific pit (pit *M*). Those dummy measurements are generated by multiplying the OLI data of pit *M* after the time of the last ILI, by the similarity index between pit *i* and pit *M*. That similarity index is defined as the average of the ratio of the estimated depth for pit *i* over estimated depth of pit *M* at ILIs times. This similarity index is modified by a location factor to consider the effect of the location (e.g. Top of line) in the pits' depth growing behavior. Estimated depth of pit *i* is obtained by fusing ILI data of m pits by using HB-GP model (data fusion, phase I) and estimated depth of pit M is obtained by using APF method on OLI data of pit M before time T. Finally, RUL of the segment which includes the i^{th} is estimated by APF method, using the dummy measurements of pit i and borrowing θ_1, θ_2, h and U_k from APF analysis for pit M (data fusion, phase II).

The proposed framework will facilitate the integrity assessment of internally corroded pipelines when the operational conditions change over time. The validation of the proposed approach is underway. In addition, pit initiation time is assumed to be the same for all pits in this study. This assumption will be relaxed in the future work. Correlating the similarity index with other potential pit specific parameters (e.g. location in this research) is also another aspect of our future work.

Acknowledgments
This work is being carried out as a part of the Pipeline System Integrity Management Project, which is supported by the Petroleum Institute, Khalifa University of Science and Technology, Abu Dhabi, UAE

Appendix. A

A.1. Hierarchical Bayesian-gamma process model

In order to fuse pigging data of all pits, a hierarchical Bayesian-gamma process model is used. Figure 3 depicts this hierarchical Bayesian-gamma process model.

A gamma process with shape function $\alpha(t)$>0 and rate parameter β (inverse of scale parameter) is a continuous-time stochastic process $\{D(t), t \geq 0\}\{D(t), t \geq 0\}$ with the following properties:

1. $D(0) = 0$ with probability one;
2. $\Delta D = D(\tau) - D(t) \sim Ga(\Delta\alpha = \alpha(\tau) - \alpha(t)), \beta)$ for all $\tau > t \geq 0\tau > t \geq 0$;
3. $D(t)$ has independent increments.

Where Ga represents probability density function of gamma distribution.

Let $D(t)$ denote the degradation level at time $t, t \geq 0$. In accordance with the aforementioned definition, the probability density function of $D(t)$ is given by Equation (1).

$$f_{D(t)}(d) = Ga(d \mid \alpha(t), \beta) = \frac{\beta^{\alpha(t)}}{\Gamma(\alpha(t))} d^{\alpha(t)-1} \exp(-\beta d) \tag{1}$$

Where $\Gamma(.)$ denotes the gamma function.

The expectation and the variance of this process can be written as below:

$$E\big(D(t)\big) = \frac{\alpha(t)}{\beta} \tag{2}$$

$$Var\big(D(t)\big) = \frac{\alpha(t)}{\beta^2} \tag{3}$$

Equation (2) shows that the shape parameter of the gamma process reflects the average degradation trend as a function of time depending on the physics of the degradation process. Hence, different degradation processes can be modeled depending on the functional form of the time mapping function of the shape parameter. Generally, power law function is a versatile function that can demonstrate different types of degradation process with increasing, decreasing, or constant degradation rate. According to (4), depth increment of each pit at each time interval ($\Delta D_{j,i}$) follows a gamma distribution with shape parameter $\Delta\alpha_{j,i}$ and rate parameter β_j.

Equation (4) depicts the relationship between shape parameter of the gamma process and general form of a power law function.

$$\Delta\alpha_{j,i} = \theta_1 \left\{ \left(t_{j,i-1} + \Delta t_{j,i-1}\right)^{\theta_2} - \left(t_{j,i-1}\right)^{\theta_2} \right\} \tag{4}$$

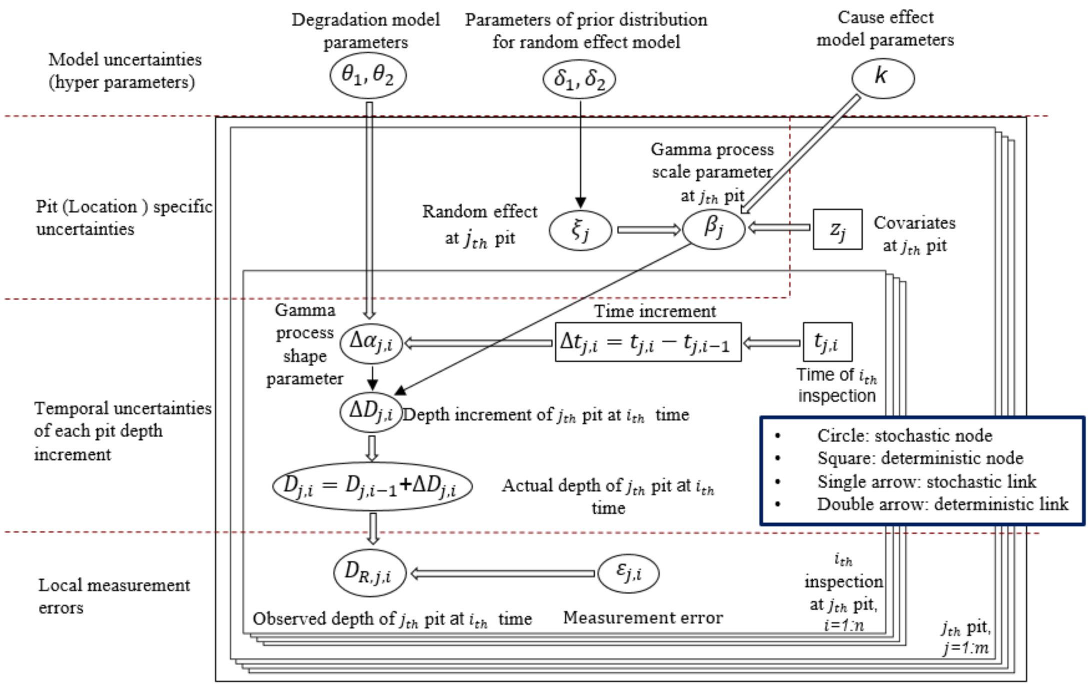

Figure 3. Hierarchical Bayes framework for heterogeneous degradation, modified from **[12]**

In this equation, the exponent θ_2=1 resembles a constant, θ_2< 1 resembles a decreasing and θ_2 >1 resembles an increasing degradation rate process. On the other hand, it is widely accepted that the pitting corrosion growth can be described by a power function with an exponent between zero and one [1][16]. Therefore, in the case of pitting corrosion, θ_2 is between zero and one. Since θ_1 and θ_2 model epistemic uncertainty of the degradation process that describes uncertain mechanistic aspects of the degradation process that are common to all pits within a pipeline, these parameters are updated by using the ILI data of all pits at all inspection times. The local heterogeneity is modeled by the rate parameter (β) of the gamma process which is defect-specific. Therefore, the rate parameter of each pit is updated by involving all ILI data of each pit. As it is shown in Figure 3, the location heterogeneity is represented by z_j, k and ξ_j. z_j are local covariates (e.g., pressure, temperature, pH), k is a vector of cause and effect regression coefficients associated with z_j and ξ_j are local aleatory effects that cannot be explained by defined covariates (e.g. top of line corrosion) [12].

Zhang et al. [16] applied this framework to predict pit depth of 62 pits by using the corresponding ILI data that was available at four inspection times. In this work it is assumed that pitting corrosion follows a homogenous gamma process (i.e., θ_2=1). Zhang et al. [16] assumed that θ_1 follows a gamma distribution with prior values of shape and scale parameters both equal to 1. Apparently having a proper prior distribution in Bayesian analysis changes the resultant posterior distribution drastically, especially when there is scarce evidence data, which is the case of ILI data of pipelines with a few inspection data for each pit (e.g., many operation companies just have one or two sets of ILI data if not zero). In this research, this HB-GP framework is modified and combined with augmented particle filtering to estimate RUL of an oil and gas pipeline when operational conditions change over time.

A.2. Augmented particle filtering

Estimating the state of a dynamic system by utilizing on-line noisy measurements has many applications in different areas such as robotic, object tracking, economy, etc. Particle filtering or sequential Monte Carlo method is a family of state estimation techniques that use recursive Bayesian approaches to update or filter the state vector at each time. Because of its flexible and powerful diagnostic and prognostic features, application of particle filtering in reliability engineering has increased rapidly in the recent years [17].

Application of particle filtering in degradation analysis is explained briefly here. Consider the evolution of the degradation level sequence of a system given by Equation (5).

$$D_k = f_k(D_{k-1}, V_{k-1}) \rightarrow p(D_k \mid D_{k-1}) \tag{5}$$

In this equation $f_k: R^{n_d} \times R^{n_v} \rightarrow R^{n_d}$ is the process or evolution function of the degradation level D_{k-1}. $\{V_k, k \in N\}$ is an i.i.d process noise sequence, and n_d, n_v are dimensions of the state and process noise vectors, respectively. Usually, the degradation level is not measurable directly and it is desired to estimate degradation level given noisy measurements. Equation (6) shows the relationship between true degradation level and the corresponding noisy measurement.

$$z_k = h_k(D_k, \omega_k) \rightarrow p(z_k \mid D_k) \tag{6}$$

In this equation $h_k: R^{n_d} \times R^{n_\omega} \rightarrow R^{n_z}$ is the measurement function that links measurement with degradation level D_k. $\{\omega_k, k \in N\}$ is an i.i.d measurement noise sequence, and n_z, n_ω are dimensions of the measurement and measurement noise vectors, respectively.

Assuming that initial probability distribution function of the degradation level is available ($p(D(0))$, the objective is to estimate the probability of being at each degradation level,D_k, based on the all available measurements up to time t_k ($p(D_k|z_{1:k})$). From a Bayesian perspective, this can be obtained following two steps: prediction and update step.

In the prediction step, assuming that the pdf $p(D_{k-1}|z_{1:k-1})$ at time t_{k-1} is available, $p(D_k|z_{1:k-1})$ can be obtained by using process function (Equation (5)) with known distribution of process noise vector (V_k) via the Chapman-Kolmogorov equation:

$$p(D_k \mid z_{1:k-1}) = \int p(D_k \mid D_{k-1}, z_{1:k-1}).p(D_{k-1} \mid z_{1:k-1}).dD_{k-1} \tag{7}$$

Assuming measurements are conditionally independent given the degradation level and also assuming first order Markovian property, the degradation level is predicted according to Equation (8).

$$p(D_k \mid z_{1:k-1}) = \int p(D_k \mid D_{k-1}).p(D_{k-1} \mid z_{1:k-1}).dD_{k-1} \tag{8}$$

In the update step, when a new measurement, z_k becomes available at time t_k, the posterior distribution of the current degradation level is obtained by using Bayesian updating:

$$p(D_k \mid z_{1:k}) = \frac{p(z_k \mid D_k).p(D_k \mid z_{1:k-1})}{p(z_k \mid z_{1:k-1})} \propto p(z_k \mid D_k).p(D_k \mid z_{1:k-1}) \tag{9}$$

Except for special cases (i.e., linear Gaussian state space models), it is not possible to evaluate these distributions analytically. Particle filtering is a powerful technique that approximates this conditional probability distribution of $p(D_k|z_{1:k})$ by a set of weighted particles as

$$p(D_k \mid z_{1:k}) \approx \sum_{i=1}^{N} w_k^i \delta(D_k - D_k^i) \tag{10}$$

where w_k^i is the normalized weight of the i^{th} particle at time t_k and δ is the Dirac's delta function. The basic idea of PF is that, firstly N samples or particle are generated based on the initial belief about degradation level ($p(D(0))$) with equal weights. Then at each time step, all particles are evolved by using process model (Equation ((5)), and subsequently, measurement at each time t_k is utilized, by using measurement model (Equation ((6)), to update the predicted degradation level and the normalized weight of each particle.

With respect to pitting corrosion, it is well accepted that maximum depth of a pit follows a power function with a positive exponent less than one [1][16] (Equation (11)).

$$D_k = \theta_1 (t_k - t_0)^{\theta_2} \tag{11}$$

Where D_k is the maximum depth of a pit at time t_k, t_0 is the pit growing initiation time, and θ_1 and θ_2 are the model parameters. By taking the first derivative of this function, for small enough time interval Δt, the process model for evolution of a pit depth can be defined according to Equation (12).

$$D_k(t) = D_{k-1}(t) + \theta_1\theta_2(t_k - t_0)^{\theta_2 - 1}\Delta t \times e^V \quad (12)$$

Where V is a white Gaussian noise.

In order to define measurement model, by considering both the bias and random scattering error, the relationship between the actual pit depth and the measurement at each time t_k can be defined as

$$z_k = a_k + b_k D_k + \omega_k \quad (13)$$

Where a_k and b_k represent constant and non-constant biases of the measurement tool employed at time t_k and ω_k represents random scattering error associated with the reported depth of a pit at time t_k. a_k, b_k and ω_k can be obtained based on the measurement tool's manufacturer specifications or by comparing the measurements and the real depths of a set of static defects [16] [18].

Although it is well accepted that pit depth grow follows a power function form, estimating the parameters of this function is a challenge. One way of estimating these parameters is to use regression analysis to correlate them with the operational parameters [9], [19]. As mentioned previously, this approach has two drawbacks; first, it requires a huge amount of recorded operational parameters data which in practice is not usually available and second, the output model would be a random-variable based model that does not represent the temporal stochasticity of the corrosion process. In this research, augmented particle filtering (APF) is used to estimate the model parameters. In APF both model parameters and state of the system are updated simultaneously at each time step. Liu and West [20], proposed a flexible approach (Kernel smoothing) to filter the state and the model parameters simultaneously which is explained briefly here.

Assume that at time t_{k-1}, the state vector (D_{k-1}), the model parameters' vector ($\boldsymbol{\theta}_{k-1}$), and the associated weight (w_{k-1}) for each particle are available. At time t_k, when a new measurement z_k gets available, the posterior distribution of state vector and model parameters can be obtained by Baye's rule:

$$p(D_k, \boldsymbol{\theta}|z_{1:k}) \propto p(z_k|D_k, \boldsymbol{\theta}).p(D_k, \boldsymbol{\theta}|z_{1:k-1}) \propto p(z_k|D_k, \boldsymbol{\theta}).p(D_k|\boldsymbol{\theta}, z_{1:k-1}).p(\boldsymbol{\theta}|z_{1:k-1}) \quad (14)$$

When model parameters ($\boldsymbol{\theta}$) are known, this equation leads to Equation (8). However, usually model parameters are unknown, and they should be estimated by using available measurement data.

The smooth kernel density form of $p(\boldsymbol{\theta}|z_{1:k-1})$ is given by Equation **Error! Reference source not found.**.

$$p(\boldsymbol{\theta}_k|z_{1:k-1}) \propto \sum_{i=1}^{n} w_{k-1}^i N(\boldsymbol{\theta}_k|m_{k-1}^i, h^2 U_{k-1}) \quad (15)$$

Where $N(.|\boldsymbol{m}, \boldsymbol{S})$ is a multivariate normal density with mean m and variance matrix $\boldsymbol{S}$, U_k is the Monte Carlo posterior variance and h is the smoothing parameter. Standard kernel methods would suggest $\boldsymbol{m}_k^i = \boldsymbol{\theta}_k^i$. However, based on this assumption, the variance of the resulting mixture of normal distributions is $(1 + h^2)U_k$ which is always larger than U_k. Liu and West [20] introduced a novel idea of shrinkage of kernel locations to solve this problem according to Equation **Error! Reference source not found.**.

$$m_k^i = \sqrt{1 - h^2}\boldsymbol{\theta}_k^i + \left(1 - \sqrt{1 - h^2}\right)\overline{\boldsymbol{\theta}}_k \quad (16)$$

With these kernel locations, the mean value of the normal mixture is $\overline{\boldsymbol{\theta}}_k$ and the variance has the correct value of U_k. Selecting a proper value for $h \in [0,1]$ is an important step in applying APF. Chen et al. [21] suggested $0<h<0.2$ when parameters are slowly varying and $0.8<h<1$ for a highly stochastic process that parameters are expected to change significantly. In this research, an approach is proposed to estimate h and U_k by using OLI data of pit M and using them for APF analysis of the other pits as is explained in the main sections.

References

[1] J. C. Velázquez, F. Caleyo, A. Valor, and J. M. Hallen, "Predictive Model for Pitting Corrosion in Buried Oil and Gas Pipelines," *CORROSION*, vol. 65, no. 5, pp. 332–342, May 2009.

[2] S. Papavinasam, R. W. Revie, W. I Friesen, A. Doiron, and T. Panneerselvan, "Review of models to predict internal pitting corrosion of oil and gas pipelines," *Corros. Rev.*, vol. 24, no. 3–4, pp. 173–230, 2006.

[3] S. Papavinasam, *Corrosion Control in the Oil and Gas Industry*. Elsevier, 2013.

[4] R. Heidary, S. A. Gabriel, M. Modarres, K. M. Groth, and N. Vahdati, "A Review of Data-Driven Oil and Gas Pipeline Pitting Corrosion Growth Models Applicable for Prognostic and Health Management," *Int. J. Progn. Health Manag.*, vol. 9, no. 1, 2018.

[5] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and Health Management: A Review on Data Driven Approaches," *Mathematical Problems in Engineering*, 2015. [Online]. Available: https://www.hindawi.com/journals/mpe/2015/793161/. [Accessed: 09-Jan-2018].

[6] A. Imanian and M. Modarres, "A thermodynamic entropy-based damage assessment with applications to prognostics and health management," *Struct. Health Monit.*, p. 1475921716689561, Feb. 2017.

[7] A. Valor, F. Caleyo, L. Alfonso, D. Rivas, and J. M. Hallen, "Stochastic modeling of pitting corrosion: A new model for initiation and growth of multiple corrosion pits," *Corros. Sci.*, vol. 49, no. 2, pp. 559–579, Feb. 2007.

[8] S. Zhang and W. Zhou, "System reliability of corroding pipelines considering stochastic process-based models for defect growth and internal pressure," *Int. J. Press. Vessels Pip.*, vol. 111–112, pp. 120–130, Nov. 2013.

[9] C. I. Ossai, B. Boswell, and I. J. Davies, "Predictive modelling of internal pitting corrosion of aged non-piggable pipelines," *J. Electrochem. Soc.*, vol. 162, no. 6, pp. C251–C259, 2015.

[10] F. Caleyo, J. C. Velázquez, A. Valor, and J. M. Hallen, "Markov chain modelling of pitting corrosion in underground pipelines," *Corros. Sci.*, vol. 51, no. 9, pp. 2197–2207, 2009.

[11] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*, First. Chapman & Hall, CRC, Boca Raton Florida, 1965.

[12] M. A. Maes, M. H. Faber, and M. R. Dann, "Hierarchical Modeling of Pipeline Defect Growth Subject to ILI Uncertainty," pp. 375–384, Jan. 2009.

[13] "Regulations | PHMSA." [Online]. Available: https://www.phmsa.dot.gov/regulations-fr/notices/2014-22201. [Accessed: 03-Apr-2018].

[14] M. R. Dann and C. Dann, "Automated matching of pipeline corrosion features from in-line inspection data," *Reliab. Eng. Syst. Saf.*, vol. 162, pp. 40–50, 2017.

[15] P. E. Dodds and S. Demoullin, "Conversion of the UK gas system to transport hydrogen," *Int. J. Hydrog. Energy*, vol. 38, no. 18, pp. 7189–7200, Jun. 2013.

[16] M. Nuhi, T. A. Seer, A. M. Al Tamimi, M. Modarres, and A. Seibi, "Reliability analysis for degradation effects of pitting corrosion in carbon steel pipes," *Procedia Eng.*, vol. 10, pp. 1930–1935, 2011.

[17] E. Rabiei, E. L. Droguett, and M. Modarres, "A prognostics approach based on the evolution of damage precursors using dynamic Bayesian networks," *Adv. Mech. Eng.*, vol. 8, no. 9, p. 1687814016666747, 2016.

[18] M. Al-Amin, W. Zhou, S. Zhang, S. Kariyawasam, and H. Wang, "Bayesian Model for Calibration of ILI Tools," pp. 201–208, Sep. 2012.

[19] J. C. Velázquez, F. Caleyo, A. Valor, J. M. Hallen, J. H. Espina-Hernandez, and A. López-Montenegro, "Statistical modeling of pitting corrosion in buried pipelines taking into account soil properties," in *2008 7th International Pipeline Conference*, 2008, pp. 251–259.

[20] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo methods in practice*, Springer, 2001, pp. 197–223.

[21] T. Chen, J. Morris, and E. Martin, "Particle filters for state and parameter estimation in batch processes," *J. Process Control*, vol. 15, no. 6, pp. 665–673, 2005.