

A hybrid model of internal pitting corrosion degradation under changing operational conditions for pipeline integrity management

Structural Health Monitoring

2020, Vol. 19(4) 1075–1091

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1475921719877656

journals.sagepub.com/home/shmRoohollah Heidary  and Katrina M Groth

Abstract

This article proposes a new framework to estimate the degradation level in oil and gas pipelines corroded by internal pitting when operational conditions change over time. Despite the fact that the operational conditions of a pipeline change at various times, this change has not been addressed in the current available pipeline corrosion degradation models. In this framework, a hierarchical Bayesian method and augmented particle filtering are used for data fusion to address this issue. This framework is applied on a case study and the results are compared with the estimations of a state of the art pitting corrosion degradation model.

Keywords

Data fusion, augmented particle filtering, pipeline integrity management, prognostics and health management, pitting corrosion

Introduction

Although pipelines are the most reliable and economical mode of transportation of oil and gas in large quantities,¹ their failures and maintenance activities can impose a high cost to industry. In order to avoid unpredicted failures and also unnecessary maintenance activities, having a high confidence estimation of pipeline degradation due to different potential failure mechanisms is critical in pipeline integrity management.

Among different failure mechanisms, corrosion is especially significant for oil and gas pipelines, and pitting corrosion is of most concern because of the high pits growth rate.² According to the available literature, 15% of all transmission pipeline incidents between 1994 and 2004 in the United States³ and 58% of oil and gas pipeline failure in Alberta, Canada, were due to internal corrosion.⁴ Furthermore, 90% of corrosion failures of transmission pipelines in the United States between 1970 and 1984 were due to localized corrosion.⁴ Therefore, investigation of internal localized corrosion is an essential task in pipeline integrity management.

While there has been significant progress in understanding uniform corrosion, localized corrosion is still not well understood.⁵ Internal pitting corrosion, as a localized corrosion mechanism, is a highly stochastic

process which is affected by a large number of dependent and independent parameters.^{6,7} Some of these parameters are pH value in the water phase, the water chemistry, the protective scale, the CO₂ partial pressure, the amount of H₂S, the effect of oil wetting, the metal alloy composition, the temperature, the multi-phase flow, and the flow rate. In addition, there is temporal and local heterogeneity in some of these parameters, and interdependence between them.⁸ There has been some progress on development of different pitting corrosion degradation models in the literature. We reviewed some of the leading probabilistic prediction models for oil and gas pipelines corroded by pitting corrosion and ranked them based on their comprehensiveness, the required data, and the level of knowledge that are required to develop each of those models.⁸ To the best of authors' knowledge, the available degradation models

Systems Risk and Reliability Analysis Lab (SyRRA), Center for Risk and Reliability, University of Maryland, College Park, USA

Corresponding author:

Roohollah Heidary, Systems Risk and Reliability Analysis Lab (SyRRA), Center for Risk and Reliability, University of Maryland, Glenn L. Martin Hall, 4298 Campus Drive, College Park, MD 20742-5031, USA.
Email: heidary@umd.edu

for internal pitting corrosion of oil and gas pipelines have been developed based on the assumption that all pits are under the same operational conditions for the operating life of the pipelines.^{9,10,11} However, operational conditions of a pipeline can change due to changing nature of the field, flow reversal, product change, or conversion to service.¹² For example, a pipeline that was constructed in 1953 to deliver crude oil, was converted to natural gas service in 2002 in Austin, TX, USA.¹³ Another example for conversion to service is the use of the current natural gas pipelines to deliver hydrogen across the United States, which is in the feasibility study phase.^{14,15} This option is under investigation in the United Kingdom as well.¹⁶ An example for change in the product is the continuous change in the product properties within an uncertain range in the natural gas pipelines.¹⁷ According to Pipeline and Hazardous Materials Safety Administration (PHMSA), changes in operational conditions may impact various aspects of a pipelines operation, maintenance, monitoring, integrity management, material compatibility, and corrosion susceptibility.¹⁸ Therefore, the focus of this article is on developing a hybrid prognostics and health management (PHM) degradation model for internal pitting corrosion in pipelines when operational conditions change over time. This model provides the main input (i.e. estimated degradation level) for condition-based maintenance optimization of the pipeline.

The rest of this article is organized as following. In the next section, related works, approach, and contributions are discussed. Two Bayesian inference methods that are used in this framework are explained afterward. Then, the problem is defined and the proposed framework is explained by applying that on a case study. The “Results” section is dedicated to discuss the results, and “Conclusion” is the last section of this article. The description of the used symbols in this article are given in Table 1.

Related work, approach, and contributions

The integrity management of piggable pipelines is commonly performed using in-line inspection data that are obtained by utilizing a non-destructive tool (e.g. magnetic flux leakage (MFL) or ultrasonic test (UT)).¹⁹ Maes et al.¹¹ proposed a hierarchical Bayesian (HB) model based on a gamma process to project pit growth in piggable pipelines. They considered four types of uncertainty in modeling: epistemic uncertainty, spatial heterogeneity, temporal variation, and measurement errors.¹¹ Zhang and Zhou¹⁰ used Maes model to estimate maximum pit depth* for a gas pipeline in Alberta,

Canada. They showed that for 90% of the 62 pits, the absolute difference between the predicted depths and the field measured depths are less than or equal to 10% of the pipe wall thickness (PWT). Zhang et al.²¹ extended Maes model by assuming pits' depth growth follow an inverse Gaussian process instead of a gamma process. By applying this new approach, on the same in-line inspection (ILI) data set in,¹⁰ they showed that the new results are essentially equivalent to those based on a gamma process. In another work, Zhang and Zhou²² again extended the Maes model using a Bayesian dynamic linear model instead of a gamma process. In this case, they showed that the absolute difference between predicted depth and the corresponding field measurement is less than or equal to 10% of the PWT for about 92% of the pits. To the best knowledge of the authors, the family of HB models are the state of the art degradation models for piggable pipelines.

In contrast to the above-mentioned family of models, there is another approach that is applicable for non-piggable pipelines. In this family of models, a generic degradation model is developed for all pits by correlating the maximum pit depth with the operational parameters. One of the most comprehensive internal pitting corrosion degradation model that has been developed based on this approach, is proposed by Ossai et al.²³ This model correlated 11 operational parameters with the maximum pit depth, by performing a nonlinear regression analysis. Ossai model was developed using 10 years of recorded pit depth data from UT, and operating parameters data that were obtained via routine quality control procedures. The Ossai model is explained in more detail in “Synthetic data generation procedure.”

These two families of models are hybrid PHM models that combine inspection and measurement data with Physics of Failure (PoF) of the pitting corrosion process, by relying on this well-accepted assumption that maximum depth of a pit follows a power function with a positive exponent less than one.^{2,24} The HB models rely more heavily on the inspection data, because in these approaches specific inspection data were available for each individual pit. In contrast, in the generic models, the PoF aspect is emphasized by taking into account the different covariates in degradation modeling. However, the data are not pit-specific in this family of models. One contribution of this article is to propose a hybrid framework that has the advantages of both approaches described above, by considering both specific ILI data of each pit and also the effect of operational parameters on PoF in degradation modeling.

Another contribution of this work, as mentioned above, is to consider changes in operational conditions

* (estimation of maximum (vs mean, etc.) pit depth is the main concern in pitting corrosion literature because the deepest pits are the first that cause leaks)²⁰

Table 1. Nomenclature.

Symbol	Description	Unit
SI	Similarity index	
ILI	In-line inspection	
OLI	Online inspection	
PF	Particle filtering	
APF	Augmented particle filtering	
HB-NHGP	Hierarchical Bayesian based on a non-homogeneous gamma process	
PWT	Pipe wall thickness	
EMPD	Estimated maximum pit depth	
RUL	Remaining useful life	
PHM	Prognostics and health management	
PM-SD	Process model standard deviation in APF	
PoF	Physics of Failure	
Metric R	Root mean squared error between actual and predicted maximum depth of all pits	
Metric N	The percentage of all pits that their predicted depths fall within the $\pm 10\%$ of their actual maximum depth	
a	Constant biased error	
b	Proportional biased error	
i	Pit index	
j	Time index	
p	Particle index	
P	Number of particles	
k	Coefficient of power law model	mm
t	Time	year
τ	Time	year
t_0	Pit initiation time	year
T	Time at which operational conditions change	year
d	Maximum pit depth	mm
θ	Vector of model parameters	
ν	Exponent of power law model	
V_q	q th operational parameter	
γ_0	Regression coefficient for the q th operational parameter	
Q	The intercept of the regression model	
Q	No. of operational parameters	
y	Measured maximum pit depth	mm
ϵ	Random scattering error	mm
m	No. of in-line inspected pits	
n	No. of ILI operations	
$Pit M$	The online inspected pit	
h	Kernel smoothing factor	
Actual depth	Synthetic actual depth of a pit without measurement error	
Measured depth	Synthetic measured depth of a pit	
Estimated depth	An estimation of the synthetic actual depth of a pit	

in pitting corrosion degradation modeling. Considering change in degradation rate in condition-based maintenance optimization is addressed by Grall and Fouladirad²⁵ by using online inspection data for one component/item. However, in the case of long pipelines, it is not feasible to install online sensors on all pits to detect change in their degradation rates in order to consider that in maintenance optimization. This article proposes a novel framework to monitor change in degradation rate (due to change in operational conditions) in the reference pit, and then make a logical and reliable inference about the change in the degradation rate and degradation level of other active pits along the pipeline.

Bayesian inference methods

The proposed framework is founded on two Bayesian inference techniques: augmented particle filtering (APF) and HB methods. APF is used to fuse online inspection (OLI) data and estimate the degradation level of the reference pit. A HB method is used to fuse ILI data and estimate degradation level of ILI pits at ILI times. These two methods are discussed in the following sections.

APF

Particle filtering (PF) or sequential Monte Carlo method is a technique that uses recursive Bayesian

approaches to estimate the state of a dynamic system that changes over time using a sequence of noisy measurements made on the system.²⁶ Because of its flexible and powerful diagnostic and prognostic features for nonlinear and non-Gaussian systems, application of PF in reliability engineering has increased rapidly in the recent years.²⁷ PF is able to process data online as it arrives, which is crucial both from the point of view of storage costs and also for rapid adaptation to changing data characteristics.²⁶ This makes it a proper choice for modeling degradation processes with change in degradation rate.

In order to make inference about a dynamic system using particle filtering, at least two models are required; the process model (equation (1)) that describes the evolution of the state with time, and the measurement model (equation (2)) that relates the noisy measurements to the state of the system²⁶

$$d_j = f_j(d_{j-1}, V_{j-1}) \rightarrow Pr(d_j | d_{j-1}) \quad (1)$$

where d represents the state of the system (in this article maximum pit depth), f represents a possibly nonlinear process function, j represents the time index, and V is an i.i.d process noise

$$y_j = g_j(d_j, \omega_j) \rightarrow Pr(y_j | d_j) \quad (2)$$

where y represents the noisy measurement of the state of the system (in this article measured maximum pit depth), g is a possible nonlinear measurement function, and ω is an i.i.d measurement noise sequence.

In order to infer the posterior density function (pdf) of the state of the system given previous noisy measurements, Bayes' rule can be used according to equation (3)

$$\begin{aligned} Pr(d_j | y_{1:j}) &= \frac{Pr(y_j | d_j) Pr(d_j | y_{1:j-1})}{Pr(y_j | y_{1:j-1})} \\ &\propto Pr(y_j | d_j) Pr(d_j | y_{1:j-1}) \end{aligned} \quad (3)$$

In this equation $Pr(y_j | d_j)$ can be calculated using equation (2) and the prior pdf of the state of the system, $Pr(d_j | y_{1:j-1})$, can be calculated using Chapman–Kolmogorov equation

$$Pr(d_j | y_{1:j-1}) = \int Pr(d_j | d_{j-1}, y_{1:j-1}) Pr(d_{j-1} | y_{1:j-1}) dx_{j-1} \quad (4)$$

Assuming that the measurements are conditionally independent and also assuming first-order Markovian property, equation (4) can be simplified as equation (5)

$$Pr(d_j | y_{1:j-1}) = \int Pr(d_j | d_{j-1}) Pr(d_{j-1} | y_{1:j-1}) dx_{j-1} \quad (5)$$

In this integral, the first term can be calculated using equation (1) and the second term can be calculated recursively forward in time by assuming that the pdf of the initial condition of the state of the system, $Pr(d_0)$, is known. The denominator in equation (3) is a normalizing factor which is independent of the state of the system and usually does not have an analytical closed form solution, and numerical solution is usually computationally expensive. In PF, there is no need to calculate the denominator.

Except for special cases (i.e. linear Gaussian state space models), it is not possible to evaluate the posterior distribution in equation (3) analytically. The key idea in PF is to approximate the pdf of the state of the system with a discrete weighted distribution of some random samples (i.e. particles) (equation (6))

$$Pr(d_j | y_{1:j}) \simeq \sum_{p=1}^P w_j^p \delta(d_j - d_j^p) \quad (6)$$

In this equation, δ represents the Dirac's delta function, w_j^p represents the normalized weight of the p th particle at the j th time step, and P is the number of particles. In order to perform PF, P number of samples or particles are generated from initial pdf of the state of the system and then at each time step, those particles are evolved using the process model (prediction step). Subsequently, the measurements corresponding to that time step will be used to update the assigned weight to each particle (updating step).²⁶ Those weights are chosen using the principle of importance sampling.^{28,29} The concept of importance sampling is as following. Suppose $e(d) \propto r(d)$ is a probability density function that is difficult to draw samples from (e.g. posterior distribution of nonlinear non-Gaussian systems in Bayes' rule in equation (3)). But, we can easily sample from another pdf, $s(d)$ (e.g. a normal distribution). In this case a weighted approximation of $e(d)$ can be obtained using equation (7)

$$e(d) \approx \sum_{p=1}^P w^p \delta(d - d^p) \quad (7)$$

where $w^p \propto r(d^p)/s(d^p)$ is the normalized weight of the p th sample (i.e. particle).

Using this concept, it can simply be proven²⁶ that the sequence of the assigned weight of particles at each time can be obtained by equation (8)

$$w_j^p \propto w_{j-1}^p \frac{Pr(y_j | d_j^p) Pr(d_j^p | d_{j-1}^p)}{G(d_j^p | d_{j-1}^p, y_j)} \quad (8)$$

Using this equation in equation (6), the posterior distribution of the state of the system can be approximated.

In the standard PF, it is assumed that the parameters of the process model are known. However, for most of the practical cases, those parameters are unknown, but the form of the process model is known based on the physics of the process. In that case, APF can be used to estimate the state of the system and the process model parameters simultaneously. The process model in APF is shown in equation (9)

$$d_k = f_k(d_{k-1}, \theta_k, V_{k-1}) \rightarrow Pr(d_k | d_{k-1}, \theta_k) \quad (9)$$

where θ represents the vector of the state model parameters.

Kitagawa³⁰ and Liu and West³¹ used a Gaussian random walk to define the evolution model for degradation model parameters to enable their adaptation to new data. It has been identified in Liu and West³¹ and Doucet and Tadić³² that using random walk results in posteriors more diffused than the actual one. To solve this issue, Liu and West³¹ proposed a kernel smoothing approach to reduce the variability in the posterior distributions. Following that approach, the posterior distribution of the model parameters can be approximated by equation (10)³¹

$$Pr(\theta_j | y_{1:j}) \simeq \sum_{p=1}^P w_j^p N(\theta_j | \mu_j^p, h^2 \zeta_j) \quad (10)$$

where $N(\cdot | \mu, S)$ is a multivariate normal density with mean μ (equation (11)) and variance S . In this equation, h is the kernel smoothing parameter and ζ is Monte Carlo posterior variance

$$\mu_j^p = h\theta_j^p + (1 - h)\bar{\theta}_j \quad (11)$$

Selection of kernel smoothing factor is also a challenge in using APF. Based on the prior knowledge, if the parameters are slowly varying or if they are fixed, the smoothing factor should be set to a small positive value (e.g. $0 < h < 0.2$) to reflect the steady property of the parameters. However, when the parameters are expected to change significantly over time, the h value should take a value close to one (e.g. $0.8 < h < 1$) to incorporate the dynamic behavior of the process.³³ The kernel smoothing factor can also be tuned on a validation data set and then be applied to the future data.³³

In this work, the latter approach is followed. We considered 70% of the OLI data, up to time T , as the validation data set and we find the optimum h value that gives the minimum root mean square error (RMSE) between the online measurements and the predictions of APF.

For the case of pitting corrosion, it is well accepted that maximum depth of a pit follows a power function

with a positive exponent less than one (equation (12))^{2,24}

$$d_j = k(t_j - t_0)^\nu \quad (12)$$

where k and ν represent the parameters of pitting corrosion degradation model and t_0 represents the pit initiation time. The recursive format of this model to be used in APF analysis is shown in equation (13). In this equation, a white Gaussian noise with mean zero and standard deviation PM-SD is assumed as the state model noise

$$d_j = d_{j-1} + k\nu(t_j - t_0)^{\nu-1} \Delta t + N(0, \text{PM} - \text{SD}) \quad (13)$$

By considering a general form of the measurement model for an inspection tool, which includes both the biased and random scattering errors, the actual and measured maximum depth of a pit are related according to equation (14)

$$y_{ij} = a_j + b_j d_{ij} + N(0, \epsilon_{ij}) \quad (14)$$

where y_{ij} represents the measured maximum depth of i th pit at the j th inspection ($j = 1, 2, \dots, n$), a_j and b_j are the constant and the proportional biases of the inspection tool employed in the j th inspection, and ϵ_{ij} denotes the SD of normally distributed random scattering error associated with the measured depth of i th pit at the j th inspection. Using equation (14) the measurement model in APF can be derived according to equation (15)

$$Pr(y_{ij} | d_{ij}) = \frac{1}{2\pi\epsilon_{ij}^2} \exp\left(-\frac{(y_{ij} - (a_j + b_j d_{i,j}))^2}{2\epsilon_{ij}^2}\right) \quad (15)$$

Using equations (13) and (14), the pseudocode in Table 2 has been used in this study for APF analysis.

HB method

Another method that is used in this framework is a HB method based on a non-homogeneous gamma process.¹¹ HB modeling is an appropriate method to make scientific inference about a population, based on many individuals, and it is called “hierarchical” because it uses hierarchical or multistage prior distributions.³⁴ This method is used in this framework to fuse ILI data of various pits along the pipeline.

Since 1975 when the gamma process was introduced in the area of reliability engineering,³⁵ it has been used widely to model degradation processes such as corrosion, wear, and fatigue, which involve monotonically accumulating damage over time in a sequence of tiny increments.^{36,37}

A gamma process is a continuous-time stochastic process $\{X(t), t > 0\}$ with the following properties:

Table 2. Pseudocode for APF.

For $p = 1:P$
 Sample d_0^p from maximum pit depth prior distribution.
 Sample θ_0^p from model parameters prior distributions.
 Normal (prior value, $0.1 \cdot \text{prior value}$).
 Calculate $\bar{\theta}_0, \text{var}(\theta_0)$.
 Assign particles' weight: $w_0^p = 1$.
 End
 For $j = 1$: number of OLI data
 $k_j = (1 - h^2)^{0.5} k_{j-1} + (1 - (1 - h^2)^{0.5}) \bar{k}_{j-1}$
 $v_j = (1 - h^2)^{0.5} v_{j-1} + (1 - (1 - h^2)^{0.5}) \bar{v}_{j-1}$
 Prediction step
 $d_j^p = d_{j-1}^p + kv(t_j - t_0)^{\nu-1} + \text{rand.N}(0, \text{PM} - \text{SD})$
 $k_j = k_{j-1} + \text{rand.N}(0, h^2 \text{var}(k_{j-1}))$
 $v_j = v_{j-1} + \text{rand.N}(0, h^2 \text{var}(v_{j-1}))$
 Updating step
 $w_j^p = w_{j-1}^p * \text{Pr}(y_j | d_j, k_j, v_j)$
 Normalize the weights
 Resample d_j, k_j, v_j
 End

APF: augmented particle filtering; OLI: online inspection.

- $X(0) = 0$ with the probability 1.
- $\Delta X = X(\tau) - X(t) \sim \text{Ga}(\Delta\alpha = (\alpha(\tau) - \alpha(t)), \beta)$ for all $0 \leq t < \tau$.
- $X(t)$ has independent increment.

In the above-mentioned properties, Ga represents pdf of gamma distribution. A random quantity (in this study maximum pit depth) has a gamma distribution with shape parameter $\alpha > 0$ and a rate parameter (inverse of scale parameter) $\beta > 0$ if its pdf is given by

$$f_{X(t)}(x) = \text{Ga}(x; \alpha, \beta) = \frac{\beta^\alpha(t)}{\Gamma(\alpha(t))} x^{\alpha(t)-1} \exp(-\beta x) \quad (16)$$

where $\Gamma(\cdot)$ denotes the gamma function. The expectation and variance of the gamma process are given in equations (17) and (18), respectively

$$E(X(t)) = \frac{\alpha(t)}{\beta} \quad (17)$$

$$\text{Var}(X(t)) = \frac{\alpha(t)}{\beta^2} \quad (18)$$

According to equation (17), the shape parameter of a gamma process reflects the average trend of the random quantity as a function of time. Therefore, by selecting an appropriate form for the shape parameter of a gamma process, it can model degradation processes with increasing, decreasing, or constant degradation rates. For pitting corrosion process, as it was mentioned previously, a well-accepted format of its degradation

model is shown in equation (12). Therefore, in that case, the shape parameter of a gamma process is correlated with the degradation model parameters according to equation (19)

$$\alpha_j = k'(t_j - t_0)^{\nu'} \quad (19)$$

Based on this assumption, the increments in degradation level follow a gamma distribution given in equation (20)

$$\Delta d_{ij} = d_{ij} - d_{i,j-1} \sim \text{Ga}(\Delta\alpha_{ij} = (k'((t_j - t_0)^{\nu'} - (t_{j-1} - t_0)^{\nu'}), \beta_i) \quad (20)$$

When new measurements are available at each inspection time, the posterior distribution of the depth increment of each pit will be updated using equation (21)

$$\text{Pr}(\Delta d_{ij} | Y_i) \propto \text{Pr}(Y_i | \Delta d_{ij}) \text{Ga}(\Delta d_{ij} | \Delta\alpha_{ij}, \beta_i) \quad (21)$$

In this equation, the likelihood of the inspection data Y_i given the increments can be written as shown in equation (22) by considering the measurement model that is given in equation (14)

$$\text{Pr}(Y_i | \Delta D_i) = (2\pi)^{-n/2} \exp(-1/2(Y_i - (A + BS_{\Delta D_i}))' \times \sum_{\epsilon}^{-1} \times (Y_i - (A + BS_{\Delta D_i}))) \quad (22)$$

where $Y_i = (y_{i1}, y_{i2}, \dots, y_{in})'$, $A = (a_1, a_2, \dots, a_n)'$, B is an n -by- n diagonal matrix with diagonal elements equal to b_j , \sum_{ϵ}^{-1} is the n -by- n diagonal covariance matrix with diagonal elements equal to the variance of the random scattering errors associated with the tool used in inspection time j , and $S_{\Delta D_i}$ is an $n \times 1$ vector with the j th element equal to $\sum_{k=1}^j \Delta d_{ik}$.

The HB model that is used to estimate these hyperparameters $(t_0, p_1, q_1, p_2, q_2, p_3, q_3)$ is shown in Figure 1.

Proposed framework

Consider a long piggable oil or gas pipeline, for which n ILI data sets are available for m number of active pits at times t_1, t_2, \dots, t_n . In addition, an active pit (the reference pit, pit M) is monitored continuously using an online inspection tool. This pipeline is in operation since time t_0 , which is assumed to be the initiation time for all pits. The operational conditions are monitored and measured as part of the routine operating condition monitoring procedure of the pipelines. The question, which is addressed in this article, is how to estimate the maximum depth of the existing pits at time t_{n+1} when the operational conditions change at time T , $t_n < T < t_{n+1}$, and there is no new ILI data for those pits.

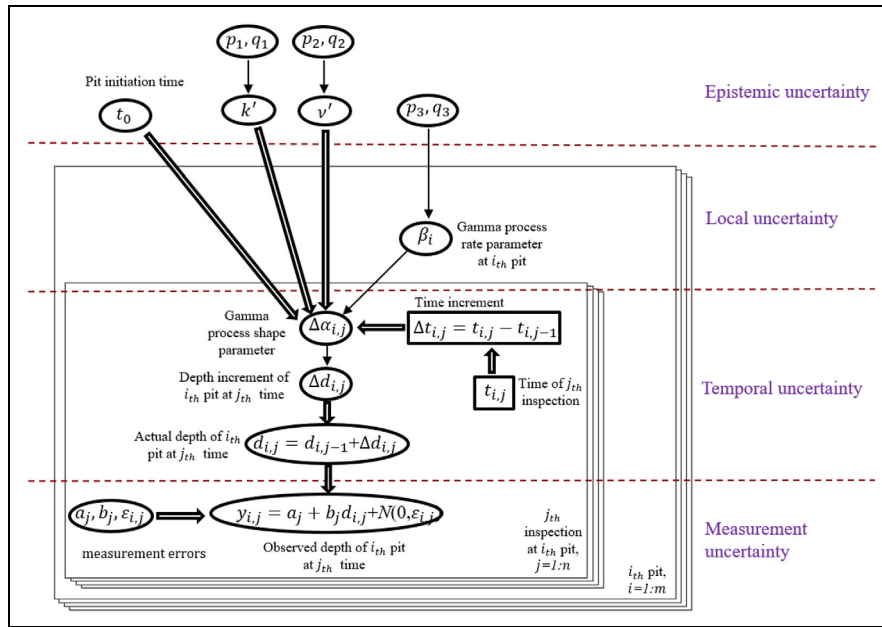


Figure 1. Hierarchical Bayesian model based on a non-homogeneous gamma process modified from the approach that is proposed by Maes et al.¹¹

In order to answer the above-mentioned question, we propose a data fusion framework that has three phases that are shown in Figure 2. The required input data for this framework are shown in the left side of this figure. In phase I, prior values for the degradation model parameters and the SD of the state model noise in APF are estimated for use in phases II and III. In phase II, a similarity index (SI) between pit i ($i = 1, 2, \dots, m$) (an ILI pit) and pit M (the reference pit) is defined. Finally in phase III using that SI, some dummy observations are generated and used to estimate the maximum depth of each ILI pit at time t_{n+1} . These phases are explained in more detail in the next paragraphs by referring to the steps in Figure 2.

Phase I: estimating the SD of the white noise of APF process model (PM-SD) and the prior values for degradation model parameters

In phase I, the SD of the process model white noise in APF analysis and also the prior values for degradation model parameters are estimated using the historical data of operational parameters and corrosion rate of the considered pipeline or pipelines under similar operational condition. Practically, operational parameters are measured by routine monitoring of the pipeline at a limited number of locations, and it is not feasible to measure them in all locations. However, valuable information about the physics of the corrosion failure mechanism are embedded in those limited data. In this phase, we propose an approach to use those data to estimate the

noise of process model in PF. In this way, a generic degradation model (equation (23)) is developed for all pits using a multivariate nonlinear regression analysis (Step I-1), to correlate the average of the maximum pits depth (\bar{d}) with the operational parameters (e.g. pressure, temperature, and pH) and time

$$\bar{d}(t) = f(P, T, pH, \dots, t) \tag{23}$$

This model is then used to simulate realizations of actual (vs measured) maximum depth growing behavior for a number of pits (Step I-2). Figure 3 shows an example of these realizations. In order to simulate those realizations, at each time interval (Δt), new samples should be extracted from the pdf of each operational parameter (from Step In-1) to be inserted in the developed generic model (equation (23)) to obtain the corresponding pit depth increment Δd . This depth increment will be used recursively to simulate those realizations (equation (24))

$$\bar{d}(t) = \bar{d}(t - 1) + \Delta \bar{d} = \bar{d}(t - 1) + \frac{\partial f}{\partial t}(P, T, pH, \dots, t) \Delta t \tag{24}$$

In Step I-3, the SD of depth increments is calculated for each pit and then the PM-SD is estimated as the average of those SDs of all pits (equation (25))

$$PM - SD = \frac{1}{m} \sum_{i=1}^m STD \text{ of } \Delta \bar{d}_{i,j} \tag{25}$$

where j is the time step index and i is the pit index.

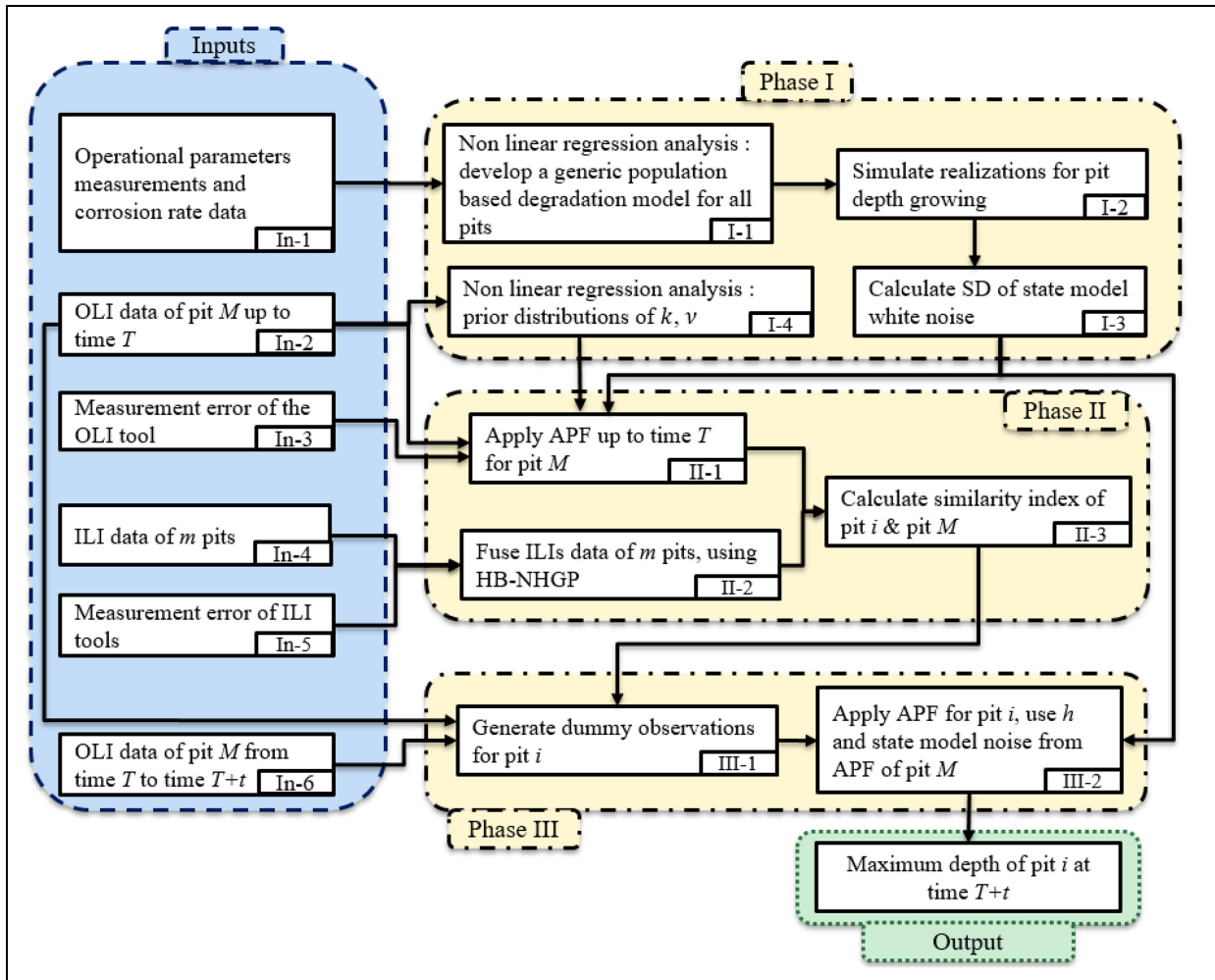


Figure 2. Flowchart of the proposed framework.

In addition, using regression analysis on the OLI data up to time T (Step In-2), and fitting a power law function, prior values for the degradation model parameters (k , ν) are obtained in Step I-4.

Phase II: defining a similarity index between pit i and pit M

In phase II, a similarity index (SI) is defined (Step II-3) between each ILI (pit i) and pit M . This SI is defined as a ratio of the estimated maximum pit depth (EMPD) of pit i over EMPD of pit M at ILI times (equation (26))

$$SI(i) = \frac{1}{nP} \sum_{j=1}^n \sum_{p=1}^P \frac{d_{jp} \text{ of pit } i \text{ by HB - NHGP}}{d_{jp} \text{ of pit } M \text{ by APF}} \quad (26)$$

where n is the number of ILI operations and P is the number of particles and p is the particle index.

The denominator is an estimation of the maximum pit depth of the reference pit M and it is obtained by using APF to fuse the OLI data (Step In-3) of this pit up to time T . In the denominator, d_{jp} is the state of particle p at the j th ILI. The other inputs for APF analysis of pit M are the preliminary estimation of k and ν (Step I-4) and the estimated PM-SD (Step I-3) and the characteristics of the OLI tool (Step In-3).

The numerator of this ratio is estimated by fusing ILI data of all pits, using the HB model based on a non-homogeneous gamma process (HB-NHGP) that is explained in the HB method section (Step II-2). Inputs of this step are ILI data of all pits (Step In-4), and the characteristics of ILI tools (Step In-5), including biased and scattering errors of those tools (a , b and ϵ in equation (14)). The output of Step II-2 is an estimate of mean and SD of the posterior distribution of EMPD of each ILI pit. We used Monte Carlo simulation to extract random samples from that posterior

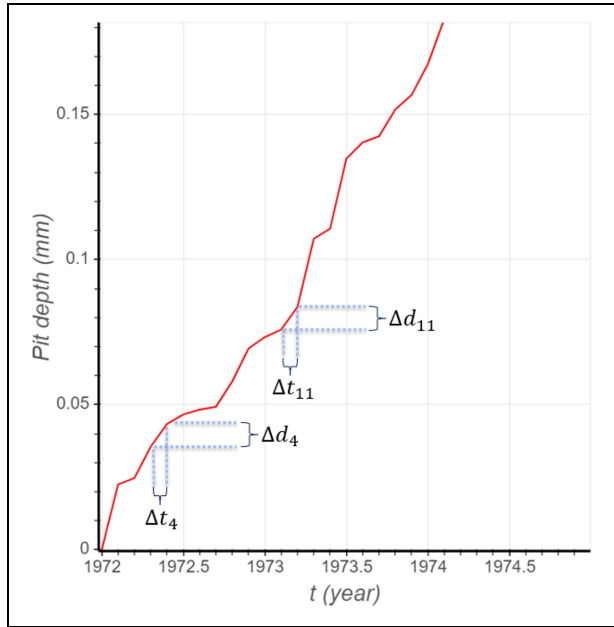


Figure 3. A realization of simulated actual maximum pit depth.

distribution to have an estimation corresponding to each particle (d_{jp} in the numerator).

Phase III: inferring the degradation level of pit i

Finally in phase III, APF is used to estimate maximum depth of pit i at time t_{n+1} (Step III-2) by fusing the generated dummy observations for that pit. Those dummy observations are generated by multiplying the real OLI data of pit M by the corresponding SI of pit i (Step III-1). In this phase, h and PM-SD that were estimated for the OLI pit previously, are used for ILI pits as well, because these two parameters show the stochasticity of a stochastic process at each time and OLI and ILI pits are exposing to the same corrosion environment at each time.

Demonstration of the proposed framework

In this section, the proposed framework is demonstrated in a case study. Consider a long oil or gas pipeline (e.g. 50 miles length) in operation since 1972. This pipeline is inspected by ILI, ultrasonic test, in years 2000, 2005, 2010, and 2015. A number of active pits are detected and monitored at those times. After the first ILI, an OLI sensor is installed to monitor the degradation behavior of an active pit (the reference pit) continuously. The operational conditions change causing change from moderate to severe corrosion condition in 2015. The goal is to estimate the maximum depth of ILI pits in 2020 when there is no new ILI data available, by inferring from OLI data of the reference pit.

Performing phase I of the proposed framework

In practice input data given in Steps In-2, In-4, and In-6 from Figure 2 should be gathered from ILI and OLI of the pipeline. These input data can hardly be found altogether in the existing literature for a pipeline. Therefore, we used the model that was developed in Ossai et al.²³ as the starting point (output of Step I-1) to generate synthetic actual (vs measured) depths in Step I-2. We added random measurement noise to those synthetic actual depths to generate synthetic ILI and OLI data for this case study. The characteristics of the inspection tools that are given in Zhang and Zhou¹⁰ are used in Steps In-3 and In-5. This synthetic data generation procedure is explained in more detail as following.

Synthetic data generation procedure. We reviewed different pitting corrosion degradation models⁸ and among them we used a model that was developed by Ossai et al.,²³ as the output of Step I-1. This model is chosen because it has been developed based on the field data (rather than experimental data) and to the best of our knowledge, that model is the most comprehensive available generic internal pitting corrosion degradation model in the literature, that correlates 11 covariates (Table 3) with the average maximum pit depth over time. This model has been developed using 10 years of measurement data from 60 X52 pipelines that were used for oil and gas pipelines in Nigeria. Ossai et al. carried out multivariate regression modeling to develop this model which is shown in equation (27)

$$\bar{d}(t) = k(t - t_0)^\nu = \exp\left(\gamma_0 + \sum_{q=1}^Q \gamma_q V_q\right) (t - t_0)^\nu \quad (27)$$

where \bar{d} represents the average maximum pit depth, t represents time of evaluation, k and ν are the power law model parameters, t_0 is the pit initiation time (Ossai et al. assumed that pit initiation time is equal to the operation initiation time for all pits), γ_0 represents the intercept, γ_q represents the mean value of the regression coefficient (Table 4) of the q th operational parameter, V_q represents q th operational parameter (Table 3), and Q represents the number of operational parameters.

Considering natural log of the mean value of the operational parameters and the mean value of the estimated regression coefficients in equation (27), the average maximum pit depth is determined using equation (28)²³

$$\bar{d}(t) = 0.732 t^{0.803} \quad (28)$$

We used this model as the degradation model for time $t > T$ ($T = 2015$ in this case study). For time $t < T$

Table 3. Best fit distribution of the operational parameters²³ and modified values for moderate corrosion rate category.

Parameters	Units	Description	Best fit distribution	Original distribution parameters Scale, shape	Modified distribution parameters Scale, shape
T	°C	Temperature	Lognormal	3.72, 0.4052	3.35, 0.12
Pc	MPa	CO ₂ partial pressure	Weibull	0.1598, 1.2797	0.065, 1.25
pH	–	pH	Extreme value	7.9418, 0.4747	8.2, 16.5
S	mgL ⁻¹	Sulfate ion	Weibull	38.9576, 0.4052	40.5, 1.5
C	mgL ⁻¹	Chloride ion	Weibull	3613.8, 1.3	1413.8, 1.5
W	–	Water cut	Lognormal	–1.7178, 1.4696	3.15, 0.8
R	Pa	Wall shear stress	Lognormal	3.447, 0.9151	3.447, 0.9151
Gs	m ³ day ⁻¹	Gas production rate	Extreme value	335,310, 120,120	335,310, 120,120
OL	m ³ day ⁻¹	Oil production rate	Weibull	136.33, 2.1145	136.33, 2.1145
Wt	m ³ day ⁻¹	Water production rate	Weibull	94.9241, 0.4847	94.9241, 0.4847
Pt	MPa	Operating pressure	Extreme value	8.1274, 3.2704	8.1274, 3.2704

Table 4. Parametric estimate for power model development.²³

Estimate	Coefficients (γ_q)	Standard error	t-stat	p-value
Log(T)	0.037	0.083	0.4465	0.6616
Log(Pc)	–0.014	0.0373	–0.3745	0.7133
Log(pH)	–0.8446	0.7418	–1.1386	0.2727
Log(S)	–0.0033	0.0835	–0.0392	0.9692
Log(C)	0.0613	0.0494	1.2388	0.2345
Log(W)	0.042	0.0337	1.2463	0.2318
Log(r)	0.0037	0.0433	0.0857	0.9329
Log(Gs)	–0.0467	0.0554	–0.8441	0.4119
Log(OL)	–0.0002	0.0657	–0.0037	0.9971
Log(Wt)	–0.0076	0.021	–0.3621	0.7223
Log(Pt)	–0.0142	0.0488	–0.2915	0.7746
Intercept γ_0	0.44	0.7572	0.5811	0.5698
Log(t)	0.8032	0.0458	17.5346	0

(from 1972 to 2015) we used the model that has been developed for moderate corrosion rate category in Ossai et al.⁹ (equation (29))

$$\bar{d}(t) = 0.269t^{0.741} \quad (29)$$

Since the best fit probability distribution of the operational parameters for the moderate corrosion rate category are not given in Ossai et al.,⁹ we modified the scale and the shape parameters of the distributions that are given in Table 3, to have approximately the same mean and SD that are given for moderate corrosion rate category in Table 1 of Ossai et al.⁹ (Ossai et al.^{9,23} are based on the same data set). These modified values are given in the last column of Table 3. We used Monte Carlo simulation to sample from the distributions of the operational parameters that are given in Table 3.

In addition, the variation in the estimated coefficients of the degradation model is also taken into account in the synthetic data generation by considering

the given standard errors. The details are given in the following pseudocode.

In this pseudocode (Table 5), std.N.rand is a positive random number generated from the standard normal distribution. DOF is the degree of freedom of the student's t -distribution which is equal to the number of samples minus one, minus number of parameters (in this case 11). According to Ossai et al.,⁹ number of samples is more than 70 and 300 for moderate and severe corrosion rate categories, respectively. Therefore, DOF is more than 30, for both cases. Which means student's t -distribution can be approximated by a standard normal distribution³⁸ and DOF does not play an important role in sampling process.

In addition, in order to take into account the temporal variation of corrosion process, we sampled from the parameters and the coefficients distribution every 0.1 year (Δt in Figure 3). We assumed that there are 100 pits on this pipeline. This pit density is selected based on examples in the literature (e.g. 62 pits in 80 km,¹⁰ 554 pits in 129 km,³⁹ 1 pit per km).⁴⁰ These examples show that the pit density is a small number and we can reasonably assume that the pits are not interacting with each other. However, in case of interaction between pits, the common practical and conservative approach can be used which is to coalesce the adjacent pits by following available codes (e.g. DNV RP-F101)⁴¹ and considering the composite pit in this framework.

Having this synthetic data, an estimation for the SD of the white noise of the state model in APF is calculated using equation (25).

In order to generate ILI data (Step In-4 in Figure 2), the measurement error of ILI tools are added to the synthetic data according to equation (14). We used the same equation to consider measurement error in generating OLI data for the reference pit (Steps In-2 and In-6) using characteristics of the OLI tool. This

Table 5. Pseudocode for synthetic data generation.

```

For  $i = 1:m$  ( $m$ , number of pits = 100 in this study)
   $d_{i,0} = 0$ 
  For  $j = 1$ : no. of time steps
    Generate  $V_{q,i,j}$  by Monte Carlo simulation (using the
    distribution parameters in Table 3).
    Generate  $V_{q,i,j} = \gamma_q$  (given in Table 4) +
     $t.\text{inv}(\text{std.N.rand}(i,j), \text{DOF}) \times$  corresponding
    standard error given in Table 4.
     $\bar{d}_{i,j} = \nu \times \exp\left(\gamma_0 + \sum_{q=1}^Q \gamma_{q,i,j} V_q\right) \times t^{\nu-1}$ 
     $\Delta \bar{d}_{i,j} = \bar{d}_{i,j} - \bar{d}_{i,j-1}$ 
  End
End
    
```

information for each ILI tool and also the OLI tool is given in Table 6. We assumed that the scattering error is independent and identically distributed for each pit at each time and it follows a white noise with mean value equal to zero and SDs ($\epsilon_{i,j}$) that are given in Table 6 for ILI and OLI tools.

Figure 4 shows an example of synthetic actual and measured maximum pit depth for the OLI and an ILI pit. This figure shows that the frequency of the OLI data is higher than the frequency of ILI data. In addition, the measurement error of OLI tool is smaller than the measurement error of ILI tool.

Performing phase II of the proposed framework

The assumptions on phase II for this case study are as following. In Step II-1, 10,000 particles are randomly selected to approximate the posterior distributions of the maximum depth and the degradation model parameters by APF. In order to select a proper value for h (kernel smoothing factor), we used 70% of OLI data up to time T (i.e. 2015 in this case) as the training data set for APF, and then we found the optimal value of h which gives the minimum RMSE between the predicted value by APF and the measured maximum pit depth by online inspection for the test data set (e.g. remaining 30% of the OLI data). Based on this approach the selected h value for this case study is 0.01. We also assumed that degradation model parameters follow

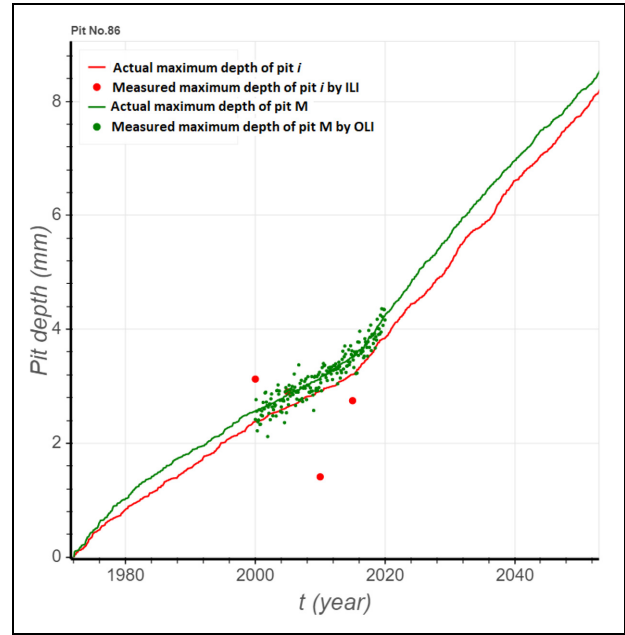


Figure 4. Generated synthetic data for an OLI and an ILI pit.

normal distributions with mean values that obtained in Step I-4 (e.g. we used 0.23 and 0.73 for a seed number) and SD equal to 0.1 of the corresponding mean value. In addition, the PM-SD of process model noise is obtained in Step I-3 (e.g. we used 0.008 for a seed number).

In Step II-2 the posterior probability distribution of maximum depth of ILI pits at ILI times are obtained. In this step, a gamma distribution with shape and scale parameters equal to ten and one, is selected as the prior distribution for k . The same assumption has been made for ν and β_i with the shape and scale parameters both equal to one.¹⁰ The generated synthetic ILI data of 100 pits are used as the evidence in the HB analysis to update the posterior distributions of the maximum pit depth (d). We employed the Markov chain Monte Carlo simulation technique using the software OpenBUGS for this analysis. Since this model is a multi-parameter model, we run two chains, starting at two different points to decide when convergence to the posterior distribution has occurred.³⁴ For each chain

Table 6. Constant (a) and proportional (b) biased error and scattering error of inspection tools.¹⁰

	First ILI	Second ILI	Third ILI	Fourth ILI	OLI
a	2.04% PWT	2.04% PWT	-15.28% PWT	-10.38% PWT	0
b	0.97	0.97	1.4	1.13	1
c	5.97% PWT	5.97% PWT	9.05% PWT	7.62% PWT	2.0% PWT

ILI: in-line inspection; OLI: online inspection; PWT: pipe wall thickness.

100,000 simulation sequences were generated and 10,000 sequences were discarded as the burn in period. A thinning interval of 10 was selected to reduce the auto-correlation between the samples.

In Step II-3, by having estimations of maximum depth of ILI and OLI pits from Steps II-1 and II-2, a similarity index between each ILI and the reference pit is defined using equation (26).

Performing phase III of the proposed framework

Finally in Step III-2, the posterior distribution of maximum depth for each ILI pit at 2020 is estimated by following the given APF pseudocode in Table 2 and using dummy observations of each in-line inspected pit. Those dummy observations were generated in Step III-1 following the procedure that is explained in phase III. The results of this case study are discussed in the next section.

Results

In this section, the results of APF analysis in estimation of maximum depth for the OLI pit is discussed first. Then, the performance of this framework is validated by comparing its results with the results of Maes model, assuming there is no change in operational conditions. Finally, assuming the operational conditions change at time T (i.e. 2015 in this case), from moderate to severe corrosion rate category, we illustrate the effects of considering this change (and using this framework) on pipeline degradation level estimation.

Figure 5 shows the estimation of the model parameters for the reference pit using APF, when the operational conditions are the same during pipeline life-cycle. This figure depicts that approximately 7 years after the first inspection, when there is enough data to update the posterior distribution of the model parameters, the variation of the model parameters reduces and they converge to constant values that can be used for prognostic purposes.

Figure 6 shows the estimated maximum depth of pit M using APF. According to this figure, between years 2000 and 2020, that OLI data are available, there is no significant error in maximum pit depth estimation. After 2020, as it was expected for particle filtering method, the estimation error increases over time because of lack of new inspection data. However, even after 2020, the actual maximum pit depth is within the lower and upper bounds of this estimation. Figures 4 and 5 indicate that APF is an appropriate method to estimate maximum pit depth, when online inspection data are available.

In order to validate the performance of this framework we defined two metrics. The first validation metric

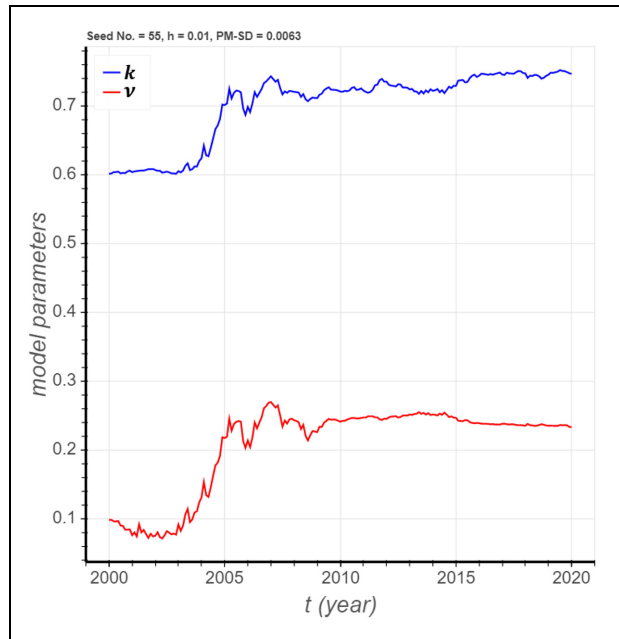


Figure 5. Estimated model parameters for the reference pit using APF for a random seed number.

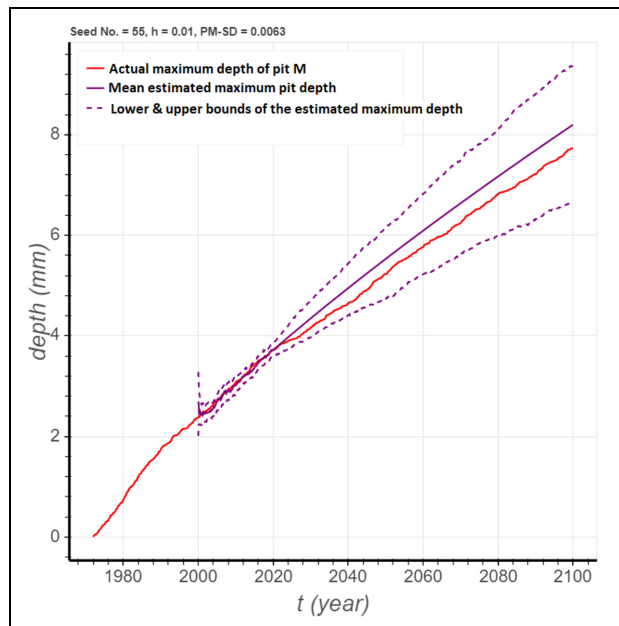


Figure 6. Estimated maximum depth of the reference pit by APF.

(Metric R) is the RMSE of maximum pit depth prediction at 2020 (equation (30)). This RMSE should be less than or almost equal to the RMSE of Maes model. The second validation metric (Metric N) is the percentage of all pits that their predicted depths fall within the $\pm 10\%$

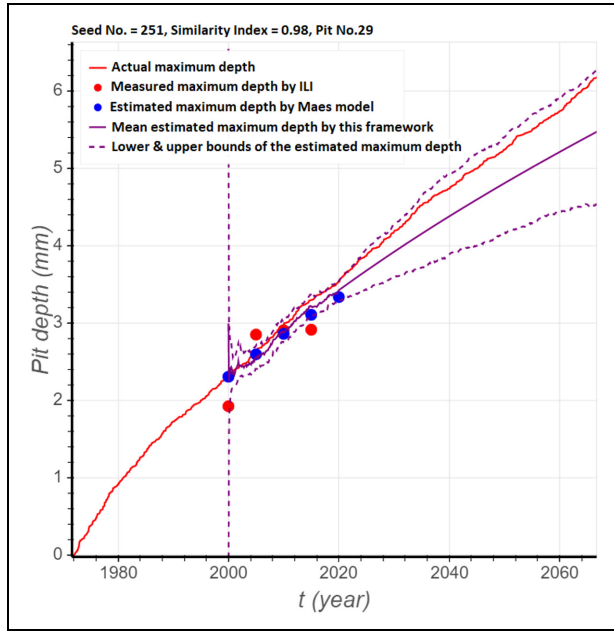


Figure 7. Estimated maximum depth of pit No. 29 by this framework and Maes model, without change in operational conditions.

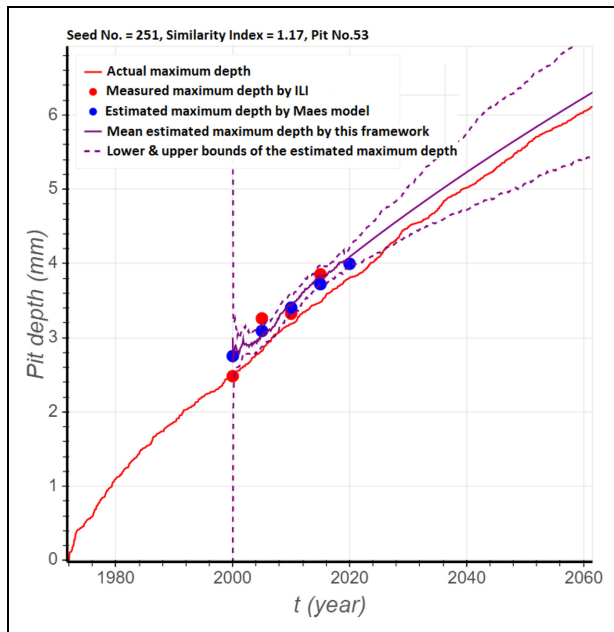


Figure 8. Estimated maximum depth of pit No. 53 by this framework and Maes model, without change in operational conditions.

of their actual maximum depth. The $\pm 10\%$ PWT is commonly used in the pipeline industry as a confidence interval for the accuracy of the inspection tools¹⁰ and we modified that in this work as a metric for accuracy

of the prediction. This metric should be greater than or equal to the one for Maes model. The second metric is defined because a pipeline is a series system and failure at each location (i.e. pit) is equal to the failure of the whole pipeline.^{42,43} Therefore having smaller RMSE, does not necessarily mean that the proposed framework has a better performance. It might be the case that the estimation errors of a few pits are so small that it causes decrement in the RMSE, but for the majority of the pits, the estimation error have increased. Hence, both of these metrics should be satisfied to conclude that the performance of the proposed framework is at least as good as the performance of Maes model when the operational conditions do not change.

$$RMSE = \sum_{i=1}^m \frac{((d_i - y_i)^2)^{0.5}}{m} \quad (30)$$

Figures 7 and 8 depict two examples of the estimated maximum depth at 2020 for two ILI pits using this framework and Maes model. The SI are 0.98 and 1.17 for pit No. 29 and pit No. 53, respectively. According to Figure 7, for pit No. 29 the estimation error of this framework is less than the estimation error of Maes model and according to Figure 8, for pit No. 53 it is vice versa. In order to quantify the estimation error for all pits to compare the performance of Maes model and this framework, the RMSE for both models is calculated. The results for some random seed numbers are given in Table 7. As shown in this table, the ratio of the RMSE of this framework over the RMSE of Maes model is around one for all seed numbers (30 seed numbers) and the average of that ratio for all seed numbers is 1.010. With respect to Metric N, on average for 73.6% and 74.9% of pits, for Maes model and this framework respectively, the predicted depth is within the $\pm 10\%$ of the actual maximum pit depth. These metrics indicate that when there is no change in operational conditions, the performance of this framework is similar to the performance of Maes model as a validated state of the art pitting corrosion degradation model for piggable pipelines.

Relying on the validation results, we used this framework to estimate maximum pit depth at year 2020 when operational conditions change at 2015. Figures 9 and 10 are two examples of the maximum pit depth estimation for pit No. 50 and pit No. 31, respectively. Pit No. 50 is an example that shows when operational conditions change, using this framework decreases the estimation error, however, the opposite is true for pit No. 31. In order to compare the estimation error for all pits, Metrics R and N are given in Table 8, for both methods and for some seed numbers. For example for seed number 3987, the RMSE of Maes model is 0.519 and for

Table 7. Comparing the results in case of no change in operational conditions.

Seed no.	Metric R (mm), Maes	Metric R (mm), Framework	Ratio	Metric N (%), Maes	Metric N (%), Framework
1653	0.324	0.308	0.951	72	75
251	0.331	0.308	0.951	69	77
2652	0.337	0.293	0.870	61	73
5,412,875	0.352	0.352	1.000	68	67
93	0.291	0.365	1.259	78	69
...
3987	0.288	0.288	0.997	79	78
Average	0.316	0.318	1.010	74	75

Table 8. Comparing the results in case of considering change in operational conditions.

Seed no.	Metric R (mm), Maes	Metric R (mm), Framework	Ratio	Metric N (%), Maes	Metric N (%), Framework
1653	0.631	0.344	0.545	31	73
251	0.596	0.407	0.683	40	67
2652	0.599	0.328	0.549	29	73
5,412,875	0.543	0.391	0.718	44	69
93	0.612	0.367	0.601	27	77
...
3987	0.519	0.322	0.621	42	75
Average	0.556	0.334	0.634	39	75

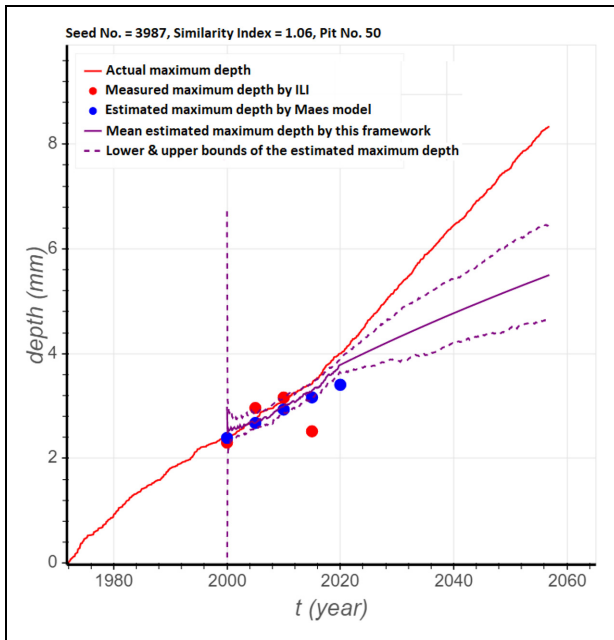


Figure 9. Estimated maximum depth of pit No. 50 by this framework and Maes model with change in operational conditions.

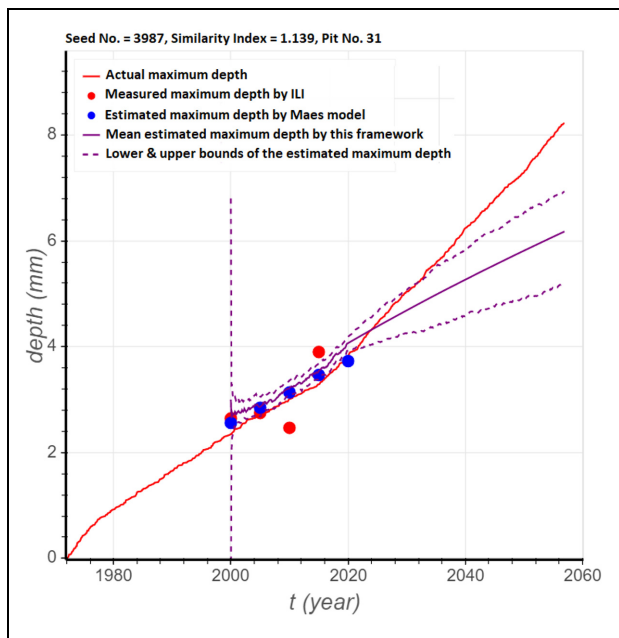


Figure 10. Estimated maximum depth of pit No. 31 by this framework and Maes model with change in operational conditions.

this framework it is 0.322. Which means the RMSE of this framework is 62% of the RMSE of Maes model, that is a significant improvement in accuracy of maximum pit depth prediction. As it was mentioned before,

this metric is necessary but not sufficient to conclude that there is an improvement in degradation level estimation for the whole pipeline. Metric N must also be considered. As it is shown in Table 8, for seed number

3987, Metric N is 42 for Maes model and it is 75 for this framework. This means that for seed 3987, out of 100 pits, the Maes model prediction for 42 of them is within $\pm 10\%$ of their actual maximum depth, whereas our framework is accurate for 75. The average of these two metrics for all seed numbers are also given in Table 8. According to this table, on average, using this framework the RMSE is 40% lower than the RMSE of using Maes model when there is change in operational conditions. In addition, for 75.5% of the pits, the predicted maximum depth is within $\pm 10\%$ of their actual maximum depth for this framework in comparison to 39.4% of the pits for Maes model. These results show a significant improvement in maximum pit depth prediction that leads to avoiding either unnecessary maintenance or unpredicted failures.

It is worth noting that when there is a change in operational conditions, going forward in time, even for the proposed framework, prediction error increases significantly because there is not enough observations to update and learn the model parameters properly. However, this error will be reduced by implementing future OLI and ILI inspections.

Conclusion

In this work, a novel data fusion framework is proposed to develop an internal pitting corrosion degradation model for oil and gas pipelines when operational conditions change over time. The change in operational conditions is taken into account by monitoring the change in degradation level of an active pit (the reference pit) and accordingly inferring about the change in degradation level of other active pits. This framework consists of three phases. In phase I, historical data of the considered pipeline or pipelines with similar operational conditions, are used to develop a generic degradation model for all pits. This model is used to generate synthetic actual maximum pit depth realizations for a number of pits. The SD of the white noise that is used in process model in APF is extracted from these synthetic data. In addition, at this phase, prior values for degradation model parameters are obtained by performing a nonlinear regression analysis on the OLI data of the reference pit. In phase II, a similarity index between each ILI pit and the reference pit is calculated as a ratio of the estimated maximum depth of the ILI pit (using a HB method based on a non-homogeneous gamma process) and the estimated maximum depth of the reference pit (using APF) at ILI times. In phase III, dummy online observations are generated for each ILI pit by multiplying its similarity index with the OLI data of the reference pit. Then, those dummy observations

are used in APF to estimate the maximum depth of ILI pits when there is no new ILI data.

The application of this framework is discussed using this framework on a number of pits and the results are compared with the results of a state of the art degradation model (Maes model), that has been validated by real field data and is available in the literature. Two metrics are used to compare the results of this framework with the results of the Maes model. The first metric is Metric R which is the average of RMSE between actual and predicted maximum pit depth for all seed numbers. This metric is necessary but not sufficient to compare the performances of this framework and the Maes model. The reason is that, it is possible that the estimation errors of a few pits are so small that it causes decrement in the RMSE, but for the majority of the pits, the estimation error have increased. We defined the second metric, Metric N, as the percentage of all pits that the corresponding predictions are within $\pm 10\%$ bounds of their actual maximum depth.

When there is no change in operational conditions, in terms of Metric R, the results are approximately the same for the new framework and the Maes model (i.e. on average of approximately 0.31). In addition, on average for 73% of all pits, for both this framework and Maes model, the predicted depth is within the $\pm 10\%$ bounds of their actual maximum pit depth (Metric N).

We also conduct this validation on case study with a change in operational conditions. Based on the presented results, by considering change in operational conditions and using this framework, Metric R, the RMSE, would be approximately 60% of the RMSE of the Maes model which is a significant improvement in maximum pit depth prediction. In addition, in the case of change in operational conditions, the predicted depth of 75.5% of pits are within the $\pm 10\%$ of their actual maximum pit depth using this framework. This number is 39.45% for Maes model. In the other words, this framework provides 91% improvement with respect to the number of pits with a high confidence level estimation. These results show that when there is a change in operational conditions, using the proposed framework resulted in decrease in the prediction error for the majority of the pits. These improvements enable avoiding either unnecessary maintenance or unpredicted failures.

In the next step of this research, we will use the results of this article to estimate the probability of occurrence of different failure modes (i.e. small leak, large leak, and rupture) in a pipeline segment and then define an optimal maintenance policy which takes into account the cost of each failure mode and also the cost of different maintenance actions (i.e. do nothing, sleeving, and replacement). That optimal policy will include

the optimal maintenance action and time for each segment and also the optimal next ILI time for the whole pipeline.

Lack of the real-field corrosion inspection data is a big challenge in PHM of the oil and gas pipelines. Hence, it is highly recommended that oil and gas pipelines' owners and pipeline-operating companies collect the operational conditions and inspection data and make them available in the public domain to make it possible for the researchers to validate their new corrosion degradation models that finally leads to the decrease in the number of unexpected failures and unnecessary maintenance.

Acknowledgements

The authors thank Austin Drisko Lewis, Ramin Moradi, and Amin Aria who reviewed this paper in depth and gave substantial feedback.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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.

ORCID iD

Roohollah Heidary  <https://orcid.org/0000-0002-9915-3335>

References

- Kishawy HA and Gabbar HA. Review of pipeline integrity management practices. *Int J Pres Ves Pip* 2010; 87(7): 373–380.
- Velázquez J, Caleyó F, Valor A, et al. Predictive model for pitting corrosion in buried oil and gas pipelines. *Corrosion* 2009; 65(5): 332–342.
- Papavinasam S, Revie RW, Friesen WI, et al. Review of models to predict internal pitting corrosion of oil and gas pipelines. *Corros Rev* 2006; 24(3–4): 173–230.
- Papavinasam S. *Corrosion control in the oil and gas industry*. Amsterdam: Elsevier, 2013.
- Nešić S. Key issues related to modelling of internal corrosion of oil and gas pipelines — a review. *Corros Sci* 2007; 49(12): 4308–4338.
- Tarantseva K. Models and methods of forecasting pitting corrosion. *Prot Met Phys Chem S* 2010; 46(1): 139–147.
- Caleyó F, Velázquez J, Hallen J, et al. Markov chain model helps predict pitting corrosion depth and rate in underground pipelines. In: *2010 8th international pipeline conference*, Calgary, AB, Canada, 27 September–1 October 2010, pp. 573–581. New York: American Society of Mechanical Engineers.
- Heidary R, Gabriel SA, Modarres M, et al. A review of data-driven oil and gas pipeline pitting corrosion growth models applicable for prognostic and health management. *Int J Progn Health Manag* 2018; 9(1).
- Ossai CI, Boswell B and Davies IJ. Predictive modelling of internal pitting corrosion of aged non-piggable pipelines. *J Electrochem Soc* 2015; 162(6): C251–C259.
- Zhang S and Zhou W. System reliability of corroding pipelines considering stochastic process-based models for defect growth and internal pressure. *Int J Pres Ves Pip* 2013; 111: 120–130.
- Maes MA, Faber MH and Dann MR. Hierarchical modeling of pipeline defect growth subject to ILI uncertainty. In: *ASME 2009 28th international conference on ocean, offshore and arctic engineering*, Honolulu, HI, 31 May–5 June 2009, pp. 375–384. New York: American Society of Mechanical Engineers.
- Heidary R, Groth KM and Modarres M. Fusing more frequent and accurate structural damage information from one location to assess damage at another location with less information. In: *PSAM 14 conference, probabilistic safety assessment and management*, Los Angeles, CA, 16–21 September 2018.
- Duckworth H and Eiber R. Assessment of pipeline integrity of Kinder-Morgan conversion of the Rancho pipeline. Report, City of Austin Texas, Austin, TX, 2004.
- Moradi R and Groth KM. Hydrogen storage and delivery: review of the state of the art technologies and risk and reliability analysis. *Int J Hydrogen Energ* 2019; 44: 12254–12269.
- Melaina MW, Antonia O and Penev M. Blending hydrogen into natural gas pipeline networks: a review of key issues. Technical report, National Renewable Energy Laboratory, Golden, CO, 2013.
- Dodds PE and Demoullin S. Conversion of the UK gas system to transport hydrogen. *Int J Hydrogen Energ* 2013; 38(18): 7189–7200.
- Wu KY and Mosleh A. Effect of temporal variability of operating parameters in corrosion modelling for natural gas pipelines subjected to uniform corrosion. *J Nat Gas Sci Eng* 2019; 69: 102930.
- PHMSA. Guidance for pipeline flow reversals, product changes, and conversion to service, 2014, <http://www.oceweb.com/PLS/2014Gas/Guide-Flo%20Rev-Prod%20Ch-Conver.pdf>
- Dann MR and Dann C. Automated matching of pipeline corrosion features from in-line inspection data. *Reliab Eng Syst Safe* 2017; 162: 40–50.
- Velázquez J, Van Der Weide J, Hernández E, et al. Statistical modelling of pitting corrosion: extrapolation of the maximum pit depth-growth. *Int J Electrochem Sci* 2014; 9(8): 4129–4143.
- Zhang S, Zhou W and Qin H. Inverse Gaussian process-based corrosion growth model for energy pipelines

- considering the sizing error in inspection data. *Corros Sci* 2013; 73: 309–320.
22. Zhang S and Zhou W. Bayesian dynamic linear model for growth of corrosion defects on energy pipelines. *Reliab Eng Syst Safe* 2014; 128: 24–31.
 23. Ossai CI, Boswell B and Davies IJ. Modelling the effects of production rates and physico-chemical parameters on pitting rate and pit depth growth of onshore oil and gas pipelines. *Corros Eng Sci Techn* 2016; 51(5): 342–351.
 24. Nuhi M, Seer TA, Al Tamimi A, et al. Reliability analysis for degradation effects of pitting corrosion in carbon steel pipes. *Procedia Engineer* 2011; 10: 1930–1935.
 25. Grall A and Fouladirad M. Maintenance decision rule with embedded online Bayesian change detection for gradually non-stationary deteriorating systems. *Proc IMechE, Part O: J Risk and Reliability* 2008; 222(3): 359–369.
 26. Arulampalam MS, Maskell S, Gordon N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE T Signal Proces* 2002; 50(2): 174–188.
 27. Rabiei E, Droguett EL and Modarres M. A prognostics approach based on the evolution of damage precursors using dynamic Bayesian networks. *Adv Mech Eng* 2016; 8(9): 1687814016666747.
 28. Bergman N. *Recursive Bayesian estimation: navigation and tracking applications*. PhD Thesis, Linköping University, Linköping, 1999.
 29. Doucet A, De Freitas N and Gordon N. An introduction to sequential Monte Carlo methods. In: Doucet A, De Freitas N and Gordon N (eds) *Sequential Monte Carlo methods in practice*. New York: Springer, 2001, pp. 3–14.
 30. Kitagawa G. A self-organizing state-space model. *J Am Stat Assoc* 1998; 93: 1203–1215.
 31. Liu J and West M. Combined parameter and state estimation in simulation-based filtering. In Doucet A, De Freitas N and Gordon N (eds) *Sequential Monte Carlo methods in practice*. New York: Springer, 2001, pp. 197–223.
 32. Doucet A and Tadić VB. Parameter estimation in general state-space models using particle methods. *Ann I Stat Math* 2003; 55(2): 409–422.
 33. Chen T, Morris J and Martin E. Particle filters for state and parameter estimation in batch processes. *J Proc Contr* 2005; 15(6): 665–673.
 34. Kelly D and Smith C. *Bayesian inference for probabilistic risk assessment: a practitioner's guidebook*. London: Springer Science & Business Media, 2011.
 35. Abdel-Hameed M. A gamma wear process. *IEEE T Reliab* 1975; 24(2): 152–153.
 36. Van Noortwijk J. A survey of the application of gamma processes in maintenance. *Reliab Eng Syst Safe* 2009; 94(1): 2–21.
 37. Frangopol DM, Kallen MJ and van Noortwijk JM. Probabilistic models for life-cycle performance of deteriorating structures: review and future directions. *Prog Struct Eng Mat* 2004; 6(4): 197–212.
 38. Walpole RE, Myers RH, Myers SL, et al. *Probability and statistics for engineers and scientists, vol. 5*. New York: Macmillan, 1993.
 39. Mustaffa Z, Gelder P, Dawotola A, et al. Reliability assessment for corroded pipelines in series considering length-scale effects. *Int J Auto Mech Eng* 2018; 15(3): 5607–5624.
 40. Pandey MD. Probabilistic models for condition assessment of oil and gas pipelines. *NDT&E Int* 1998; 31(5): 349–358.
 41. Det Norske Veritas. *Recommended practice DNV-RP-F101 corroded pipelines*, vol. 11. Hovik: Det Norske Veritas, 2004, pp. 135–138.
 42. Zhou W, Hong H and Zhang S. Impact of dependent stochastic defect growth on system reliability of corroding pipelines. *Int J Pres Ves Pip* 2012; 96: 68–77.
 43. Zhou W. System reliability of corroding pipelines. *Int J Pres Ves Pip* 2010; 87(10): 587–595.