Reliability Engineering and System Safety 214 (2021) 107740

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

A hybrid population-based degradation model for pipeline pitting corrosion

Roohollah Heidary *, Katrina M. Groth

Systems Risk and Reliability Analysis Lab (SyRRA), Center for Risk and Reliability, University of Maryland, College Park, MD, 20742, USA

ARTICLE INFO

Keywords:
Population-based degradation model
Pitting corrosion
Hierarchical Bayesian
Data fusion
Hybrid prognostics and health management

ABSTRACT

This paper presents a novel algorithm to develop a population-based pitting corrosion degradation model for piggable oil and gas pipelines. The algorithm is designed to estimate and predict the distribution of actual depth of existing pits on a pipeline segment, given two or more sets of in-line inspection data that have uncertainty in size and number of the detected pits. This algorithm eliminates the need for a defect-matching procedure for those pits that are not critical, that is required in developing defect-based pitting corrosion degradation models. A hierarchical Bayesian model based on a non-homogeneous gamma process is developed to fuse the uncertain in-line inspection data and physics of failure knowledge of pitting corrosion process. Measurement error (ME), probability of detection (POD), and probability of false call (POFC) are addressed in the developed algorithm. The application of the developed algorithm is demonstrated by implementing it on a simulated case study and the results are compared with the simulated data from a generic degradation model that is available in the literature. Results indicate that this algorithm can predict the degradation level of the pipeline with a high accuracy.

1. Introduction

Having a high confidence estimation of pipelines' degradation level plays an important role in pipeline integrity management. Estimated degradation level is the main input for time to failure or remaining useful life estimation and subsequently condition-based maintenance optimization of the pipelines. In this way, taking into account all potential failure mechanisms is necessary. Among different potential failure mechanism of pipelines, pitting corrosion is one of the main concerns because of the high rate at which pits can grow [1] and cause major failures that may impose a huge cost to the industry and environment, or unnecessary maintenance.

Two types of algorithms are present in pitting corrosion degradation modeling literature: defect-based and population-based. The defect-based algorithms, in which sequential data from individual pits are used to evaluate the growing pattern of each pit are more common in literature. However, defect-based algorithms are inefficient and resource intensive when the pit density is high. When a pipeline contains a combination of high and low density pitting segments, it is necessary to use a combination of population-based algorithms and defect-based algorithms. [2]. This motivates the development of the population-based algorithms, wherein the growing behavior of a population of pits (rather than of individual pits) is analyzed.

In this paper, we propose a novel hybrid population-based algorithm to estimate the degradation level of piggable pipelines due to pitting corrosion. We use the word hybrid to denote that the algorithm takes into account both physics of failure (POF) and inspection data. By developing this hybrid model, the disadvantages of a pure physics-based model (e.g., a lot of simplifying assumptions) or a pure data driven-based model (e.g., the need for a large amounts of data) are minimized, and their advantages (e.g., long-term damage behavior prediction of physics-based models and flexibility of data-driven models) are maximized [3].

Compared to previous literature on hybrid population-based models, our proposed model addresses temporal stochasticity of pitting corrosion process, practical cases of having more than two ILI datasets, and initiation of new pits after the last ILI. In this algorithm, we use non-homogeneous gamma process (NHGP) as the underlying stochastic process. In addition, we propose an innovative clustering approach to cluster detected pits based on their initiation times. This approach is based on comparing the probability mass function (PMF) of those pits that have been initiated before a particular inspection time and **are expected to be detected** at that inspection time and the PMF of those pits that **are detected** at that inspection time. In order to validate our algorithm, we simulate the actual maximum depth for a number of pits, as the existing pits, based on a non-homogeneous gamma process and account for probability of detection and measurement error to consider uncertainty in pit measurement and detection. Finally, we compare the distributions of the simulated and estimated maximum pit depth, the

* Corresponding author.
E-mail addresses: heidary@umd.edu (R. Heidary), kgroth@umd.edu (K.M. Groth).

https://doi.org/10.1016/j.ress.2021.107740
Received 12 December 2019; Received in revised form 1 January 2021; Accepted 29 April 2021
Available online 21 May 2021

Nomenclature

ILI	In-line inspection
KLD	Kullback–Leibler divergence
MFL	Magnetic flux leakage
NHGP	Non-homogeneous gamma process
PMF	Probability mass function
POD	Probability of detection
POF	Physics of Failure
POFC	Probability of false call
PWT	Pipe wall thickness
RMSE	Root mean squared error
RUL	Remaining useful life
SKLD	Symmetric Kullback–Leibler divergence
Actual depth	Actual depth of a pit without measurement error (mm)
Measured depth	Measured depth of a pit with measurement error (mm)
Estimated depth	Mean of estimated actual depth of a pit (mm)
a	Constant biased error (mm)
b	Proportional biased error
c	Pit initiation time index, 1 for those pits that are initiated before ILI_1, 2 for those that are initiated between ILI_1 and ILI_2, ...
d	Maximum pit depth (mm)
d_{ijc}	Actual depth of a true call i at t_j with the pit initiation time's index c (mm)
d_d	Minimum depth detection threshold of an inspection tool (mm)
D_j	Actual maximum pit depth of true calls at t_j (mm)
h	Bin index
H	Number of bins of a histogram
i	Pit index
j	ILI order index
k	Parameter of the power law model (mm)
m_j	Number of detected pits at ILI_j at t_j
m_{jc}	Number of detected pits at ILI_j that have been initiated at t_{c-1}
m'_{jc}	Number of pits that have been initiated at t_{c-1} and are expected to have been detected at ILI_j
m'_{jc}	Number of pits that have been initiated at t_{c-1} and are expected to have been detected at ILI_j
M_j	Actual number of existing pits at t_j, $M_0 = 0$
M_{jc}	Actual number of existing pits at ILI_j that have been initiated at t_{c-1}
n_h	Number of pits in bin h
r_{ij}	Reported depth of pit i at t_j (mm)
R_j	Number of reported pit at t_j
t_0	Pit initiation time (year)
t	Time (year)
t_j	Time of ILI_j (year)
t_{c-1}	Pit initiation time (year)
x_{ijc}	Actual depth of an existing pit i at t_j with the pit initiation time's index c (mm)
X_j	Vector of the actual maximum pit depth of existing pits at t_j (mm)
y	Measured maximum pit depth (mm)
y_{ijc}	Measured depth of a true call i at t_j with the pit initiation time's index c (mm)
Y_j	Vector of the measured maximum pit depth of true calls at t_j
Y_{jc}	Matrix of the measured maximum pit depth of true calls at t_j that have been initiated at t_{c-1}
Y'_{jc}	Expected measured maximum pit depth of true calls at t_j that have been initiated at t_{c-1}
y_c	Mean credible measured depth of the inspection tool (mm)
y'	Predicted value for the measured depth of a pit (mm)
n_h	Number of pits in bin h
ν	Exponent of the power law model
θ	Vector of degradation model parameters, k, ν, β
α'_{jc}	Shape parameter of the gamma process at t_j for those pits that have been initiated at t_{c-1} by using k, ν that have been estimated based on the inspection data at $ILI_1, ILI_2, \ldots, ILI_{j-1}$
α_{jc}	Shape parameter of the gamma process at t_j for those pits that have been initiated at t_{c-1} by using k, ν that have been estimated based on the inspection data at ILI_1, ILI_2, ..., ILI_j
β	Scale parameter of the gamma process
ϵ	Random scattering error (mm)
λ	Rate parameter of the HPP for pits initiation times
$\hat{\lambda}$	Point estimate of the rate parameter of the HPP for pits initiation times

number of the existing pits, and the pit initiation rate of the homogeneous Poisson process.

The input data used in this work is based on in-line inspection (ILI) reports of pipeline inspection. The ILIs are commonly based on a magnetic flux leakage (MFL) or ultrasonic testing (UT) techniques [4] and those reports contain the measured maximum depth of the detected pits after processing the raw collected data (i.e., MFL or UT signals) which have uncertainty in depth measurement (i.e., ME) and number of pits (i.e., POD and POFC) [5].

Despite significant advancement in the smart pigs technologies and the continuous improvement in their accuracy, their measurements still include of different types of uncertainty including measurement error in sizing, probability of not detecting some of the existing defects, and false call (false positive) for some defects that do not exist [6]. In order to address these types of uncertainty in pitting corrosion degradation modeling, two categories of algorithms are developed in the literature. The defect-based algorithms, in which the results of sequential ILIs of each individual pit are used to evaluate the growing pattern of that pit (are more common in literature) vs. population-based algorithms, in which the growing behavior of a population of pits (and not individual pits) is analyzed.

Existing defect-based algorithms require matching the results of sequential ILIs based on the location of the pits. Assuming the matched ILI datasets are available, Maes and Dann [7] proposed a defect-based hierarchical Bayesian (HB) model to estimate the degradation level of

oil and gas pipelines due to localized corrosion. To address the temporal stochasticity of the pitting corrosion process, they used a gamma process as the underlying stochastic process of their model. Zhang and Zue [8] validated Maes model by applying that model on four real pitting corrosion data sets of sixty two pits on an 80 km natural gas pipeline in Alberta, Canada. They also extended Maes model by assuming inverse-Gaussian [9] and Bayesian dynamic linear model [10] as the underlying stochastic process. In all those models, the underlying assumption is that the operational conditions of the pipelines remains the same for the operating life of the pipelines. Heidary and Groth [11] developed a defect-based pitting corrosion degradation model to cover the case when the operational conditions of a pipeline change over time.

Despite the fact that defect-based algorithms are more common in the oil and gas industry, they are less suitable when the pit density is high, because the matching procedure is time consuming and prone to error [2,12]. A population-based algorithm can be applied as soon as new ILI data become available without further matching procedure, and the results of this algorithm can be used to evaluate the criticality of a pipeline segments to decide about necessity of extra effort of a local corrosion growth analysis using matched features [2]. In other words, in case of a high pit density, it is not cost-effective to match the ILI data of all pits at the first place, and it is more time and cost effective to use a population-based algorithm to recognize the critical segments first, and then match the ILI data of just those segments.

To eliminate the matching procedure step in the case of having high density pits, for those pits that are not critical (e.g., as experienced in upstream and subsea pipelines [2]), a few population-based methodologies have been proposed in the literature. The most recent one was proposed by Dann and Maes [2]. They used KL divergence method to estimate the hyper-parameters of the degradation model and used a homogeneous gamma process as the underlying stochastic process. They applied their model in a case study with two sets of ILI data and did not consider the initiation of new defects after the most recent inspection. Lu [13] developed a population based pitting corrosion model, for nuclear power plants, using gamma process based on a "repair-on-detection" strategy, which at least for pipelines, is not applicable.

More details about this algorithm, the simulated case study, and the results are presented in this paper as follows. Section 2 is dedicated to the requirements and assumptions. In this section theoretical background about gamma process and HB model and the reasons that they are suitable for this degradation modeling are explained. In addition, the assumptions behind this algorithm are articulated. In Section 3 the developed algorithm is presented in detail and in Section 4, it is demonstrated by a simulated case study. The conclusion is presented in Section 5.

2. Assumptions and methods

2.1. Assumptions

This algorithm is developed based on the following assumptions:

- Operational conditions (i.e., probability density function of temperature, pressure, flow rate, etc.) of the pipeline do not change over time and all pits are under the same operational conditions at each time [7,8,14].
- Pits are not interacting each other [2].
- The number of new initiated pits between each two ILI follows a homogeneous Poisson process (HPP) [15,16]. Which means pits' initiation times follow the corresponding uniform distribution and the time between pits' initiation follow the corresponding exponential distribution.
- The detected pits are not mitigated by the maintenance activities (no maintenance).
- n ($n \geq 2$) sets of population-based ILI datasets are available. These datasets are reported by the ILI service companies that had performed the n ILI operations at $t_1, t_2, \ldots, t_n$. And those reported ILI datasets include both true calls and false calls of the uncertain measurement (y_i) of the maximum depth of each pit of the population.
- The measurement model of the inspection tools is available and follows Eq. (1) [17].

$$y_i = a + b * d_i + N(0, \epsilon) \tag{1}$$

where a and b are the biased errors and ϵ is the random scattering error of the inspection tool which are assumed to be given by the ILI service companies. d_i is the actual depth and y_i is the measured depth of pit i, and N stands for a normal distribution. This equation is used to estimate the actual depth of a pit given its measured depth.

- The detection threshold and the credible pit depth of the inspection tools are given by the ILI service companies. They are used to calculate the POD (Eq. (2)) of each pit given its actual depth and the probability of false call (Eq. (3)) corresponding to each measured depth.

$$POD(Pit_i \mid d_i, d_d) = 1 - exp(-d_i/d_d) \tag{2}$$

where d_d represents the minimum detection threshold of the inspection tool [8,12].

$$POFC(Measurement_i \mid y_i, y_c) = exp(-y_i/y_c) \tag{3}$$

where y_i represents the i_{th} measured depth and y_c represents the mean credible measured depth of the inspection tool [2].

2.2. Homogeneous Poisson Process (HPP)

It is well accepted that pitting corrosion comprises two main processes: pit initiation and stable pit growth. The pit initiation process can be a consequence of the breakdown of the passive layer (a protective layer that can be created by the corrosion products) caused by random fluctuations in local conditions which takes some time, usually called induction (nucleation or initiation) period [18]. The pit initiation time varies depending on the corrosive environment and the material properties. In some experimental works, it has been confirmed that the distribution of the pits' initiation times follows an exponential distribution, and therefore pit initiation can be modeled by using a homogeneous Poisson process (HPP) [15,16]. In some other experiments, it has been observed that the pits' initiation times are not distributed uniformly and pits initiation rate is a decreasing function of time for long duration corrosion tests [19,20]. To model this behavior, non-homogeneous Poisson process (NHPP) has been used to model the stochasticity in pits initiation times [21,22]. In this paper, we assumed that pits initiation times follow an exponential distribution and the number of initiated pits at each time interval follows the corresponding HPP with a rate parameter λ.

2.3. Non-homogeneous gamma process

The probabilistic pitting corrosion models can be categorized as random-variable based and stochastic-process based models [8]. The main difference between these categories is that, the latter deals with the temporal variability of the pitting corrosion process, while the former does not capture it [8]. Among different stochastic-process based models, gamma process is more appropriate to model pitting corrosion process [23]. Gamma process has been used widely to model degradation processes such as wear, fatigue, and corrosion, which involve monotonically accumulating damage over time in a sequence of tiny increments [24,25]. In this paper, a non-homogeneous gamma process is used to take into account the temporal uncertainty (i.e., time

dependent variation) in pitting corrosion process. Another reason for using non-homogeneous (vs. homogeneous) gamma process is that, it makes it possible to consider physics of failure knowledge about pitting corrosion process, that is embedded in this well-accepted assumption, that growing behavior of maximum depth of pits follows a power function with a positive power less than one [26–30].

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

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

where Ga represents probability density function (PDF) of a gamma distribution. A random quantity (x) (in this study a pit's maximum depth) has a gamma distribution with a shape parameter $\alpha > 0$ and a rate parameter $\beta > 0$, if its PDF is given by:

$$f_{X(t)}(x) = Ga(x; \alpha, \beta) = \frac{\beta^{\alpha(t)}}{\Gamma(\alpha(t))} x^{\alpha(t)-1} exp(-\beta x) \tag{4}$$

where $\Gamma(.)$ denotes the gamma function. Eqs. (5) and (6) show the expectation and variance of the gamma process respectively.

$$E(X(t)) = \frac{\alpha(t)}{\beta} \tag{5}$$

$$Var(X(t)) = \frac{\alpha(t)}{\beta^2} \tag{6}$$

According to Eq. (5), the scale parameter (β) of a gamma process is constant and time independent. Hence, the shape parameter, $\alpha(t)$, of a gamma process can address the temporal trend of the average of a random variable that follows a gamma process. Hence, different degradation rate behavior(i.e., increasing, decreasing, or constant) can be modeled by selecting an appropriate form for the shape parameter of a gamma process [7].

In the case of pitting corrosion, it is well-accepted that the mean value of the maximum depth of an active pit follows a power law function with a positive exponent less than one [1,26,27]. It has been found that pitting corrosion growth in stainless, mild steels and aluminum alloys follows this form of function [31,32]. Therefore, in this case, the shape parameter of the underlying gamma process follows a power function according to Eq. (7) and the corresponding gamma process is a non-homogeneous (non-stationary) gamma process, i.e., $\nu \neq 1$ [25].

$$\alpha(t) = k(t - t_0)^{\nu} \tag{7}$$

where k and ν represent the parameters of the pitting corrosion degradation model and t_0 represents the pit initiation time. Accordingly, the distribution of the actual depth of a pit population at t follows a gamma distribution given in Eq. (8).

$$f_{X(t)}(x) = Ga(\alpha, \beta) = Ga(k(t - t_0)^{\nu}, \beta) \tag{8}$$

2.4. Hierarchical Bayesian (HB) modeling

In this study, we use a HB model to estimate the pitting corrosion degradation model parameters (k, ν, β), and the actual depth of the existing pits. HB modeling is an appropriate method to make scientific inference about hyper-parameters of the distribution of an unknown of a population, based on the observations of many individuals. It is called "hierarchical", because it uses hierarchical or multistage prior distributions [33]. This method has been used in the literature to develop defect-based algorithms to model different types of uncertainty (temporal, spacial, epistemic [34], measurement) related to corrosion growth in the pipelines [7–9].

The general population-based HB model that is developed in this study is depicted in Fig. 1. In this figure, the temporal plate indicates the time of the ILI_j. Based on these assumptions that operational conditions do not change over time and all pits are under the same

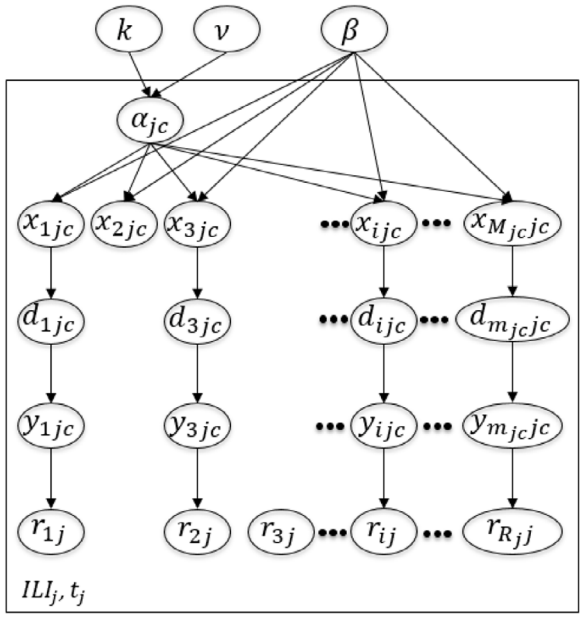

Fig. 1. General HB model for each category of pits at each ILI time based on their initiation time.

operational condition, a same gamma process is used for all pits for the entire life of the pipeline. Therefore, the degradation model parameters (k, ν, β) of that gamma process are outside of the temporal plate. However, the shape parameter of the gamma process, α, is time dependent (because of the change in t_j and t_{c-1} in Eq. (7)). Therefore, α_{jc} is inside of the temporal plate and it varies at each inspection time for each category of pits based on their initiation time. Primarily, it has been assumed that for those pits that have been initiated before the first ILI, the initiation time is $t_{c-1} = t_0 = 0$, and for those that have been initiated between each two ILIs, the initiation time of all of them is exactly after the most recent ILI. Finally at the prediction phase, the pits initiation times are distributed as an HPP. Hence, after each ILI, it is necessary to figure out which detected pits are new and which pits have been initiated previously that may or may not be detected in the previous ILIs. This clustering step will be discussed in more details in Section 3.3.

In Fig. 1, x_{ijc} indicates the actual depth of an existing pit i at t_j with the initiation time index c, d_{ijc} indicates the actual depth of a detected pit i at t_j with the pit initiation time index c, y_{ijc} indicates the measured depth of a detected pit i at t_j with the initiation time index c, and r_{ij} indicates the measured depth of a reported pit i at t_j. In addition, R_j indicates the number of reported pits at ILI_j including both false and true calls, m_{jc} indicates the number of pits that have been initiated at t_{c-1} and are truly detected (i.e., true calls) at ILI_j, M_{jc} indicates the number of existing pits at ILI_j that have been initiated at t_{c-1}. As it is shown in this figure, some existing pit (e.g., x_{2jc}) might be missed to be detected by the ILI tool at t_j. In addition, some reported pits might be false calls (e.g., r_{3j}).

Since at each ILI there are some false calls and also some existing pits that are not detected, the original HB model is modified to the one that is shown in Fig. 2. The process of filtering true calls from the reported pits is explained in Section 3.

This HB model is used to estimate the degradation model parameters and actual depth of the detected pits by using the Bayes rule according to Eq. (9).

$$\begin{aligned} Pr(d_{ijc}, k, \nu, \beta \mid y_{ijc}) &= Pr(d_{ijc}, \theta \mid y_{ijc}) \\ &\propto Pr(y_{ijc} \mid \theta, d_{ijc}) \times Pr(\theta, d_{ijc}) \end{aligned} \tag{9}$$

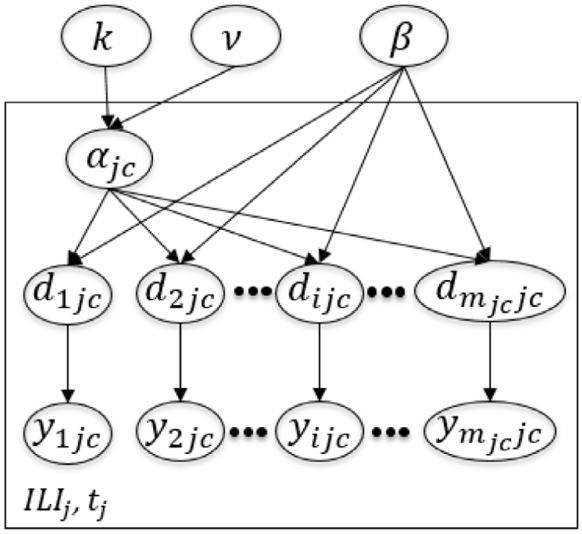

Fig. 2. Modified general HB model for each category of pits at each ILI time based on their initiation time.

3. Developed algorithm

3.1. General structure of the developed algorithm

Fig. 4 depicts the general structure of the algorithm which includes three phases. According to this figure, in phase I, the first ILI dataset is used to estimate the degradation model parameters, the number of existing pits at t_1, and the PMF of the actual depth of those pits at that time. In phase II, the other ILI datasets are used to update the degradation model parameters, and to estimate the number of existing pits and the PMF of their actual depth at $t_2, \ldots, t_n$. The main difference between phase I and phase II is in a clustering step. Since it is assumed that all existing pits at t_1, has been initiated at $t = 0$, hence all of them follow a same degradation model. However, at other inspections times and the prediction time, there are clusters of pits with different initiation times. Therefore, we use a clustering algorithm to categorize the detected pits at each inspection time based on their initiation time to be used on their corresponding Bayesian models. In phase III, the rate parameter of the assumed HPP and the number of pits that are expected to be initiated between the last ILI and the prediction time are estimated. Finally, the PMF of the actual depth of the existing pits at prediction time (t_{n+1}) is estimated. The details of these three phases are given in the following subsections.

3.2. Phase I: Using the first ILI dataset

In phase I, the first set of reported ILI data is used to find the first estimate of the degradation model parameters (k, ν, β) of the NHGP, the number of the existing pits, and the PMF of their actual depth at t_1, by performing the following steps:

- Step I-1: Filter the true calls from the reported pits:
  For each reported measured depth, generate a random uniform number between zero and one. If the probability of false call of that measured pit depth (use Eq. (3)) is higher than that random number, remove that measurement from the dataset and if not, keep it.
- Step I-2: Estimate the degradation model parameters and the actual depth of the detected pits:
  Use the measured depth of the m_{11} true calls at ILI_1 and the prior values ($p_{1,2,3}$, $q_{1,2,3}$) in the corresponding HB (Fig. 3) to obtain the first estimate of the degradation model parameters and the actual depth of the true calls at t_1 by using the Bayes rule (Eq. (9)). Non informative prior distributions (i.e., uniform distributions) can be selected for prior distributions of the hyper-parameters in this Bayesian model (Fig. 3); $p_{1,2,3} > 0$ are the lower bounds of the prior uniform distributions and $q_{1,2,3} > 0$ are the upper bounds. The only prior information is that the upper bound of ν, (q_2), is equal to one based on physics of failure knowledge of internal pitting corrosion process.

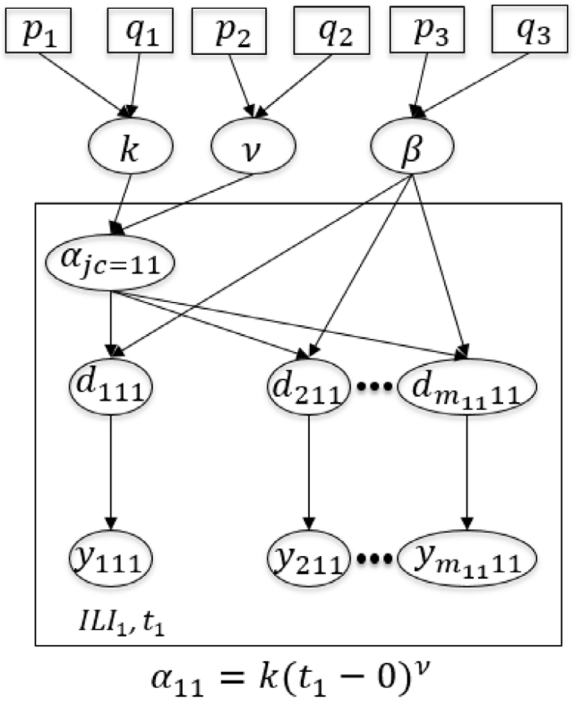

Fig. 3. Modified HB model for ILI_1.

- Step I-3: Estimate the number of existing pits at t_1 (M_1) given the number of true calls (m_{11}):
  Discretize the PMF of d_{i11} to H number of bins and estimate the number of existing pits at each bin by using Eq. (10) [5,35].

$$n_h(PMF(X_1)) = n_h(PMF(D_1))/POD_h \quad h = 1, 2, \cdots, H \tag{10}$$

  where n_h is the number of pits in bin h and POD_h is the corresponding POD which can be calculated by using the mean value of bin h in Eq. (2). The total number of existing pits at t_1 can be estimated by summing the frequency of all bins of the PMF of x_{i11} by using Eq. (11).

$$M_1 = \sum_{h=1}^{H} n_h(PMF(x_{i11})) \tag{11}$$

- Step I-4: Estimate the PMF of the actual depth of the existing pits:
  For those bins that the frequency in PMF of x_{i11} is higher than the frequency of PMF of d_{i11}, generate $n_h(PMF(x_{i11})) - n_h(PMF(d_{i11}))$ number of random uniform values between the start and end depth of that bin.

3.3. Phase II: Incorporating additional datasets

In this phase, as mentioned previously, a clustering step is required to categorize the detected pits based on their initiation time. Generally speaking, (for ILI_2), first we filter the true calls from the reported measurements at t_2. Then we estimate how many of the true call at

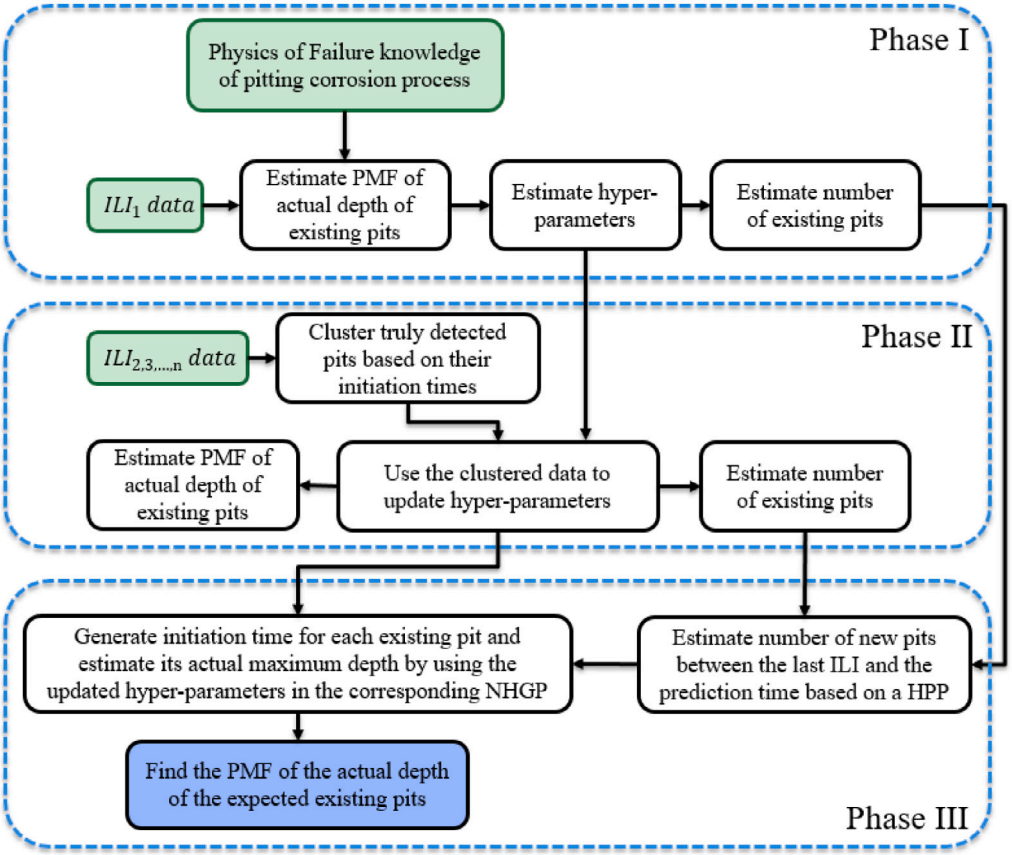

Fig. 4. General structure of the developed algorithm.

t_1, we expect to have been detected at t_2. We subtract this expected number from the true calls at t_2 to obtain the number of pits that have been initiated between t_1 and t_2. The next step is to assign these estimated number of pits to their corresponding clusters (i.e. initiated pits before t_1 or initiated pits between t_1 and t_2). This phase is explained in detail for ILI_2 as following and the same approach is applicable for the next ILIs.

- Step II-1: Filter the true calls from the reported pits at ILI_2:
  For each reported measured depth, generate a random uniform number between zero and one. If the probability of false call of that measurement (use Eq. (3)) is higher than that random number, remove that measurement from the dataset and if not, keep it.
- Step II-2: Estimate the number of pits of each cluster that are expected to have been detected at t_2:
  In this step, we want to estimate how many of those pits that have been initiated before ILI_1 (M_1) are expected to have been detected at ILI_2 (m'_{21}), given the first estimation of the degradation model parameters that are obtained in Phase I, and the number of true calls at t_2 (m_2). Then, we subtract that number from the number of true calls at ILI_2 (m_2) to find the expected number of those pits that have been initiated between ILI_1 and ILI_2, and are expected to have been detected at ILI_2. The general idea of the this step is shown in Fig. 5. According to this figure, at each inspection time, there are a number of true calls and undetected pits. Since in a gamma process the rate parameter is time independent, the estimated rate parameter at t_1, can be used at t_2. However, for the shape parameter, which is time dependent, the estimated k and ν in phase I, in conjunction with t_2 are used in Eq. (7) to find $\alpha'_{21} = k(t_2 - 0)^{\nu}$. By using those rate and shape parameters, the expected PDF of the actual maximum pit depth of the pit population can be estimated by using Eq. (8) ($g'_{21}(x) = Ga(\alpha'_{21}, \beta)$). Then m'_{21} can be obtained by integrating the expected PDF times the corresponding POD on all possible values of maximum pit depth (x), multiplied by the number of pits that have been initiated before t_1 (M_1) (Eq. (12)). Based on the assumption of NHGP, the expected PDF of the actual depth of the pits that have been initiated before ILI_1, follows a gamma distribution ($g'_{21}(x)$) at t_2 with a shape parameter equal to $\alpha'_{21} = k(t_2 - 0)^{\nu}$ and a scale parameter equal to β (Eq. (8)).

$$m'_{21} = M_1 \int g'_{21}(x) POD(x) d(x) \tag{12}$$

And m'_{22} can be calculated by subtracting m'_{21} from the true calls (m_2) at t_2 by using Eq. (13).

$$m'_{22} = m_2 - m'_{21} \tag{13}$$

- Step II-3: Assign each true call to a cluster:
  At this step, m'_{21} number of true calls at t_2 are assigned to one cluster and m'_{22} of them are assigned to another cluster. To do so:
  - Generate m'_{21} random number from $g'_{21}(x)$ and use Eq. (1) to find the PMF of Y'_{21}.
  - Generate m'_{22} random number from $g'_{22}(x) = Ga(\alpha'_{21}, \beta) = Ga(k(t_2 - t_1)^{\nu}, \beta)$ and use Eq. (1) to find the PMF of Y'_{22}.
  - Discretize PMF of Y_2 and PMF of Y'_{21} with the same bins intervals.
  - Use the following pseudo code to assign each true call to a cluster:
    Divide the pits' depth domain to H number of bins.
    Set the frequency of each bin of PMF(Y_{21}) equal to zero

Set the frequency of each bin of PMF(Y_{22}) equal to zero
For $i = 1$ to m_2:
BI_i = corresponding bin to pit i (find out each pit belongs to which bin).
If PMF($Y_{21}[BI_i]$).freq < PMF($Y'_{21}[BI_i]$).freq:
$ClusterIndex_i = 1$.
PMF($Y_{21}[BI_i]$).freq = PMF($Y_{21}[BI_i]$).freq + 1.
else:
$ClusterIndex_i = 2$.
PMF($Y_{22}[BI_i]$).freq = PMF($Y_{22}[BI_i]$).freq + 1.
End

In this pseudo code, BI_i stands for the bin index of pit i and .freq stands for the frequency of bin BI_i. The idea behind this pseudo code is that, for each pit we find the corresponding bin. Then, if the frequency of that bin in PMF(Y_{21}) is less that the frequency of the same bin in PMF(Y'_{21}), we assign that pit to PMF(Y_{21}), and if not, we assign that to PMF(Y_{22})

- Step II-4: Update the degradation model parameters and estimate the actual depth of the detected pits at ILI_2:
  Use the clustered measurement data in ILI_2 in the HB model that is shown in Fig. 6 to update the estimation of the degradation model parameters and also to estimate the actual depth of the true calls. It is worth noting that in this HB model, the prior values for the degradation model parameters are their estimated values in Phase I.
- Step II-5: Estimate the number of actual pits (M_2) given the number of true calls (m_2) at ILI_2:
  Follow the same approach that has been used in Phase I to estimate M_2 according to Eq. (14).

$$M_2 = \sum_{h=1}^{H} n_h(PMF(X_2)) = \sum_{h=1}^{H} n_h(PMF(D_2))/POD_h \tag{14}$$

- Step II-6: Estimate the PMF of the actual depth of the existing pits at ILI_2:
  For those bins that the frequency of PMF(X_2) is higher than the frequency of PMF of D_2, generate $n_h(PMF(X_2)) - n_h(PMF(D_2))$ number of random uniform values between the start and end depth of that bin.

For $ILI_j, j > 2$, the same steps should be followed. To shorten the paper, the algorithm is not demonstrated for those $ILIs$ in details and the related equations are generalized as follows:

$$\alpha'_{jc} = k(t_j - t_{c-1})^{\nu} \tag{15}$$

$$m'_{jc} = (M_c - M_{c-1}) \int g'_{jc}(x) POD(x) d(x), c = 1, \ldots, j-1 \tag{16}$$

$$m'_{jj} = m_j - \sum_{c=1}^{j-1} m'_{jc} \tag{17}$$

$$M_j = \sum_{h=1}^{H} n_h(PMF(X_j)) = \sum_{h=1}^{H} n_h(PMF(D_j))/POD_h \tag{18}$$

3.4. Phase III: Prognostics

In phase III, the rate parameter of the HPP is estimated and the PMF of the actual depth of the existing pits at any time after the last ILI can be predicted. This phase includes the following steps:

- Step III-1: Estimate the rate of HPP:
  By having the estimated number of existing pits at each ILI ($M_1, M_2 - M_1, M_3 - M_2, \ldots$) a point estimate of the rate parameter of the HPP can be estimated by using the maximum likelihood estimation:

$$\hat{\lambda} = \frac{\sum_{j=1}^{n}(M_j - M_{j-1})}{\sum_{j=1}^{n}(t_j - t_{j-1})} = \frac{M_j}{t_j} \tag{19}$$

- Step III-2: Estimate the number of new pits between the last ILI and the prediction time (t_{n+1}):
  The mean value of the number of initiated pits between the last ILI and t_{n+1} is calculated by using Eq. (20), which is the mean value of a Poisson distribution.

$$M_{n+1} - M_n = \hat{\lambda} \times (t_{n+1} - t_n) \tag{20}$$

- Step III-3: Estimate the PMF of the actual depth of the existing pits at t_{n+1}:
  The final step is to predict the PMF of the existing pits at t_{n+1}. To do so, for each cluster of pits, pits initiation times are randomly generated from the corresponding uniform distribution (since the number of pits follows a HPP, the pit initiation times follows a uniform distribution and the initiation time intervals follow an exponential distribution). The lower and upper bound of each uniform distribution is equal to the start and the end point of each inspection interval and the number of initiation times at each time interval is equal to the numbers that are estimated previously (i.e., $M_1, M_2, \ldots, M_{n+1}$). Those generated initiation times in conjunction with the updated degradation model parameters are used in the corresponding NHGP for each cluster according to Eq. (21), to estimate the PMF of the existing pits at t_{n+1}.

$$x_{i(n+1)c} = random.gamma(k[t_{n+1} - random.uniform(t_c, t_{c-1})]^{\nu}, \beta), \quad i = 1, \ldots, M_{n+1} \tag{21}$$

4. Demonstration of the developed algorithm

In this section, the developed algorithm is demonstrated in a simulated case study. Consider a pipeline that has been inspected three times ($n = 3$) by ILI, after $t_1 = 30, t_2 = 37, t_3 = 42$ years of operation. ILI datasets (i.e., reported number of pits and their measured depth (a combination of false calls and true calls)) for a segment of the pipeline are available. The ILI datasets are not matched because of the high density of the existing pits. The goal is to estimate the number of the new pits that are expected to be initiated between the time of the last ILI and 50 years (t_4, prediction time) of operation and then using this result, estimating the distribution of the actual depth of the existing pits at prediction time to be used in pipeline reliability estimation and consequently condition-based maintenance optimization.

4.1. Simulated case study data

Since the real field ILI data for a population of pits are not available in the literature, we simulate three sets of ILI data by using a generic internal pitting corrosion degradation model that has been developed by Ossai et al. [14] to demonstrate the application of the developed algorithm and validate the results. Ossai's model was developed by using ten years (from 1999 to 2008) measurement of pit depth (using UT) and operating parameters for sixty X52 non-piggable oil and gas pipelines in Nigeria. To the best of authors' knowledge, Ossai's model is the most comprehensive generic population based model that has been developed based on field data (rather than experimental data). This model correlates eleven operational parameters (Table 1) with the average of the maximum depth of a population of pits.

The simulated data are generated by using the parameters that are given in Table 2. In this table the degradation model parameters are based on Ossai's model [14]. The characteristics of the sizing error of inspection tools (a, b, ϵ in Eq. (1)) are the characteristics of an UT tool that has been used at 2004 in Alberta, Canada [8]). The HPP rate is selected in a way to have the same order of magnitude of number of pits as is given in [2], and d_d and y_c are selected according to [12] and [2], respectively.

Given the assumed values in Table 2, following approach is used to generate simulated ILI data. The number of initiated pits at each

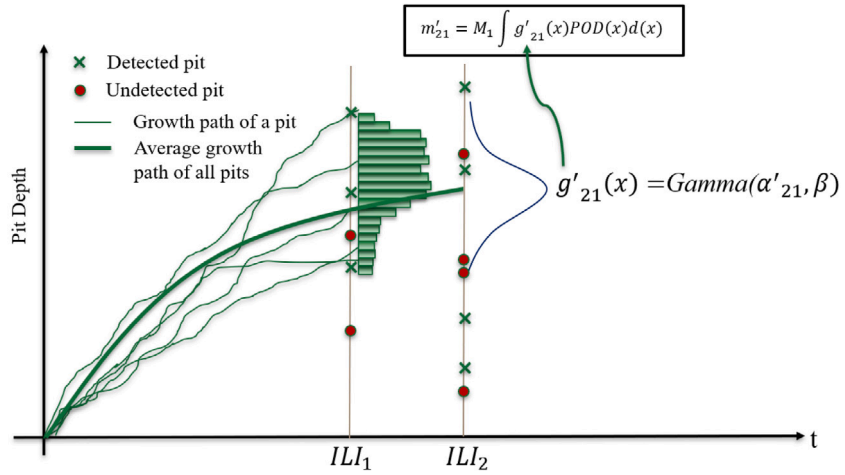

Fig. 5. Estimating number of the pits that were initiated before ILI_1 that are expected to have been detected at ILI_2.

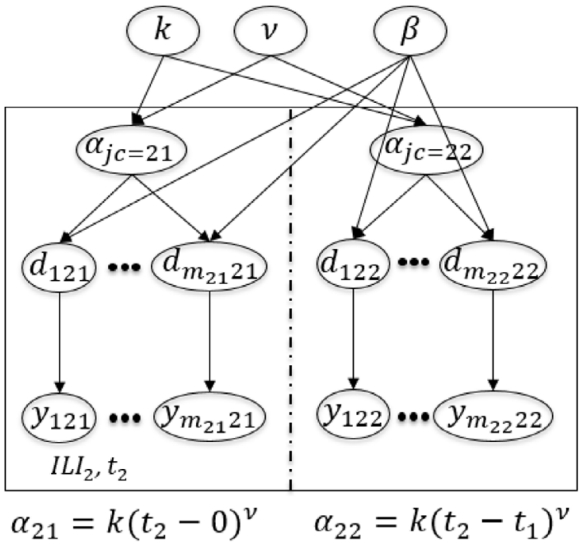

Fig. 6. Modified HB model for ILI_2.

Table 1
Operational parameters considered in the underlying generic model for data simulation [14].

Parameters	Units	Description
T	C	Temperature
Pc	MPa	CO_2 partial pressure
pH	–	pH
S	Mg L^{-1}	Sulfate ion
C	Mg L^{-1}	Chloride ion
W	–	Water cut
r	Pa	Wall shear stress
Gs	m^3 day^{-1}	Gas production rate
OL	m^3 day^{-1}	Oil production rate
Wt	m^3 day^{-1}	Water production rate
Pt	MPa	Operating pressure

time interval is calculated by generating a random number from the corresponding Poisson distribution with the rate parameter equal to $\lambda\Delta t$ (Table 3). The actual depth of each initiated pit, at each evaluation time (inspection times and prediction time), is calculated by generating a random number from the corresponding gamma distribution. The shape parameter of the gamma distribution corresponding to each pit is equal to $k(t_j - t_0)^{\nu}$. Where t_j is the evaluation time, and t_0 is the initiation time of that pit which is a uniformly distributed (i.e., HPP)

Table 2
Assumed values for input data simulation.

Param.	Value	Param.	Value
k (mm)	0.12	a (% PWT)	2.04
ν	0.771	b	0.97
β	3.5	ϵ (% PWT)	5.97
λ	80	d_d (% PWT)	10
PWT (mm)	8.41	y_c (% PWT)	20

Table 3
Number of existing and reported pits at each time interval.

Time (yr)	Exist. pits No.	Repor. pits No.
t_1: 30	2374	2516
t_2: 37	2955	3484
t_3: 42	3336	3781
t_4: 50	3967	–

random number between the start and end point of the time interval in which that specific pit is initiated in. The scale parameter of the gamma process is β. In order to consider POD in simulated data, a uniform random number is generated for each initiated pit and if that random number is less than the POD of that pit (Eq. (2)), it is considered as a detected pit, otherwise it is considered as an undetected one (hit-miss POD model [17]). For each detected pit, the measurement error is added by using Eq. (1). At this stage the PMF of the detected pits at each evaluation time is available. In order to add the false call error to the simulated dataset, the PMF of the detected pits is discretized to a number of bins. Then the frequency of each bin is divided by (1-POFC) of the mean value of that bin (using Eq. (3)) to find the frequency of the corresponding bin in the PMF of reported pits (r_{ij} in Fig. 1). For those bins that the frequency of the PMF of the reported pits is higher than the frequency of the PMF of the detected pits, uniformly distributed random measurements, between the lower and upper bound of each bin, are added to the reported dataset as false calls. The number of those false calls for each bin is equal to the difference between the frequency of the PMF of the reported pits and the frequency of the PMF of the detected pits. The PMF of the reported pits are assumed as the reported data from the ILI service companies.

4.2. Phases of the algorithm for the simulated case study

4.2.1. Phase I for the simulated case study

- Filter the true calls from the reported pits of ILI_1:
  Following the procedure that is explained in Step I-1, the PMF of the true calls is extracted from the PMF of the reported

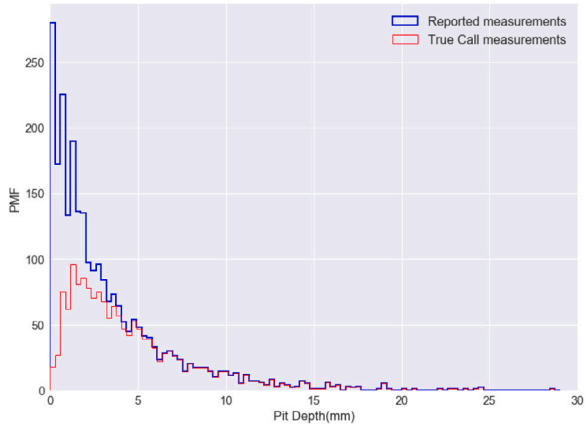

Fig. 7. PMF of the reported and true call measurements at ILI_1.

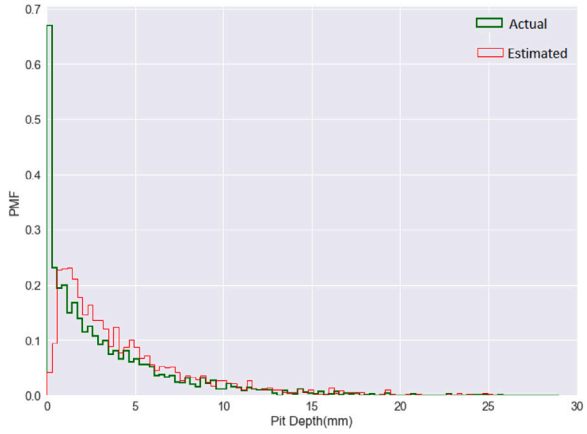

Fig. 8. PMF of actual and estimated pits' depths at $t_1(X_1)$.

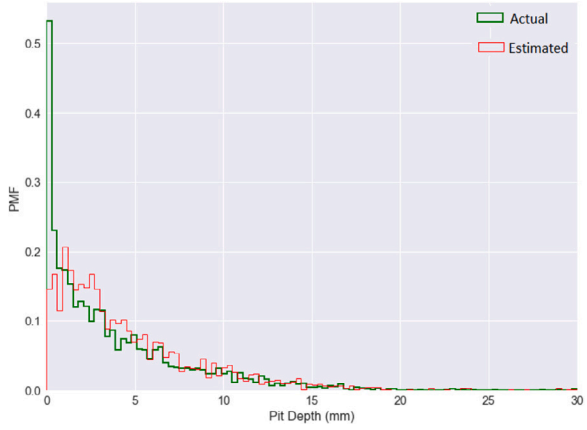

Fig. 9. PMF of actual and estimated pits' depths at $t_2(X_2)$.

measurements as shown in Fig. 7. According to this figure, the number of false measurement for pits with smaller depths is higher than the one for the deeper ones. Hence, the number of the filtered measurements is higher for the smaller pits.

- Estimate d_{i11} and a first estimate for the degradation model parameters:
  The marginal distribution of the actual depth of the true calls (d_{i11}) and the degradation model parameters are obtained by using HB model in Fig. 3. The estimated degradation model parameters with 95% confidence level are given in Table 4.
- Estimate M_1 and PMF of the existing pits at t_1:
  By discretizing the PMF of d_{i11} to $H = 100$ bins, and using Eqs. (10) and (11), the estimated number of existing pits at t_1 is obtained as $M_1 = 1816$. Then the PMF of the exiting pits at t_1 is estimated by following the procedure that is explained in Section 3.2. Fig. 8 shows a visual comparison between the simulated and estimated PMF of the existing pits at t_1. This figure shows that for the first ILI dataset, the performance of the developed algorithm is visually acceptable. A quantified comparison between these two PMFs is given in the Discussion section.

Table 4
Estimated degradation model parameters at t_1.

	Mean	SD.Dev	2.5%Qt.	97.5%Qt.
k	0.51	0.29	0.11	1.04
ν	0.37	0.20	0.11	0.75
β	3.07	0.13	2.83	3.33

Table 5
Updated degradation model parameters at t_2.

	Mean	SD.Dev	2.5%Qt.	97.5%Qt.
k	0.13	0.01	0.11	0.16
ν	0.72	0.03	0.66	0.76
β	3.02	0.08	2.88	3.19

4.2.2. *Phase II for the simulated case study*

- Estimate $m\prime_{21}$ and $m\prime_{22}$:
  The first step in phase II, is to filter the true calls from the reported pits by following the procedure that explained in Step II-1. Then it is desirable to estimate the number of pits that have been initiated before t_1 and are expected to have been detected at t_2 ($m\prime_{21}$). In this way $\alpha\prime_{21}$ is equal to 1.97 by using the estimated k and ν in Phase I, $t_j = t_2$, and $t_{c-1} = t_0 = 0$ in Eq. (15). Consequently, $m\prime_{21}$ is calculated as 1708 by using this shape parameter and the scale parameter that is estimated at t_1 in Eq. (12). Accordingly, based on Eq. (13), $m\prime_{22} = m_2 - m\prime_{21} = 528$. By having $\alpha\prime_{21}$, β, $m\prime_{21}$, $m\prime_{22}$, and following the procedure that is given in Step II-3, the PMF of $Y\prime_{21}$ and $Y\prime_{22}$ are obtained. Then, by using the given pseudo code, the detected pits at t_2 are clustered.
- Update the degradation model parameters and estimate the actual depth of the detected pits at ILI_2:
  The next step in Phase II is to use the clustered measurements in the HB model that is shown in Fig. 6. It is worth noting that the posterior distributions at Phase I are the prior distributions of this HB model. The updated degradation model parameters are shown in Table 5.
- Estimate the number of existing pits (M_2) and the PMF of their actual depth at t_2:
  By following Step II-5 and II-6, the number of existing pits at t_2 is obtained as 2541. The PMF of the simulated and estimated actual depth of the existing pits at t_2 are depicted in Fig. 9.

In the interest of brevity, the steps are not discussed in detail for ILI_3. The updated degradation model parameters are given in Table 6 and Fig. 10 and the other results are presented briefly as following.

$$\alpha'_{31} = k(t_3 - t_0)^{\nu} = 1.93 \tag{22}$$

$$\alpha'_{32} = k(t_3 - t_1)^{\nu} = 0.78 \tag{23}$$

$$\alpha'_{33} = k(t_3 - t_2)^{\nu} = 0.41 \tag{24}$$

$$m'_{31} = (M_1) \int g'_{31}(x) POD(x) d(x) = 1704 \tag{25}$$

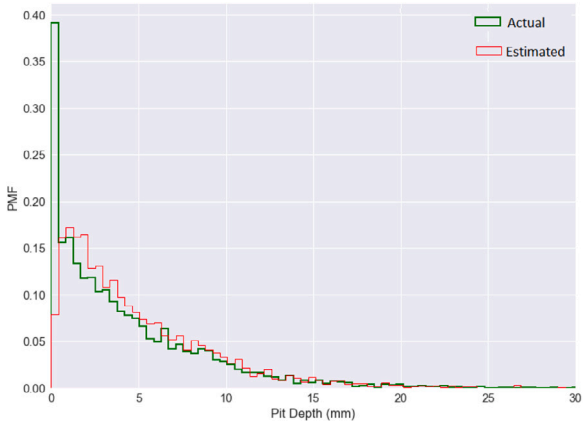

Fig. 10. PMF of actual and estimated pits' depths at $t_3(X_3)$.

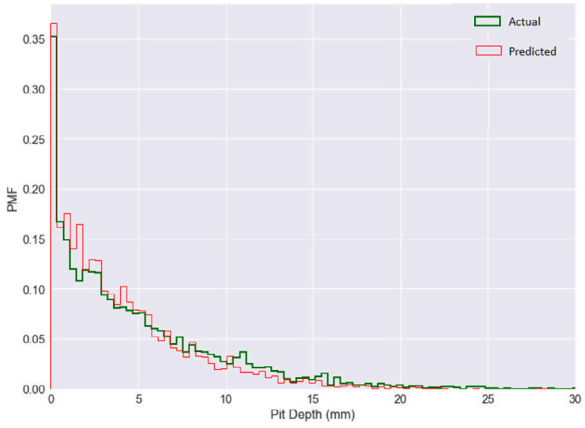

Fig. 11. PMF of actual and predicted pits' depths at $t_4(X_4)$.

$$m'_{32} = (M_2 - M_1) \int g'_{32}(x) POD(x) d(x) = 425 \tag{26}$$

$$m'_{33} = m_3 - \sum_{c=1}^{2} m'_{3c} = 502 \tag{27}$$

$$M_3 = \sum_{h=1}^{H} n_h(PMF(X_3)) = \sum_{h=1}^{H} n_h(PMF(D_3))/POD_h = 2870 \tag{28}$$

4.2.3. Prognostics phase for the simulated case study

Having those three sets of ILI data at 30, 37, and 42 years of pipeline operation, the goal is to estimate the actual depth of the existing pits at $t_4 = 50$. To do so, by using maximum likelihood estimation (Eq. (19)), the rate parameter of the assumed underlying HPP (λ) is obtained as 68. Consequently the expected number of new pits is equal to $M_4 = \lambda(t_4 - t_3) = 68 \times (50-42) = 544$. By using the updated degradation model parameters at t_3 (Table 6) and the number of initiated pits at each time interval, Eq. (21) is used to find the actual depth of the existing pits at t_4. The PMF of simulated and estimated actual depth of existing pits at t_4 are shown in Fig. 11

4.2.4. Discussion

For the sake of comparison, the mean value of the degradation model parameters and the shape parameters of the gamma process at evaluation times are given in Table 7. According to this table, although the estimated k and ν, are not so close to the assumed ones, the absolute error of the estimated β and shape parameters (α_{jc}) are less than 10% that indicates a good prediction performance.

Table 6
Updated degradation model parameters at t_3.

	Mean	SD.Dev.	2.5%Qt.	97.5%Qt.
k	0.16	0.001	0.158	0.162
ν	0.67	0.004	0.66	0.68
β	3.17	0.013	3.14	3.19

Table 7
Comparing degradation model parameters.

	Assumed	ILI_1	ILI_2	ILI_3
k	0.12	0.51	0.13	0.16
ν	0.771	0.37	0.72	0.67
β	3.5	3.07	3.03	3.17
α_{11}	1.65	1.50	–	–
α_{21}	1.94	–	1.75	–
α_{22}	0.54	–	0.53	–
α_{31}	2.14	–	–	1.97
α_{32}	0.81	–	–	0.85
α_{33}	0.41	–	–	0.47

Table 8
Comparing No. of pits.

	Assumed Pit.No.	Estimated Pit.No.
0-ILI_1	2374	1816
ILI_1-ILI_2	577	725
ILI_2-ILI_3	381	329
ILI_3-Prediction	631	544

It is also worth noting that the standard deviation of k and ν are decreasing by implementing new evidences (new ILI data) as shown in Tables 4, 5, and 6. In addition, the standard deviation of k and ν in Table 6 are small, however, in reality, it is expected to have a higher uncertainty in the estimated parameters. The reason is related to this assumption that all pits are under the same operational conditions (assumed constant values in Table 2, without uncertainty) for the whole life-cycle of the pipeline. However, in reality, the operational conditions are different for different locations and times. The authors have addressed change in operational conditions in a defect-based algorithm [11]. Considering change in operational conditions in this population-based algorithm is one aspect of the future work of this research.

The assumed and estimated number of exiting pits at each time interval is given in Table 8. Which relatively is a good performance considering both probability of detection and probability of false calls.

The next comparison is on the rate parameter of the HPP. As mentioned in phase III of the simulated case study, the estimated rate for the underlying HPP is 72 which is in the ±10% boundary of the assumed value (80). This ±10% criteria is selected based on the ±10% boundary that is accepted for depth estimation in the oil and gas industry [8].

In order to quantify the difference between simulated and estimated distribution of the actual depth of the existing pits at each evaluation time, Kullback Leibler divergence (KLD) [36] measure is used. Kullback Leibler divergence is a measure to find the relative entropy between two probability distribution of a random variable. Eq. (29) shows this measure between two probability distribution functions (P(x) and Q(x)).

$$KLD(P \parallel Q) = \sum_{x \in X} P(x) log \frac{P(x)}{Q(x)} \tag{29}$$

In this equation, usually P is considered as the true distribution of data (in this case simulated actual depth of the existing pits) and Q comes from the proposed model that is used to approximate the

Table 9
Symmetric Kullback Leibler divergence and χ^2 measures.

	X_1	X_2	X_3	X_4
SKLD	0.187	0.226	0.112	0.030
χ^2	0.233	0.075	0.397	0.033
χ^2 (Shifted distributions)	1.994	1.999	1.990	1.984

true distribution (in this paper estimated actual depth of the existing pits(X_j)) [37]. Larger Kullback Leibler divergence indicates less similarity between the two distributions. Since Kullback Leibler divergence is not a symmetric distance, we used symmetric Kullback Leibler divergence to compare similarity between two distributions [6,38] according to Eq. (30).

$$SKLD(P \parallel Q) = 1/2[KLD(P \parallel Q) + KLD(Q \parallel P)] \tag{30}$$

In case of having two distributions that are exactly the same, the Kullback Leibler divergence and symmetric Kullback Leibler divergence are equal to zero. In other words, the smaller value for Kullback Leibler divergence or symmetric Kullback Leibler divergence the two distributions are more similar. The symmetric Kullback Leibler divergence for the normalized PMF of the simulated and estimated actual depth of the existing pits at evaluations times are shown in Table 9. According to this results, the developed population-based algorithm has an acceptable performance (relatively small values for symmetric Kullback Leibler divergence) and its performance is improved by arriving new ILI data (i.e., values for X_3 and X_4 in comparison to the values for X_1 and X_2).

In addition to symmetric Kullback Leibler divergence measure, we compared the results with χ^2 distance measure (Eq. (31)).

$$\chi^2 = \sum_h \frac{(P_h - Q_h)^2}{(P_h + Q_h)} \tag{31}$$

where P and Q are the normalized histograms to be compared and index h refers to h_{th} bin.

This distance measure for different evaluation times are given in Table 9. Analogous to KLD measure, the smaller χ^2 shows that the two compared distributions are more similar. In order to have a criteria to compare the calculated χ^2 distance, this measure is also calculated for the case when there is no overlap between the distributions of the simulated and the estimated actual depth (by shifting the simulated distributions from the estimated distribution in a way to have no overlap). According to Table 9, the χ^2 results also show that the developed algorithm have a good performance in terms of predicting the PMF of the actual depth of the existing pits at inspections times and also at prediction time after the last ILI.

5. Conclusion

The paper presents a novel hybrid, population-based algorithm to evaluate pipeline degradation due to internal pitting corrosion when the pit density is high. This algorithm, in combination with defect-based models, can be used to estimate a pipeline's degradation level when a pipeline contains a combination of high and low density corroded segments. This combined approach is more cost effective by eliminating the need of matching the ILI data for those pits that are not critical.

This algorithm is a hybrid one that takes into account both physics of failure and inspection data. It employs the NHGP to address the temporal uncertainty and the physics of failure knowledge about pitting corrosion process and homogeneous Poisson process to model the variation in pits' initiation times. This algorithm incorporates the sequential ILI data in degradation analysis by using the HB method and Markov Chain Monte Carlo simulation. A clustering algorithm is developed to cluster the detected pits based on their initiation times to be used in the corresponding HB model. Different types of measurement uncertainty (measurement error, POD, and POFC) are taken into account in this algorithm.

An example involving simulated ILI data used to illustrate the developed algorithm. We assumed that three ILI datasets are available for a pipeline segment. In addition to degradation model parameters, we estimated the number of existing pits and the distribution of their actual depth at each ILI time. In prognostic phase, the rate parameter of the underlying HPP is estimated by using maximum likelihood estimation and subsequently the number of new initiated pits in the future. Finally we estimated the PMF of the actual depth of the existing pits in the future which is the main input in pipeline reliability analysis and condition-base maintenance optimization.

We verified the result of this algorithm, for the simulated case study, by comparing the assumed and the estimated degradation model parameters, number of initiated pits, and the rate parameter of the HPP. In addition, we compared the PMF of the simulated and estimated actual depth of the exiting pits at each evaluation time by using the symmetric Kullback Leibler divergence and Chi-Square methods.

It is worth noting that lack of the real ILI 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.

Future research should estimate degradation level to evaluate the reliability of the pipeline (i.e., the probability of different failure modes (small leak, large leak, and rupture)) to be used in maintenance optimization. The recommendation for future work of this study is to use non-homogeneous Poisson process to model the variation in pits' initiation times. Considering different inspection tools with different characteristic will be investigated as well. Another potential area to extend this study is to find an optimal number of bins to discretize the PMFs, since this algorithm is sensitive to this number. Furthermore, this model has considered the mean value of the estimated depth in the calculation. It is worthwhile to add another level of complexity and consider the variance of the estimated depth to take into account the pits' size uncertainty in the estimation.

CRediT authorship contribution statement

Roohollah Heidary: Conceptualization, Methodology, Coding, Visualization, Writing - original draft. **Katrina M. Groth:** Supervision, Investigation, Validation, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

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.

References

[1] Velazquez J, Caleyo F, Valor A, Hallen J. Predictive model for pitting corrosion in buried oil and gas pipelines. Corrosion 2009;65(5):332–42.

[2] Dann MR, Maes MA. Stochastic corrosion growth modeling for pipelines using mass inspection data. Reliab Eng Syst Saf 2018;180:245–54.

[3] An D, Kim NH, Choi J-H. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. Reliab Eng Syst Saf 2015;133:223–36.

[4] Heidary R, Groth KM, Modarres M. Fusing more frequent and accurate structural damage information from one location to assess damage at another location with less information. In: Proceedings of the 14th probabilistic safety assessment and management conference (PSAM 14). 2018.

[5] Dann MR, Maes MA, Salama MM. Pipeline corrosion growth modeling for in-line inspection data using a population-based approach. In: ASME 2015 34th international conference on ocean, offshore and arctic engineering. American Society of Mechanical Engineers Digital Collection; 2015.

[6] Ramírez J, Segura JC, Benítez C, De La Torre A, Rubio AJ. A new Kullback-Leibler VAD for speech recognition in noise. IEEE Signal Process Lett 2004;11(2):266–9.

[7] Maes MA, Faber MH, Dann MR. Hierarchical modeling of pipeline defect growth subject to ILI uncertainty. In: ASME 2009 28th international conference on ocean, offshore and arctic engineering. American Society of Mechanical Engineers; 2009, p. 375–84.

[8] Zhang S, Zhou W. System reliability of corroding pipelines considering stochastic process-based models for defect growth and internal pressure. Int J Press Vessels Pip 2013;111:120–30.

[9] Zhang S, Zhou W, Qin H. Inverse Gaussian process-based corrosion growth model for energy pipelines considering the sizing error in inspection data. Corros Sci 2013;73:309–20.

[10] Zhang S, Zhou W. Bayesian dynamic linear model for growth of corrosion defects on energy pipelines. Reliab Eng Syst Saf 2014;128:24–31.

[11] Heidary R, Groth KM. A hybrid model of internal pitting corrosion degradation under changing operational conditions for pipeline integrity management. Struct Health Monit 2019.

[12] Dann MR, Dann C. Automated matching of pipeline corrosion features from in-line inspection data. Reliab Eng Syst Saf 2017;162:40–50.

[13] Lu D. Estimation of stochastic degradation models using uncertain inspection data. University of Waterloo; 2012.

[14] Ossai CI, Boswell B, 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 Technol 2016;51(5):342–51.

[15] Shibata T. 1996 WR whitney award lecture: Statistical and stochastic approaches to localized corrosion. Corrosion 1996;52(11):813–30.

[16] Tsukaue Y, Nakao G, Takimoto Y, Yoshida K. Initiation behavior of pitting in stainless steels by accumulation of triiodide ions in water droplets. Corrosion 1994;50(10):755–60.

[17] Chatterjee K, Modarres M. A probabilistic approach for estimating defect size and density considering detection uncertainties and measurement errors. Proc Inst Mech Eng O 2013;227(1):28–40.

[18] Valor A, Caleyo F, Alfonso L, Velázquez J, Hallen J. Markov chain models for the stochastic modeling of pitting corrosion. Math Probl Eng 2013;2013.

[19] Baroux B. The kinetics of pit generation on stainless steels. Corros Sci 1988;28(10):969–86.

[20] Aziz P. Application of the statistical theory of extreme values to the analysis of maximum pit depth data for aluminum. Corrosion 1956;12(10):35–46.

[21] Valor A, Caleyo F, Alfonso L, Rivas D, Hallen J. Stochastic modeling of pitting corrosion: a new model for initiation and growth of multiple corrosion pits. Corros Sci 2007;49(2):559–79.

[22] Zhang S. Development of probabilistic corrosion growth models with applications in integrity management of pipelines. 2014.

[23] Heidary R, Gabriel SA, Modarres M, Groth KM, Vahdati N. A review of data-driven oil and gas pipeline pitting corrosion growth models applicable for prognostic and health management. Int J Progn Health Manage 2018;9(1).

[24] Abdel-Hameed M. A gamma wear process. IEEE Trans Reliab 1975;24(2):152–3.

[25] Van Noortwijk J. A survey of the application of gamma processes in maintenance. Reliab Eng Syst Saf 2009;94(1):2–21.

[26] Nuhi M, Seer TA, Al Tamimi A, Modarres M, Seibi A. Reliability analysis for degradation effects of pitting corrosion in carbon steel pipes. Procedia Eng 2011;10:1930–5.

[27] Vanaei H, Eslami A, Egbewande A. A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models. Int J Press Vessels Pip 2017;149:43–54.

[28] Bazán FAV, Beck AT. Stochastic process corrosion growth models for pipeline reliability. Corros Sci 2013;74:50–8.

[29] Romanoff M. Underground corrosion, Vol. 579. US Government Printing Office; 1957.

[30] Scarf P, Cottis R, Laycock P. Extrapolation of extreme pit depths in space and time using the r deepest pit depths. J Electrochem Soc 1992;139(9):2621.

[31] Hunkeler F, Bohni H. Mechanism of pit growth on aluminum under open circuit conditions. Corrosion 1984;40(10):534–40.

[32] Zhang W, Frankel G. Localized corrosion growth kinetics in AA2024 alloys. J Electrochem Soc 2002;149(11):B510–9.

[33] Kelly D, Smith C. Bayesian inference for probabilistic risk assessment: A practitioner's guidebook. Springer Science & Business Media; 2011.

[34] Der Kiureghian A, Ditlevsen O. Aleatory or epistemic? Does it matter?. Struct Saf 2009;31(2):105–12.

[35] Pandey MD. Probabilistic models for condition assessment of oil and gas pipelines. NDT E Int 1998;31(5):349–58.

[36] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22(1):79–86.

[37] Rabiei E, Droguett E, Modarres M. Fully adaptive particle filtering algorithm for damage diagnosis and prognosis. Entropy 2018;20(2):100.

[38] Tabibian S, Akbari A, Nasersharif B. A new wavelet thresholding method for speech enhancement based on symmetric Kullback-Leibler divergence. In: 2009 14th international CSI computer conference. IEEE; 2009, p. 495–500.