

RESEARCH

Open Access



# Fatigue life prediction of pipeline with equivalent initial flaw size using Bayesian inference method

Milad Salemi and Hao Wang \*

## Abstract

Majority of pipeline infrastructure are old and susceptible to possible catastrophic failures due to fatigue. Timely maintenance is the key to keep pipeline in serviceable and safe condition. This paper proposed a Bayesian inference methodology based on the observed crack growth measurements and cycle data that predicts the probability density of failure after initially estimating the equivalent initial flaw size (EIFS). The model was first developed based on one-dimensional crack growth problem in plate with edge crack. Then the model was expanded to two-dimensional crack growth problem in pipe wall. Stress intensity factors (SIF) at the crack tip in pipe model were calculated using finite element (FE) analysis for different crack lengths and depths. Polynomial function and Gaussian process were used to develop surrogate models of SIF. The analysis demonstrated that the proposed Bayesian inference method with hyperparameters generated accurate inferred results for probability density function (PDF) of both EIFS and the number of cycles to failure.

## Introduction

Fatigue cracking is an inherently stochastic problem affected by various sources of variability and uncertainty. In the particular case of oil and gas pipeline fatigue-induced cracks, it is important to effectively and efficiently monitor and inspect pipelines. Due to the old age and large size of pipeline transmission network, maintaining these infrastructures within safe and serviceable conditions is not an easy task. As a result, it is important to contentiously improve and update efficiency of means and interval of monitoring and maintenance by use of innovative scientific methodologies. Such methodologies should be able to account for probabilistic nature of fatigue process, such as uncertainties in material properties, variability of internal pressure within the pipe, and reliability and accuracy of inspection data about the condition of pipe structure.

Traditionally probabilistic models within the *classical statistics* scope deal with analyzing statistics of the error

due to differences between model predictions and measurement/observation data. The accuracy of such models rely upon the quality and quantity of available observation data. On the other hand, *Bayesian statistics* view on model definition quantifies the extent to which the model represents the data and determines the probability of the model being correct itself. This makes Bayesian models suitable for cases in which the relevant data is limited.

In recent years there has been a great focus on data driven solutions with help of machine learning methods in detecting patterns and performance prediction in various fields of science and engineering. Aeronautical engineering and aircrafts structural failures has been the major front for doing research related to fatigue in various metallic alloys [1–6]. Knowledge about equivalent initial flaw size (EIFS) in material is key to estimate crack initiation and progression process [7]. Different researchers have studied various aspects of applying Bayesian inference in fatigue modeling. Makeev et al. [3] investigated Bayesian inference of EIFS using Weibull and log-normal Distribution. In the follow up work, Cross et al. [1] extended the inference of Makeev's work

\* Correspondence: [hwang.cee@rutgers.edu](mailto:hwang.cee@rutgers.edu)

Department of Civil and Environmental Engineering, School of Engineering, Rutgers, The State University of New Jersey, Piscataway, USA

to multivariable inference of other crack growth parameters in addition to EIFS using Markov Chain Monte Carlo (MCMC) method. Liu et al. [2] used a fracture-mechanics based approach to estimate EIFS based fatigue limit of the material and compared the results with the observed material microstructure and subsequently presented prediction on fatigue life. Sankararaman et al. [4, 5] performed detailed inference of multiple parameters involved in crack growth model in conjunction with finite element (FE) simulations under variable amplitude loading. Xie et al. [8] used Bayesian inference to infer parameters of crack growth model in addition to fatigue life. They tested their model with pipeline field data and reported desirable results.

Gobbato et al. [9] introduced a recursive Bayesian prognosis methodology to predict and update remaining fatigue life (RFL) of structural components in aerospace structures using continues non-destructive evaluation (NDE) data. Ribeiro et al., [10] further employed Bayesian inference to estimate RFL of fixed offshore structures. Babuška et al. [11] used Bayesian network to predict fatigue parameters of 75S-T6 aluminum alloys by using strain-life (S-N) curve data. Arzaghi et al. [12] used Bayesian network inference to propose a dynamic decision making and planned maintenance framework for sub-sea pipelines. Li et al. [13] recommended a probabilistic health prognosis for aircraft wing structure due to an edge crack using dynamic Bayesian networks (DBN). They considered uncertainties and proposed methodology which reduced computational cost by modifying DBN updating interval requirements.

## Objective and scope

This study aims to employ a combination of these approaches to establish a hybrid model that improves its prediction for future state of pipeline condition as more inspection data is entered into the model. The model has backward and forward inference. At each addition of data point on number of cycle and crack size first the model infers the equivalent initial flaw size (EIFS) probability density (PDF) and subsequently uses that to infer the PDF for number of cycles to failure. The model can be used to prioritize maintenance orders by isolating sections that are at greatest risk of failure.

First, a base Bayesian framework model is established based on synthetically generated data sets generated from crack growth in steel plate from available equations in literature for edge crack in plates. Secondly, FE simulations were performed on pipe model for several cases of crack length and depth in pipe wall thickness. Surrogate models were fitted to compute SIF for intermediary points among actual FE simulated SIF values. The proposed model predicts the failure based on 2-dimssional crack growth in the pipe wall thickness. Results of the proposed

methodology are presented and ground truth values were compared with inferred values.

## Base model development and results for steel plate

### Equivalent initial flaw size (EIFS) definition concept

Using the available empirical solution for crack propagation in plates with a single edge notch in a load controlled boundary conditions (constant remote tensile stress) the base model was established. To calculate stress intensity factor (SIF) at a given crack length, the equation proposed by [14] was used to calculate shape factor in the general SIF formula:

$$K_I = \sigma \sqrt{\pi a} F\left(\frac{a}{b}\right) \quad (1)$$

$$F\left(\frac{a}{b}\right) = \sqrt{\frac{2b}{\pi a} \tan\left(\frac{\pi a}{2b}\right)} \times \frac{0.752 + 2.02\left(\frac{a}{b}\right) + 0.37\left(1 - \sin\left(\frac{\pi a}{2b}\right)\right)^3}{\cos\left(\frac{\pi a}{2b}\right)} \quad (2)$$

Where  $K_I$  is mode-I SIF,  $\sigma$  is applied nominal constant remote stress,  $a$  is crack length,  $b$  is plate width, and  $F$  is shape factor given by Tada's equation and is valid for any  $\frac{a}{b}$  ratio.

The crack growth model will follow Paris' crack growth regime [15]:

$$\frac{da}{dN} = C1(\Delta K)^{C2} \quad (3)$$

Where,  $C1$  and  $C2$  are material properties calibrated based on experimental data, and  $\Delta K$  is stress intensity factor range between minimum and maximum applied remote stress.  $\frac{da}{dN}$  is the rate of crack growth with the change of loading cycle.

Combining the equation for SIF and Paris's regime:

$$N = \int_c^a \frac{da}{C1(\Delta K)^{C2}} \quad (4)$$

$$N = \int_c^a \frac{da}{C1(\sigma \sqrt{\pi a} F\left(\frac{a}{b}\right))^{C2}} \quad (5)$$

Where,  $c$  is initial crack length. This integral can be numerically calculated using Simpson's method.

The experimental data follows a general function in the form of:

$$N = f(a, \phi)$$

Where,  $\phi$  is the vector of model parameters such as loading, material properties, and geometry. The elements of  $\phi$  are considered random variables in a probabilistic scheme. The parameters governing the distribution of each random variable in  $\phi$  are called hyper-parameters, .

The probability of observing target cycle for a given data point of  $(N, \phi)$  can be framed using normal distribution.

$$p(N|a, \theta, \beta) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{[N-f(a, \phi)]^2}{2\beta^2}\right) \quad (6)$$

Where,  $p$  is conditional probability of observing  $N$  cycle given model parameters, which includes EIFS, crack length  $a$  at cycle  $N$ , and associated noise in the crack growth model with standard deviation of  $\beta$ :

$$\log(\hat{N}) = \log(N) \pm \beta \quad (7)$$

Where,  $\hat{N}$  is the noisy measurement of the cycle compared to the true cycle.

Using Bayesian inference the joint probability of target data,  $N$ , given model for  $k$  number of data points can be derived as:

$$\begin{aligned} p(N|a, \{\phi_k\}, \beta) &= \prod_{k=1}^m p(N_k|a_k, \phi_k, \beta) \\ &= \prod_{k=1}^m \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{[N_k-f(a_k, \phi_k)]^2}{2\beta^2}\right) \end{aligned} \quad (8)$$

The initial flaw size (IFS) in fatigue analysis refers to the small flaws that exist in material's grain size scale. Such flaws do not necessarily follow crack growth regimes such as Paris' equation which relates *long crack* propagation rates to material properties. The small crack growth rate exhibits oscillatory behavior compared to generally monotonic behavior in long crack growth rates.

The equivalent initial flaw size (EIFS) is the concept that allows interpretation of IFS into long crack analysis realm. This allows implementation of long-crack based crack propagation models from the beginning life of a material undergoing cyclic loading. In other words, the number of cycles required to reach a certain crack length  $a_f$  considering IFS and EIFS are equal. Their respective short-crack and long-crack growth models  $g_s(a)$  and  $g_l(a)$  are as follows [2]:

$$N_f = \int_{IFS}^{a_f} \frac{da}{g_s(a)} = \int_{EIFS}^{a_f} \frac{da}{g_l(a)}$$

#### EIFS direct probability density inference - $p(\theta)$

To develop and verify base model, it was assumed that all the uncertainties are the result of EIFS or initial crack length,  $\theta$ . The modified distribution is as follows:

$$p(N|a, \theta, \beta) = \prod_{k=1}^m p(N_k|a_k, c_k, \beta)$$

$$= \prod_{k=1}^m \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{[N_k-f(a_k, c_k)]^2}{2\beta^2}\right) \quad (9)$$

Where now only EIFS,  $c$ , is the only parameter whose distribution is directly inferred.

To solve for  $p$  for each data point,  $k$ , we have to solve for  $\hat{N}$  by solving for  $f(a_k, c_k)$  following Eq. (5). To calculate the one dimensional integral in Eq. (5) Simpson's quadratic integration rule can be employed which is discussed in detail in [16].

To evaluate the proposed base model, various sets of data points were generated and  $\beta = 10\%$  noise was added to the data to account for uncertainty within crack growth model. The training data were generated according to the following framework:

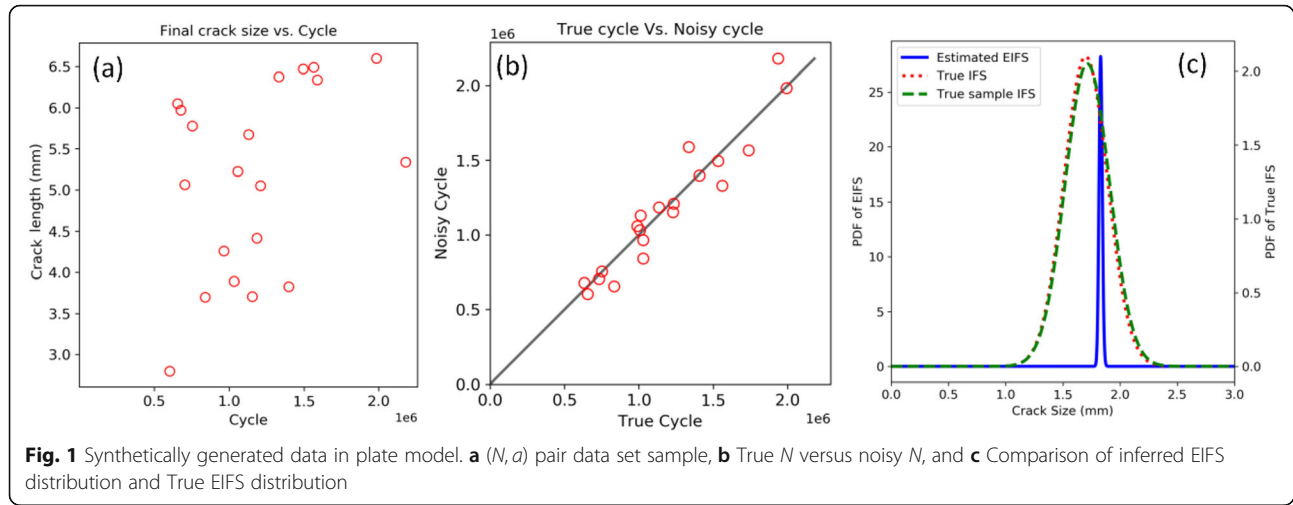
- 1- It was assumed that the initial flaw size,  $\theta$ , follows a normal distribution with assumed mean and standard deviation.  $k$  samples were drawn from this distribution.
- 2- Uniform distributions was assumed for final crack size,  $a$ , and  $k$  samples were drawn from this distribution. The lower bound of this uniform distribution was assumed to be at least 3 times of the assumed standard deviation of the normal distribution for  $c$ , larger than the mean assumed for  $c$ .
- 3- Given  $a$  and  $c$  for each data point the corresponding number of cycles,  $N$ , for that data point was calculated following Eq. (5) and using Simpson's quadratic numerical integration method.
- 4- A random noise with zero mean and  $\beta$  standard deviation was added to the calculated  $N$ .

Figure 1 shows a sample of synthetically generated data points. In Fig. 1(a) final crack size and its corresponding cycle number is visualized for  $k$  data point, in which for this example they are 20 data points.

Table 1 shows the selected properties from previous study [8] and assumed parameters for crack growth in the base model for a steel plate with an edge crack.

For each data point  $(N, a)$  an EIFS likelihood probability density is calculated and the product of all data point likelihoods as shown in Eq. (9) will establish estimated probability density of EIFS. Figure 2 shows box plots of EIFS PDF for the dataset containing 20 data points. The variation of the inferred EIFS distribution at each data point is clearly distinguishable.

The final inferred EIFS distribution shown in Fig.1(c) and summarized in Table 2 based on 20 data points shows that the estimation for mean is very close to the mean of the true distribution but the standard deviation or spread is not captured very well. To improve the inferred distribution, we will try to infer the distribution



parameters instead of distribution itself in the next section.

#### EIFS probability density inference with hyper parameters- $p(\theta|\alpha)$

In this section instead of inference on EIFS itself, the parameters of the assumed governing distribution for EIFS conditioned on its distribution parameters will be inferred:

$$p(\theta|\alpha) = \prod_{k=1}^m p(\theta_k|\alpha) \quad (12)$$

$$\begin{aligned} p(N|\alpha, \beta) &= \int p(N|\alpha, \theta, \alpha, \beta) p(c|\alpha, \beta) d\mathbf{c} \\ &= \int p(N|\alpha, \theta, \beta) p(c|\alpha) d\theta \end{aligned} \quad (13)$$

$$p(N|\alpha, \beta) = \prod_{k=1}^m \int \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{[N_k - f(a_k, \theta)]^2}{2\beta^2}\right) p(\theta|\alpha) d\theta \quad (14)$$

Now in this equation where we have introduced a prior probability for EIFS,  $p(\theta|\alpha)$ , which is conditioned on  $\alpha$ . This forms a hierarchical Bayes model. In this approach an additional integration is introduced. This

integration can be solved by applying Simpson's quadratic integration twice.

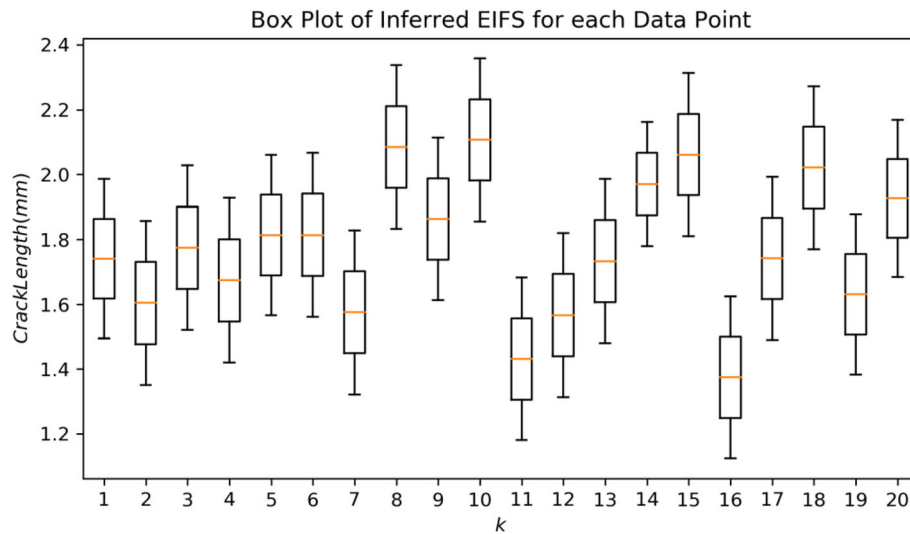
Let us elaborate inference process of hyper-parameters of EIFS distribution. We select some possible lower and higher bounds for both mean and standard deviation. Mathematically the bounds for mean can be from  $-\infty$  to  $+\infty$ , and for standard deviation it can be from 0 to  $+\infty$ ; but to save a huge computation cost especially if we are considering a finer step size or mesh, with a good guess based on engineering judgement, we can select appropriate bounds. The physics of the problem such as maximum possible crack depth and minimum possible crack depth, a non-negative value, would give initial intuition.

In the next step we select a small enough step size to capture a smooth outcome. This can be achieved by running few trial cases. After selecting the appropriate step size for both standard deviation range and mean range a  $\theta$  distribution can be generated (bounds of  $\theta$  (EIFS) was chosen from zero to close to maximum possible crack depth or plate thickness, b).

Integrating the resulting joint probability density along mean and standard deviation separately yields the marginal probability densities of mean and standard deviation of the EIFS. As it can be seen from Fig. 3(d) and the accuracy of EIFS distribution estimation using hierarchical model and hyper-parameters has increased significantly. It is clear that inference directly on EIFS distribution can estimate the mean within an acceptable range with less than 0.1 mm error (less than 6%), but it fails to predict standard deviation and it has more than 90% error in standard deviation prediction. On the other hand, the hierarchical model was able to predict both mean and standard deviation with high accuracy. The error for estimating EIFS distribution mean was 2% and error for estimating standard deviation of EIFS distribution was 4.6%.

**Table 1** Material and model parameters used in the plate model

Parameter	Definition - Value
C1	3.39e-13
C2	2.9
$\beta$	10%
$\sigma$	40 MPa
b (plate width/pipe wall thickness)	7.1 mm



**Fig. 2** Box plot illustrating EIFS distribution for individual data points

Comparison of results between simple maximum likelihood (direct EIFS inference) and hierarchical Bayesian analysis are shown in Table 2.

#### Estimating probability and cycle of failure

To estimate probability of failure we need to set a failure crack length/depth criteria,  $a_f$ . This could be a percentage of the width in which crack is being propagated. From previous section we determined probability density of the EIFS for the plate problem. For the plate problem only  $a$  is the variable of crack that is growing and crack length does not apply.

Now, having the EIFS distribution, distribution of number cycles required to reach failure can be proactively estimated as inspection data is gathered over time. In the case of pipe problem, the inspection data could include length and depth of the cracks detected in the wall of pipe section.

To test this hypothesis, synthetic inspection data points corresponding the number of cycles and respective crack length at that cycle (or time) needed to generated. Paris' crack growth model was used to progress crack length from the assumed initial size to assumed final crack size for each synthetically generated data point.

Initial crack length is randomly sampled from EIFS distribution. As for the failure crack depth, a certain

percentage of plate width,  $b$ , (and later pipe wall thickness in the pipe model) is assumed as the mean of failure crack depth (maximum allowable crack length). To include uncertainty for the failure crack length, a standard deviation is added to the assumed mean of failure crack depth to generate a normal distribution for the final crack length. In addition, for each randomly sampled EIFS, a corresponding failure crack depth is sampled from failure crack depth distribution and hence the number of cycles were calculated as shown in the flow chart in Fig. 4. The flow chart shown in this figure can be used in either plate problem or pipe problem. In the plate problem, being a 2D model, crack only grows in one dimension (one crack front), while in the pipe model crack grows along length of pipe and depth (through pipe thickness). Consequently, their corresponding SIF functions are single variate and bivariate, respectively.

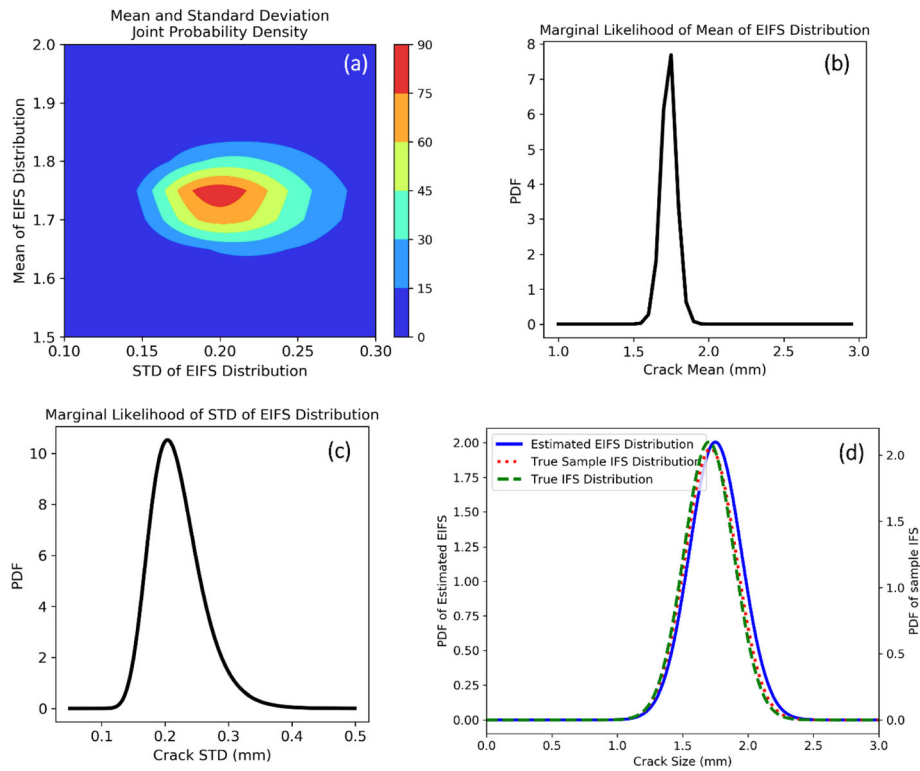
For each synthetic data point generated using the algorithm that was introduced earlier, the number of cycles to the failure point corresponding to the sampled EIFS and failure crack depth is estimated. Using Bayesian analysis two forecast scenarios are viable:

- 1- Estimating the distribution of number of loading cycles to failure ( $N_f$ ) based on observed numbers of data points sampled from EIFS and failure crack

**Table 2** Comparison of EIFS inference results using discussed methods

Parameters	True Value	Inference on EIFS	Inference on Hyper Parameters
Mean	1.715 mm	1.831	1.750 mm
Standard deviation	0.195 mm	0.014	0.204 mm





**Fig. 3** **a** Joint probability density function of mean and standard deviation of EIFS, **b** Marginal probability density of mean of EIFS, **c** Marginal probability density of standard deviation of EIFS, **d** Estimated EIFS density vs. True assumed density for EIFS

depth distributions. This can only be done on the synthetic data or field data where data was collected after observing actual failure in the field or lab test.

- 2- Estimating the distribution of number of loading cycles to failure ( $N_f$ ) based on observed numbers of data points sampled from EIFS and any random final crack depth that is sampled from a uniformly distributed final crack size from the sampled EIFS to any value smaller than assumed value for mean of failure crack depth. i.e. We will be able to predict probability distribution of number of loading cycles to failure from observed data (loading cycles vs crack depth) as new data points are added.

It should be mentioned the distributions resulted from scenario-1 may be used as an informative prior in the Bayesian process of estimating the distribution of number of cycles to failure in scenario-2. The mathematical expression is as follows:

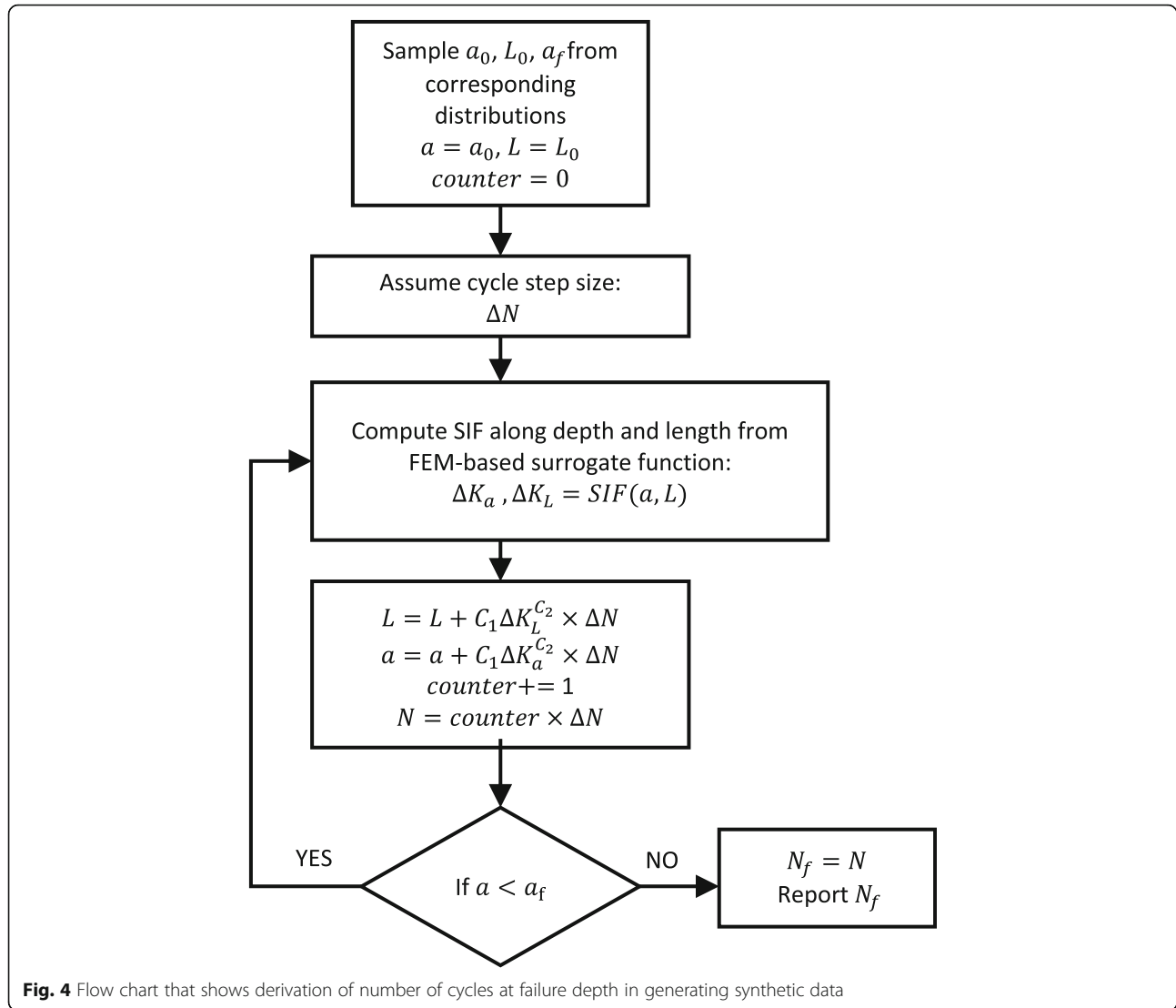
$$p(N|N_{range}, \beta) = \prod_{k=1}^m p(N_k|N_{range}, \beta)$$

$$= \prod_{k=1}^m \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{[N_k - N_{range}]^2}{2\beta^2}\right)$$

Where,  $N_{range}$  is a vector with lower bound and higher bound that covers the range of the number of cycles that may cause failure.

The upper bound of  $N_{range}$  can be calculated by assuming an extreme case of smallest possible EIFS and largest possible crack depth. The probability of such combination is very low in reality and even in case of synthetically generated data. For example, Fig. 5, shows an illustration of possible failure cycle outcomes for a steel plate with an edge crack. This example includes 250,000 data points which was generated by sampling 500 EIFS and 500  $a_f$  from distributions  $EIFS \sim N(1.5, 0.12)$  and  $a_f \sim N(6, 0.1)$ , respectively. As it can be seen in Fig. 5, for this particular example the number of cycles does not reach 3 million. While calculating the number of cycles for the extreme case by considering  $IFS = 0.1mm$ , and  $a_f = 6.99mm$ , yields the maximum number of cycles at about 3.8 millions.

With this overview of methodology, we now investigate the estimations of number cycles to failure for



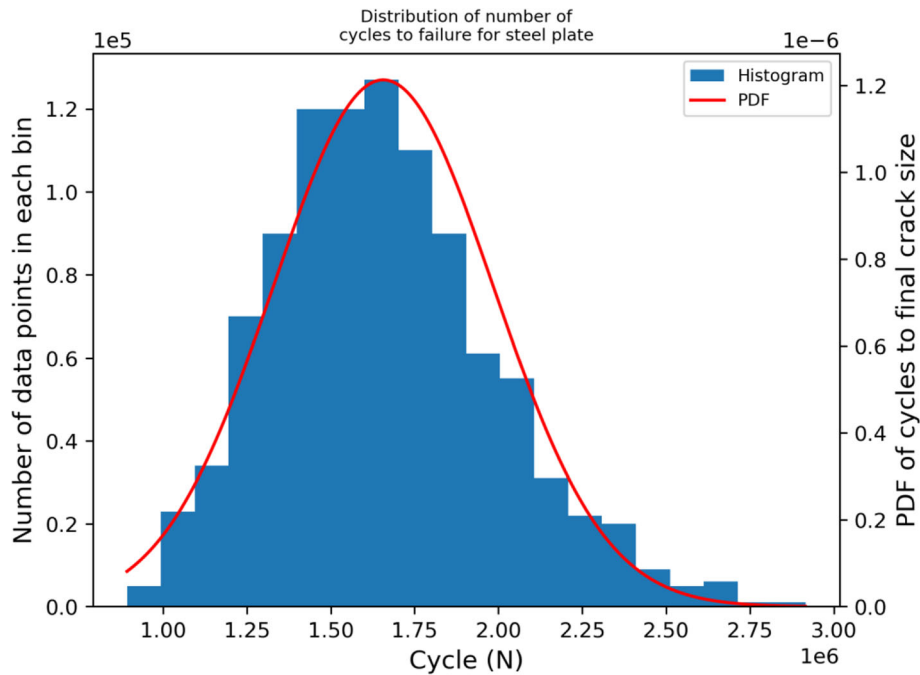
various synthetic data sizes by directly estimating  $N_{range}$  distribution for the plate problem with edge crack. The following flowchart, shown in Fig. 6 demonstrates the process for estimating the distribution of number of cycles to failure.

The non-informative prior was selected as unit integer. The informative prior may be a PDF generated the similar way as the example shown in Fig. 5 by sampling from EIFS and an assumed failure crack length. In this particular method using a good informative prior will reduce the standard deviation of the final posterior PDF.

To demonstrate this methodology with example, results of various estimations for number of cycles to failure is shown in Fig. 7 (top). In this example the inference was performed on four datasets comprising 5, 15, 30, and 50 data points.

From the Fig. 7 (left) we can observe that as more data points are added (blue 5 samples and red 50 samples) the distribution becomes sharper at peak and closer to the mean of sample distribution shown in black dash lines. In the case of *EIFS* inference this method is better if only a single estimate is desired rather than the whole distribution. Table 3 lists the parameters of the distributions.

To have the standard deviation also represented in our inference process we will again resort to inferring hyper parameters of the  $N_{range}$  distribution as we did for *EIFS*. Figure 7 (middle) shows an example of resulting joint likelihood probability density for mean and standard deviation of the  $N_{range}$  distribution with 50 samples. To the right we can see progression of the  $N_{range}$  distribution inference with 5 samples (blue curve) to 50 samples (red curve). The dashed black distribution is observed sample distribution. As it can



**Fig. 5** An illustrative example of possible distribution of number cycles to failure for a steel plate with thickness of 7 mm. assumptions:  $EIFS \sim N(1.5, 0.12)$ , and  $a_f \sim N(6, 0.1)$

be seen clearly there is very good match between predictions and observed samples. Table 4 summarizes the parameters of this distributions.

It is worth noting that final number cycles were more affected by the distribution of the *EIFS* than the final crack size during process of generating synthetic data. This another instance that the exemplifies importance of *EIFS* distribution inference.

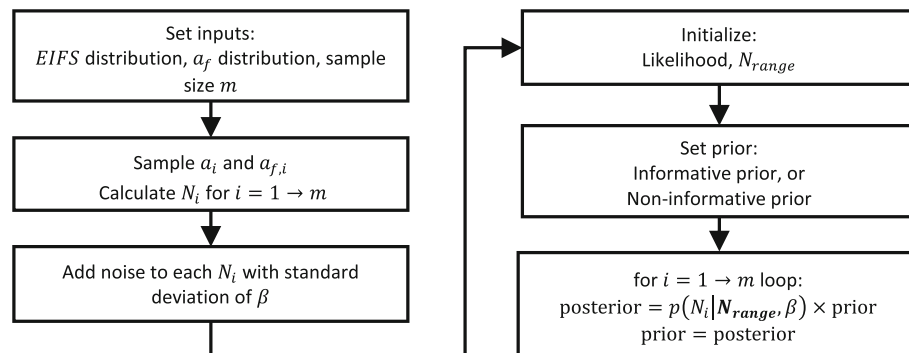
### Pipe model development and Bayesian inference results

In this section first the methodology to compute stress intensity factor (SIF) using FE models is discussed. In

addition, surrogate models were introduced to interpolate SIF values at the crack length and depths that FE simulation were not performed. First, EIFS for crack depth and were inferred and those inferred values were used to estimate the distribution for number of cycles to estimate cycle at final crack length.

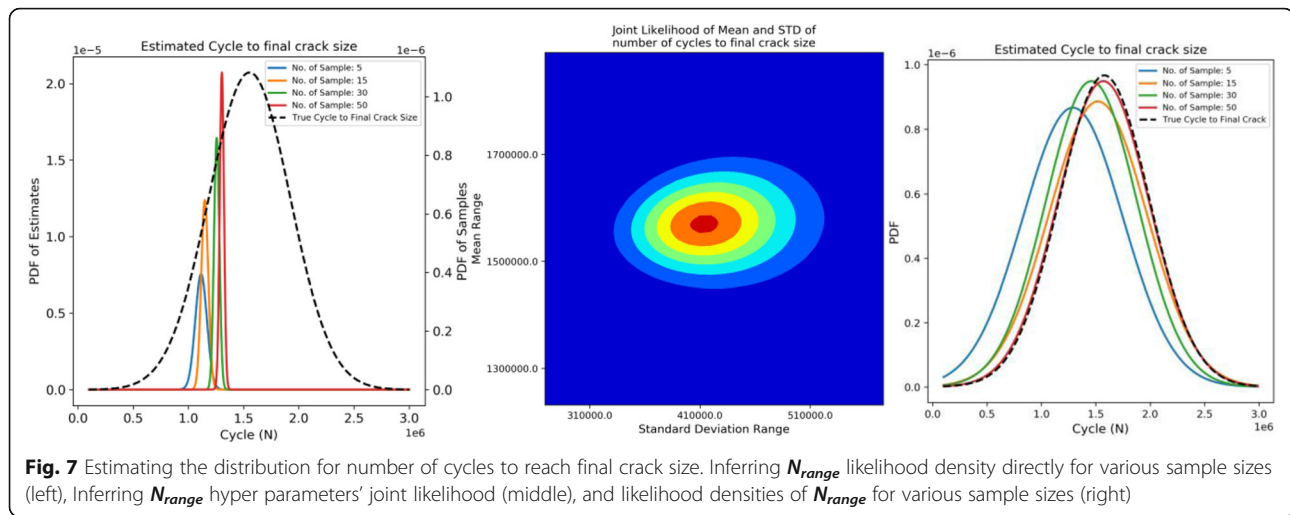
### Finite element (FE) modeling

The finite element modeling of this research included 3D simulation to calculate at the crack tip. Commercially available multi-physics software ABAQUS was used to perform FE simulations. For purpose of this study typical steel material properties ( $E = 200 \text{ GPa}$ ,  $\nu =$



**Fig. 6** Flowchart of deriving estimated distribution for number cycles for final (or failure) crack size distribution





0.3) were assumed for steel and it was assumed that steel has elastic behavior within the scope of SIF simulations.

To investigate crack growth and the respective SIF evaluation more accurately a 3D model was developed and a semi-elliptical (halved ellipse) shape was assumed for the crack shape as previously it was reported in the literature to emulate realistic crack shape more closely [8, 17, 18]. The diameter of pipe was selected as  $D = 863.6\text{mm}$  and pipe wall thickness of  $t = 7.1\text{mm}$ . These numbers were adopted from [8]. The length of the pipe section model was chosen to be  $1000\text{mm}$ .

To reduce running time of each analysis case a preliminary mesh sensitivity analysis was performed to determine largest mesh size where solution reaches stable state and will not be improved by any further mesh size reduction at the vicinity of the crack. This analysis was performed for the crack with length of  $L = 10\text{mm}$  and depth of  $=1\text{mm}$ . The analysis concluded that a mesh size of approximately  $0.5\text{mm}$  (equivalent to 23 nodes) is small enough to produce the stable response.

To further optimize required running time for each analysis case, the symmetry of the model was taken advantage of, to reduce model size. To this end, as shown in Fig. 8, the full pipe model was once reduced in half due to symmetry along y axis (x-z plane) and a second time along z axis (x-y plane). In addition, bias mesh sizing was used to incrementally increase mesh size of

the pipe as we go farther away from location of embedded crack to reduce computation effort.

It should be noted that the region where crack was embedded was always meshed uniformly in size and corresponding to the appropriate mesh size derived from mesh sensitivity analysis that was pointed earlier. This incremental progression of mesh sizing can be observed in Fig. 8(b) where the elliptical crack is embedded in the quartile model (the model with symmetry along Y and Z axis) and mesh dimension along perimeter of pipe and long Z axis is increased as we go away from the crack region. To verify validity of this model reduction approach we performed analysis for all models introduced in Fig. 8. The observed corresponding SIF calculations results in the quartile model was validated and verified as an alternative to the full model.

Figure 8(c) shows a snippet of cross section of the pipe along Z axis where the crack is embedded (crack is positioned in the middle of the pipe). In this cross section we can calculate the SIF for Mode-I fracture along the perimeter of semi-elliptical crack. All combination crack length (10, 15, 20, 25, 30, 35, and 40 mm) and depth (1, 2, 3, 4, 5, and 6 mm) generates 42 simulation cases.

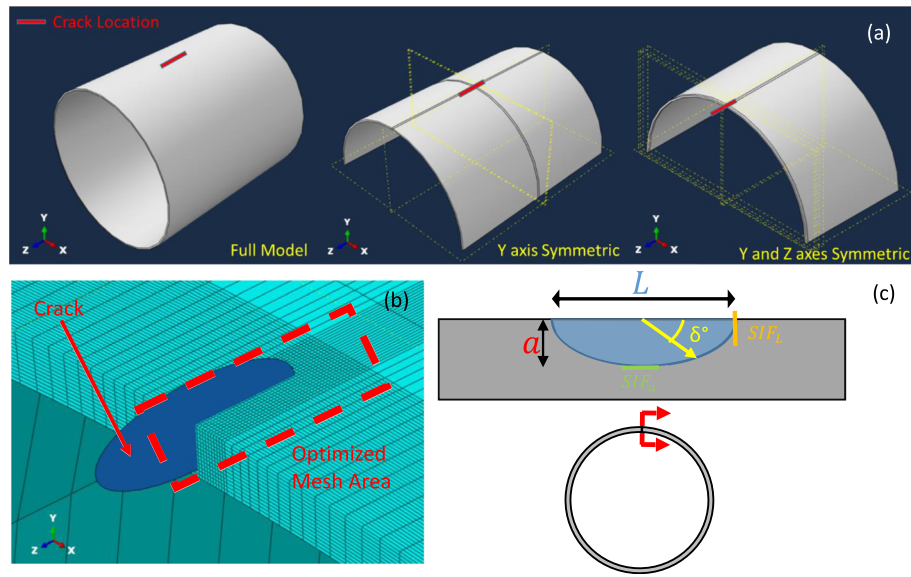
It is worth noting that in Fig. 4 flowchart the criteria for finding maximum number of cycles is controlled by the maximum depth of crack through wall thickness of the pipe. There are two reasoning for this criteria: first

**Table 3** Comparison of direct inference of  $N_{range}$  on different sample sizes

No Samples	STD		Mean	
	True	Estimate	True	Estimate
5	515,847	52,757	1,422,988	1,115,177
15	530,064	32,175	1,569,746	1,148,483
30	425,478	24,247	1,572,589	1,256,449
50	368,879	19,258	1,553,850	1,303,446

**Table 4** Comparison of inference of  $N_{range}$  with hyper parameters on different sample sizes

No Samples	STD		Mean	
	True	Estimate	True	Estimate
5	482,924	460,000	1,318,528	1,290,000
15	425,447	450,000	1,516,247	1,520,000
30	428,778	420,000	1,493,829	1,460,000
50	412,633	420,000	1,578,909	1,570,000



**Fig. 8** 3D pipe model with three different representations for modeling

as the ratio of crack length to crack depth ( $\frac{L}{a}$ ) increases the stress intensity factor along crack length becomes less critical and the stress intensity factor along crack depth is dominant. Second, in the crack growth process, due to physics of the problem the room for crack growth along the length of the pipe is many times larger than the room for crack growth along crack depth through the pipe wall thickness. This methodology can be reduced to only account for crack growth along the depth and neglect the crack length growth evolution if the final crack length is not of the interest for a particular problem such as edge crack growth in the plate problem discussed earlier. It is important to note that while crack length would be neglected as a failure assessment criterion, updating crack length is necessary to have correct crack depth growth progress.

The simulation results reveal that at shorter crack lengths as the depth of crack increases the SIF becomes more critical at the endpoints along the crack length, as shown in Fig. 9(a). As the length of the crack becomes longer the critical SIF remains at the deepest point of the crack. In other words, the longer the crack length becomes the more dominant crack depth becomes. This is the assertion and justification to an earlier explanation about the implementation made in the algorithm in Fig. 4 that crack depth is the main driver behind crack growth progression. The cases illustrated in Fig. 9(a) the critical points of stress intensity are clearly visible at red areas along crack perimeter.

#### Surrogate model of SIF

Due to high computation cost and time, evaluating SIF values for all possible combinations of crack lengths and

depths are not feasible. In a crack growth model and the flowchart shown in Fig. 4 in each cycle SIF needs to be calculated for the updated  $L$  and  $a$ . To this end a function was defined that can continuously compute SIF at any  $L$  and  $a$  within the lower and upper bounds of the crack length and depth that was evaluated in the FE model. This function is known as *surrogate function*.

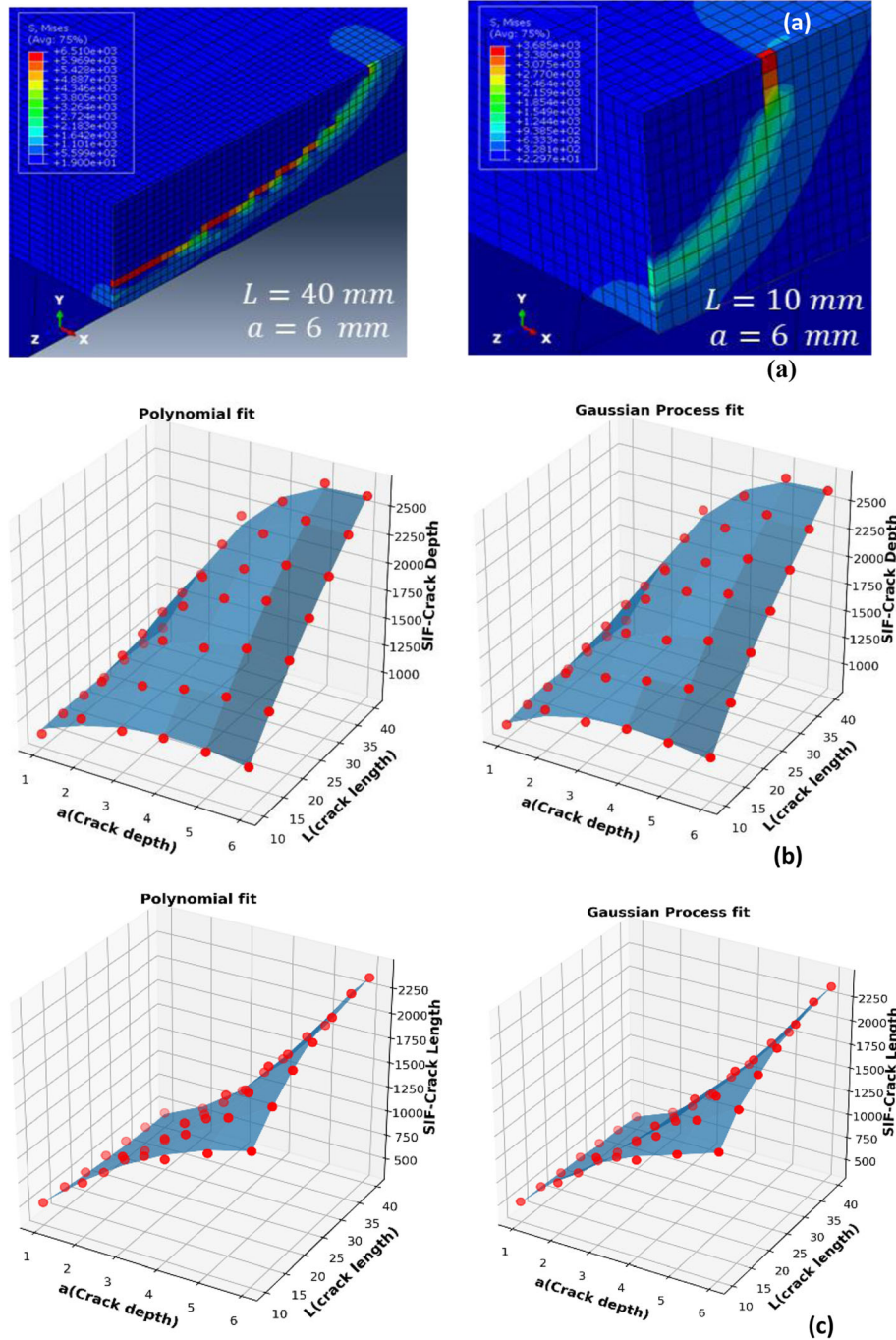
The surrogate function forms a 3D surface. This surface can be estimated using various functions such as polynomial based function or probabilistic based function such as Gaussian process (GP) model. Here we will investigate both of these function for the simulated cases and compare their fitting.

#### Polynomial surface fitting

A bivariate polynomial function with variables  $L$  and  $a$  was used as inputs which are crack length and depth, respectively. Consequently, the SIF value can be calculated as follows:

$$\widehat{SIF}(a, L) = \sum_{i=0}^{\deg_L} \sum_{j=0}^{\deg_a} c_{i,j} L^i a^j$$

Where  $\deg_a$  and  $\deg_L$  are the highest degree for each variable ( $L$  and  $a$ ) in the bivariate polynomial. Here we chose  $\deg_a = \deg_L = 2$ . In addition,  $c_{i,j}$  are constant coefficients of the polynomial that will be evaluated using linear optimization. As we have two sets of SIF this function needs to be determined twice: once as  $SIF_a$  and second time as  $SIF_L$ ; which are SIF values at the front of the crack along the length and SIF values at the front of the crack along the depth, respectively. See Fig. 8(c) for



**Fig. 9** a Stress contour at crack front for different crack lengths and depths. b Fitted surrogate functions using 42 data points for SIF along crack depth. c Fitted surrogate functions using 42 data points for SIF along crack length (Crack lengths are in mm and SIF in  $\text{MPa}\cdot\text{mm}^{-1}$ )

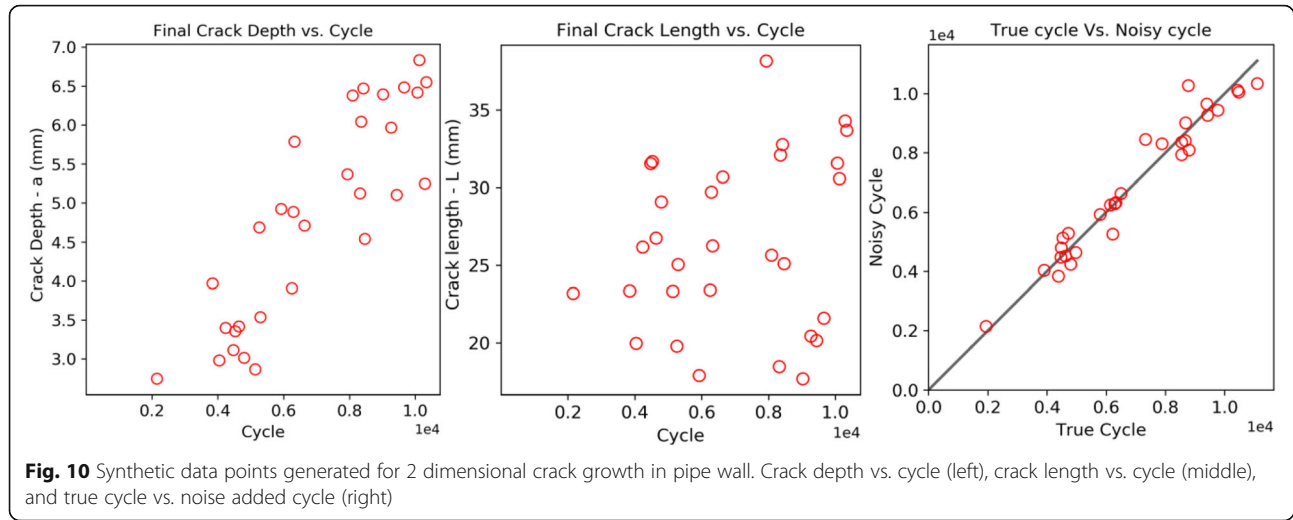
visualized locations of crack fronts annotated with  $SIF_a$  and  $SIF_L$ .

#### Gaussian process (GP) fitting

In the GP fitting method each SIF data point is fitted to a normal distribution and the expected value of the distribution is chosen as fitted value at that point.

Assuming input variables ( $L$  and  $a$ ) in form of a vector  $X_i = X_1, X_2, \dots, X_m$  for  $m$  data points, the output,  $\widehat{SIF}(a, L)$ , would be in form of  $Y(X_1), Y(X_2), \dots, Y(X_m)$ . At each non training point,  $X^*$ :

$$K(X_i, X_j) = \text{Cov}(f(X_i), f(X_j)) = e^{-\frac{1}{2\sigma^2}|X_i - X_j|^2}$$



$$\begin{aligned}\widehat{SIF}(a, L) &= Y^*(X^*) = \mathbb{E}[Y^*(X^*)|X^*, X, Y] \\ &= K(X^*, X)K(X, X)^{-1}Y\end{aligned}$$

$$\begin{aligned}\text{Var}[Y^*(X^*)|X^*, X, Y] &= K(X^*, X^*) \\ &\quad - K(X^*, X)K(X, X)^{-1}K(X, X^*)\end{aligned}$$

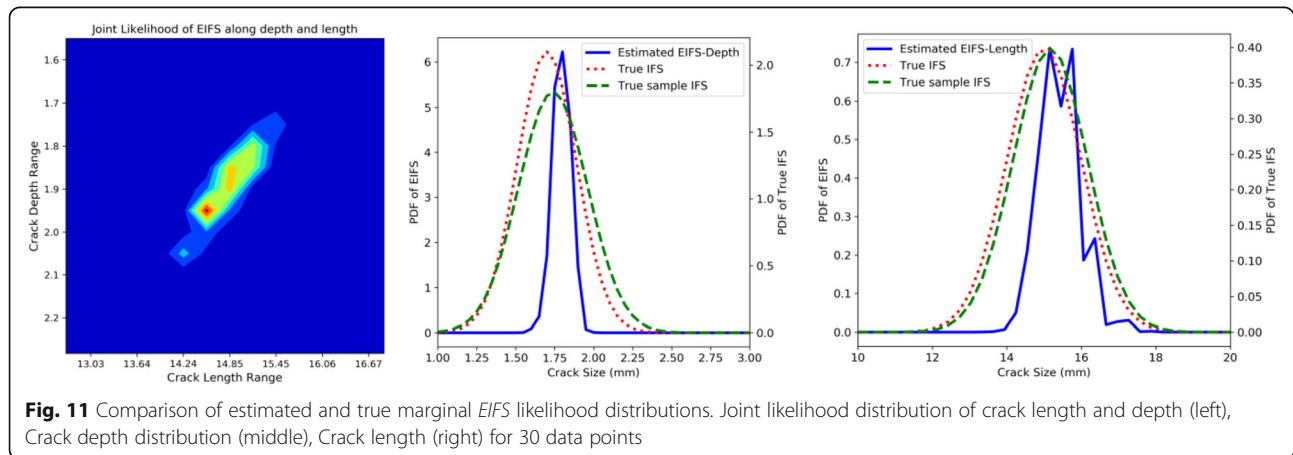
Where,  $K$  is the kernel or covariance function,  $f$  is the process function, and  $\ell$  is the characteristic length of the covariance function. In this study Radial-basis function (RBF) kernel (squared-exponential (SE) kernel) was used.

Too small  $\ell$  values cause oscillatory behavior between training data points as result of faster variations of the function. For details on Gaussian process implementation refer to [19, 20].

Totally 42 cases of FE simulations were conducted. These two surface fitting methods' accuracy were compared with different number of training points. Simulations were performed at 6 crack depths and 7 crack length, which in total yields 42 cases of simulation.

Figure 9 illustrates results of fitting surface to the FE simulations. In these figures the fitted surface is shown in blue and the all data points are shown in red circles.

Overall it can be said that both methods perform very good with predictions. But caution is needed when dealing with GP method as the fitting in this method is very sensitive to the length scale parameter ( $\ell$ ). In addition, GP method has a tendency of overfitting if not tuned well with a good covariance function (kernel) [20]. The polynomial based fitting method showed that it consistently gets better as more training data is included in the fitting process. It is worth noting that neither of these methods (especially GP method) are capable of having good extrapolated predictions so it is necessary to have as many boundary points as possible to make the fitted model more accurate. Nevertheless, both these methods were incorporated in the developed algorithm that computes the SIF values for predicting number of cycles in the crack growth model.





### Inference of EIFS for pipe

Applying the aforementioned method to crack growth in pipeline problem brings a few more challenges. The case for crack growth in an edge crack in steel plate only assumes crack growth in single dimension while the crack growth in pipe wall is assumed as a semi-elliptical shape that has two growth fronts; along the depth and along the length of the crack (minor and major axis of ellipse respectively). This property makes this problem similar to the problem of *EIFS* distribution with hyper-parameters (introduced in the section “*EIFS probability density inference with hyper parameters-  $p(\theta|\alpha)$* ”) which would require cubic likelihood array. Applying the 2-dimensional version of crack growth algorithm illustrated in Fig. 4 we will generate 30 data points as shown in Fig. 10. The finer the step size for cycles and mesh size for crack length and depth, the longer it takes for the code to yield the results.

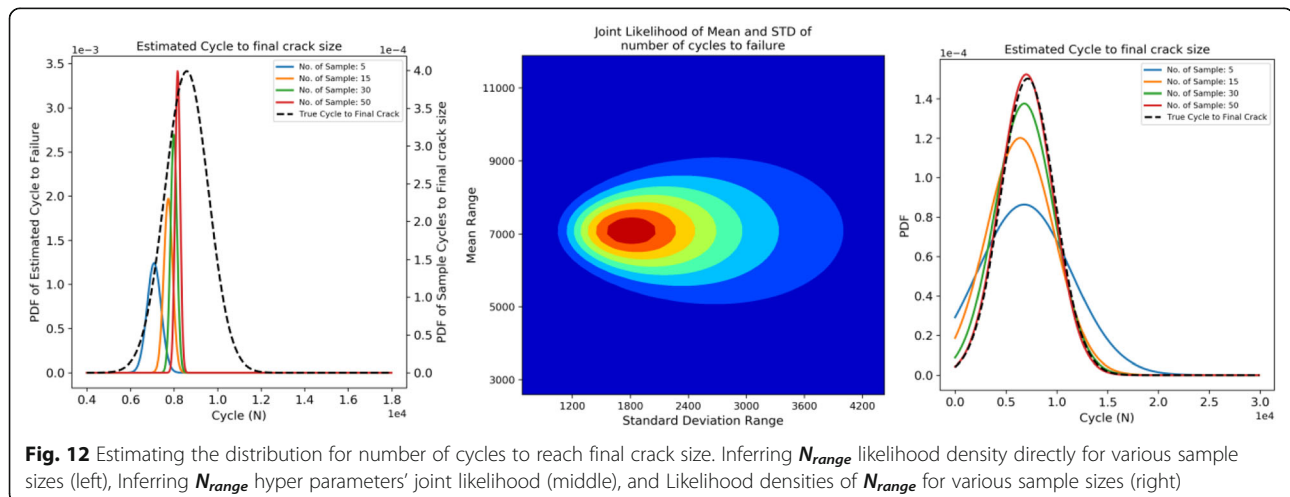
Consequently, there exists a likelihood distribution for every possible pair of crack length and crack depth. So, in the end we will find a joint likelihood distribution of *EIFS* for crack length and crack depth. The example of such joint likelihood distribution is shown in Fig. 11 (left). We can solve for marginal likelihoods of *EIFS* for crack length and crack depth by integrating out the other variable using Simpson's quadrature technique introduced earlier. The computed marginal likelihoods of crack length and depth are shown in Fig. 11 for crack depth (middle) and crack length (right). The results show that the algorithm has good accuracy in estimating (represented by blue curves) mean of the true distribution (represented by red as ground truth distribution and green as sample distribution). Similar to the plate problem, the standard deviation of crack depth and length is not well estimated in this method.

**Table 5** Estimated vs. true distribution values for *EIFS* along depth and length

<i>EIFS</i>	Mean		STD		
	True	Estimate	True	Estimate	Assumed
Depth	1.74	1.794	0.222	0.052	0.19
Length	15.166	15.429	1	0.555	1

To improve inference on standard deviation we could use hyper-parameter estimation as we did for plate problem. In comparison to the plate problem where crack growth was single dimensional, we only had to solve marginal likelihoods once for each data point. In the case of pipe crack growth given we consider hyper parameters (mean and standard deviation) for both crack depth and crack length, we will have 4 hyper parameters to estimate. Solving this problem with Simpson's quadrature will be extremely expensive. To solve this problem Monte Carlo simulations would work more efficiently as also mentioned in [1, 2, 4]. We chose to assume standard deviation for our problem in the section where we infer number of cycle to final crack size. This is because we did not perform hyper parameter estimates for mean and standard deviation of crack depth and length separately as we did in case of plate problem (see Fig. 3, and Table 2). Such computation can be performed using MCMC [21].

The properties of estimates for *EIFS* are summarized in Table 5. The estimated values of standard deviation in this table are not capturing the variability compared to true values. The assumed values were selected based on many observations and engineering judgments. Values between 0.1 to 0.5 mm for depth and 0.5 to 1.5 mm can be considered good guess range for standard deviations according to author's observations in other cases. We



**Table 6** Comparison of direct inference of  $N_{range}$  on different sample sizes

No of Samples	STD		Mean	
	True	Estimate	True	Estimate
5	836	321	8240	7080
15	898	202	8455	7732
30	1039	147	8626	7986
50	1000	116	8580	8165

will use these estimated values as input for the model to do inference of the likelihood distribution for the number of cycles to failure in the next section.

### Estimating probability and cycle of failure

Here we will follow the similar steps as we did for plate problem to estimate the distribution for the number of cycles to failure,  $N_{range}$ . Using the methodology discussed in section “[Estimating probability and cycle of failure](#)” for plate problem we will estimate the distribution of number of cycles to failure. As shown in Fig. 12 (top) the direct inference on  $N_{range}$  distribution is shown for various sample sizes. We can see the progressive approach of the predictions towards the true sample distribution from 5 samples (blue) to 50 samples (red). The ground truth of the sample distribution is shown in black dashed line. As with previous observation using direct inference we can see that while the standard deviation is not accurately captured the mean or expected value of number of cycles to final crack size is very well captured. The summary of data for this model is presented in Table 6.

Using hyper parameters to perform estimations we will have the joint marginal likelihood distribution of mean and standard deviation of number of cycles to failure, shown in Fig. 12 (bottom left).

Integrating along mean and standard deviation separately we have marginal distribution for mean and standard deviation of number of cycles to failure separately. Using the expected value of the latter distributions we construct the distribution for number of cycles to failure. These distributions for various number data points or samples are shown in Fig. 12 (bottom right). We can see that as more data points are

**Table 7** Comparison of inference of  $N_{range}$  with hyper parameters on different sample sizes

No of Samples	STD		Mean	
	True	Estimate	True	Estimate
5	3885	4620	6966	6800
15	3195	3320	6363	6400
30	2876	2900	6918	6800
50	2657	2620	7136	7000

added to the model, from samples (blue curve) to 50 samples (red curve), the estimated distribution very closely matches the true sample distribution shown in black dashed line. The results for various sample sizes are summarized in Table 7.

### Conclusions and discussions

In this paper a Bayesian inference methodology was implemented to accurately estimate the time (cycle) when the structure (pipe or plate) has the most probability of failure based on observed crack growth measurements and cycle data (equivalent to field inspection) which was generated synthetically. A methodology to estimate Equivalent Initial Flaw Size (EIFS) was introduced and the distribution for number of cycles to failure was predicted.

A base model was initially developed based on edge crack growth in a steel plate and verified the methodology. Then the method was expanded to model predictions for two-dimensional crack growth in pipe wall thickness. Finite Element (FE) Modeling was employed to calculate stress intensity factor (SIF) at finite pints of crack length and depth combinations. Two surrogate models were used to interpolate the SIF values at the points in which FE simulations were not conducted. The Bayesian inference was performed with and without hyper parameters and the result demonstrated the gain in accuracy with use of hyperparameters.

Comparison of estimation and true values showed that the proposed methodology has strong grounds for accurately predicting most critical time (cycle) when the structure may become susceptible to failure as more inspection data is collected and input into the model. The model can be further customized and more variables can be inferred simultaneously. Consequently, the more variables of the model to be inferred, the more complex and certainly more computationally expensive it becomes.

### Abbreviations

3D: three-dimensional; DBN: dynamic Bayesian networks; EIFS: equivalent initial flaw size; FE: finite element; GP: Gaussian process; IFS: initial flaw size; MCMC: Markov Chain Monte Carlo; NDE: non-destructive evaluation; PDF: probability density function; RFL: remaining fatigue life; SIF: stress intensity factor; S-N: strain-life

### Acknowledgements

The authors acknowledge the financial support provided by the USDOT Pipeline and Hazardous Materials Safety Administration (PHMSA).

### Authors' contributions

MS conducted the analysis and wrote the draft. HW developed the study plan and revised the draft. The authors read and approved the final manuscript.

### Funding

USDOT PHMSA Grant DTPH5615HCAP04L.

### Availability of data and materials

All data used or generated by this study is available from the corresponding author by reasonable request.



**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 19 December 2019 Accepted: 24 February 2020

Published online: 23 March 2020

**References**

1. Cross R, Makeev A, Armanios E (2007) Simultaneous uncertainty quantification of fracture mechanics based life prediction model parameters. *Int J Fatigue* 29(8):1510–1515
2. Liu Y, Mahadevan S (2009) Probabilistic fatigue life prediction using an equivalent initial flaw size distribution. *Int J Fatigue* 31(3):476–487
3. Makeev A, Nikishkov Y, Armanios E (2007) A concept for quantifying equivalent initial flaw size distribution in fracture mechanics based life prediction models. *Int J Fatigue* 29(1):141–145
4. Sankararaman S, Ling Y, Mahadevan S (2010) Statistical inference of equivalent initial flaw size with complicated structural geometry and multi-axial variable amplitude loading. *Int J Fatigue* 32(10):1689–1700
5. Sankararaman S, Ling Y, Mahadevan S (2011) Uncertainty quantification and model validation of fatigue crack growth prediction. *Eng Fract Mech* 78(7):1487–1504
6. Zarate BA, Caicedo JM, Yu J, Ziehl P (2012) Bayesian model updating and prognosis of fatigue crack growth. *Eng Struct* 45:53–61
7. Fawaz SA (2003) Equivalent initial flaw size testing and analysis of transport aircraft skin splices. *Fatigue Fracture Eng Mater Struct* 26(3):279–290
8. Xie M, Bott S, Sutton A, Nemeth A, Tian Z (2018) An integrated prognostics approach for pipeline fatigue crack growth prediction utilizing inline inspection data. *J Press Vessel Technol* 140(3):031702
9. Gobbato M, Kosmatka JB, Conte JP (2014) A recursive Bayesian approach for fatigue damage prognosis: an experimental validation at the reliability component level. *Mech Syst Signal Process* 45(2):448–467
10. Ribeiro T, Borges L, Rigueiro C (2019) A case study on the use of Bayesian inference in fracture mechanics models for inspection planning. *J Fail Anal Prev* 19(4):1043–1054
11. Babuška I, Sawlan Z, Scavino M, Szabó B, Tempone R (2016) Bayesian inference and model comparison for metallic fatigue data. *Comput Methods Appl Mech Eng* 304:171–196
12. Arzaghi E, Abaei MM, Abbassi R, Garaniya V, Chin C, Khan F (2017) Risk-based maintenance planning of subsea pipelines through fatigue crack growth monitoring. *Eng Fail Anal* 79:928–939
13. Li C, Mahadevan S, Ling Y, Choe S, Wang L (2017) Dynamic Bayesian network for aircraft wing health monitoring digital twin. *AIAA J* 55(3):930–941
14. Tada, H., Paris, P. C., & Irwin, G. R. (1973). *The stress analysis of cracks*. Handbook, Del Research Corporation
15. Paris P, Erdogan F (1963) A critical analysis of crack propagation laws
16. Atkinson KE (2008) *An introduction to numerical analysis*. Wiley, Hoboken, p 203
17. Newman JC Jr, Raju IS (1984a) Prediction of fatigue crack-growth patterns and lives in three-dimensional cracked bodies. In: *Proceedings of the 6th International Conference on Fracture (ICF6)*, New Delhi, India, pp 1597–1608
18. Newman JC Jr, Raju IS (1984b) Stress-intensity factor equations for cracks in three-dimensional finite bodies subjected to tension and bending loads
19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12(Oct): 2825–2830
20. Williams CK, Rasmussen CE (1996) Gaussian processes for regression. In: *Advances in neural information processing systems*, pp 514–520
21. Geyer CJ (1992) Practical markov chain Monte Carlo. *Stat Sci* 7(4):473–483

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)