ESTIMATION OF STRESS INTENSITY FACTOR FOR SURFACE CRACKS IN THE FIRTREE GROOVE STRUCTURE OF A TURBINE DISK USING POOL-BASED ACTIVE LEARNING WITH GAUSSIAN PROCESS REGRESSION

KAIMIN GUO

Research Institute of Aero-Engine, Beihang University, Beijing, China

HONGZHUO LIU, HAN YAN, ZIYUAN SONG, SHENGMING ZHANG

School of Energy and Power Engineering, Beihang University, Beijing, China

DAWEI HUANG

School of Energy and Power Engineering, Beihang University, Beijing, China, and

Beijing Key Laboratory of Aero-Engine Structure and Strength, Beijing, China

e-mail: huangdawei@buaa.edu.cn

XIAOJUN YAN

School of Energy and Power Engineering, Beihang University, Beijing, China, and

Beijing Key Laboratory of Aero-Engine Structure and Strength, Beijing, China, and

National Key Laboratory of Science and Technology on Aero-Engine Aero-thermodynamics, Beijing, China

e-mail: yanxiaojun@buaa.edu.cn

Calculation of the stress intensity factor K is a crucial and difficult task in linear elastic fracture mechanics. With the capacity to solve complex input-output problems of an underlying system, machine learning is especially useful in the calculation of K. However, when faced with complex systems, such as the firtree groove structure of a turbine disk, the data-consuming issue has always been a thorny problem in K-solutions combined with machine learning studies for a long time. In this paper, a novel K-solution method called PA-GPR (Pool-based Active learning with Gaussian Process Regression) for the calculation of the stress intensity factor for surface cracks in the firtree groove structure of a turbine disk is proposed. Using the pool-based active learning strategy, the proposed K-solution method could make the GPR model have a great regression performance with a few samples required. In the pool-based active learning strategy analysis, the learning function based on greedy sampling is proposed to select samples with a high contribution to the training of the GPR model. The calculation of K for a semi-elliptical surface crack in the firtree groove structure is evaluated to verify the accuracy and effectiveness of the proposed method. The results show that this novel method is accurate, time-saving and effective.

Keywords: damage tolerance, stress intensity factor solutions, machine learning, active learning

1. Introduction

Probabilistic damage tolerance analysis is essential for the assessment of integrity and reliability in practical engineering applications such as high-temperature components in aero engines (Basista and Węglewski, 2006; Huang *et al.*, 2018, 2019; P. Li *et al.*, 2018; Y. Li *et al.*, 2019; Rinaldi *et al.*, 2006), high-speed railway components, etc. In the probabilistic damage tolerance assessment of a turbine disk, it is necessary to carry out crack growth calculations in locally complex structures (such as the firtree groove structure, etc.) many times (up to 10^6) under

the assumption of the existence of an initial defect or crack, the convergence failure risk can be obtained by a statistical method.

The firtree groove structure is one of the most critical and complicated regions of a turbine disk (Yuan *et al.*, 2021). In the aspect of structure, the firtree groove structure is composed of many pairs of tenons with complex geometric characters. Meanwhile, the firtree groove structure bears many complicated loads, such as the centrifugal force of the blade, aerodynamic bending moment, thermal load and vibration load (Yang *et al.*, 2017). The complexity of geometry and load makes the firtree groove structure one of the most vulnerable parts of the turbine disk.

Consequently, due to the complexity of geometry and load of the firtree groove structure, it is difficult to evaluate the crack growth, especially the stress intensity factor K for (Cui and Wang, 2011; Witek, 2012).

Up to now, lots of K-solution methods have been developed for cracks in the firtree groove structure of a turbine disk, such as the finite element method (FEM) (Boulenouar *et al.*, 2014; Huang *et al.*, 2021; Moustabchir *et al.*, 2015, 2017; Shlyannikov *et al.*, 2016) and machine learning (ML). The finite element method usually becomes a reliable method to generate accurate K solutions. However, the accuracy of the finite element method is greatly influenced by the number of mesh elements of the model. When faced with a complex structure like the firtree groove structure, this number must be large. That means, consequently, a high computational cost of the finite element method.

Recently, K-solutions combined with ML have become a topic of growing interest (Keprate *et al.*, 2017; Liu *et al.*, 2020; Muñoz-Abella *et al.*, 2015; Xu *et al.*, 2021). With the capacity to solve complex input-output problems of an underlying system, ML is especially useful in the calculation of K. These solutions proceed in three steps (in this paper, the method that follows these steps is called the traditional ML method): (1) calculating the limit number of K data as the training data set; (2) training an ML model using training the data set and (3) predicting values of K using the trained ML model.

However, the data-consuming issue has always been a thorny problem in the K-solutions combined with ML studies for a long time. For regression problems, the generalization ability of a ML model is always related to a sample size of the training data set. A large quantity of data is usually required to get accurate results in the calculation of K. This problem is particularly acute in the face of complex structures, such as turbine disk grooves, etc. In some studies, the sample size of the training data set has even reached 10^6.

The active learning strategy is an efficient solution to this data-consuming problem. This strategy first uses the initial training data set T to train the regression ML model. Then the initial training data set T is enriched by adding new sample points based on the defined learning function. And finally, a new surrogate model is retained using the updated T, then the above-mentioned steps are repeated till the convergence criterion is satisfied.

This study presents a novel K-solution method called PA-GPR (Pool-based Active learning with Gaussian Process Regression) for the calculation of the stress intensity factor for surface cracks in the firtree groove structure of a turbine disk. The paper is organized as follows. Firstly, some problems of K-calculation for surface cracks in the firtree groove structure will be described. Then, details and principles of the proposed PA-GPR are elaborated, including the learning function and the convergence criterion. To demonstrate the accuracy and effectiveness of the proposed method, a calculation of K case is presented, i.e., calculation of K for a semi-elliptical surface crack in the firtree groove structure of a turbine disk. Furthermore, the mechanism of the learning function of active learning is discussed.

2.   Problem description

The firtree groove structure is one of the most critical and complicated regions of turbine disks. As shown in Fig. 1, the geometry features include the number of firtree teeth n, flank length l, contact angle α, flank angles β and γ, outer radius R_0 and inner radius R_1 (Meguid *et al.*, 2000). Moreover, different kinds of loads including the centrifugal force, the thermal load, vibration load are applied to the firtree groove structure. The intricate structure and load conditions make the stress and fracture analysis a hard task. Figure 2 shows a typical stress distribution of the firtree groove structure.
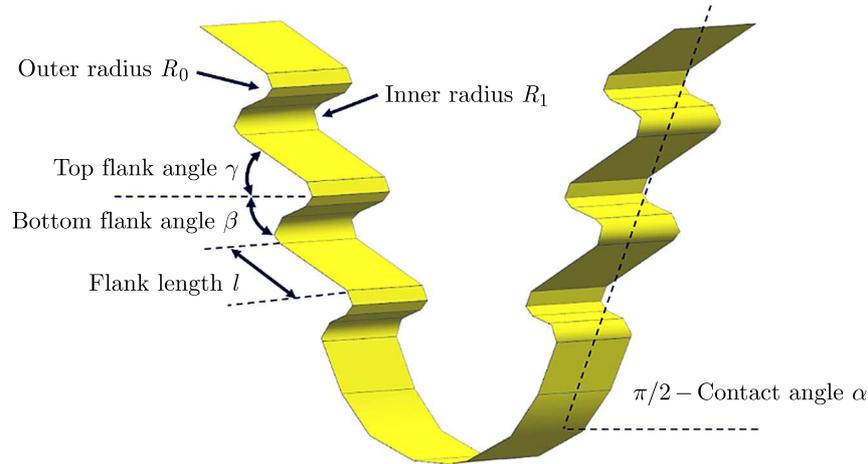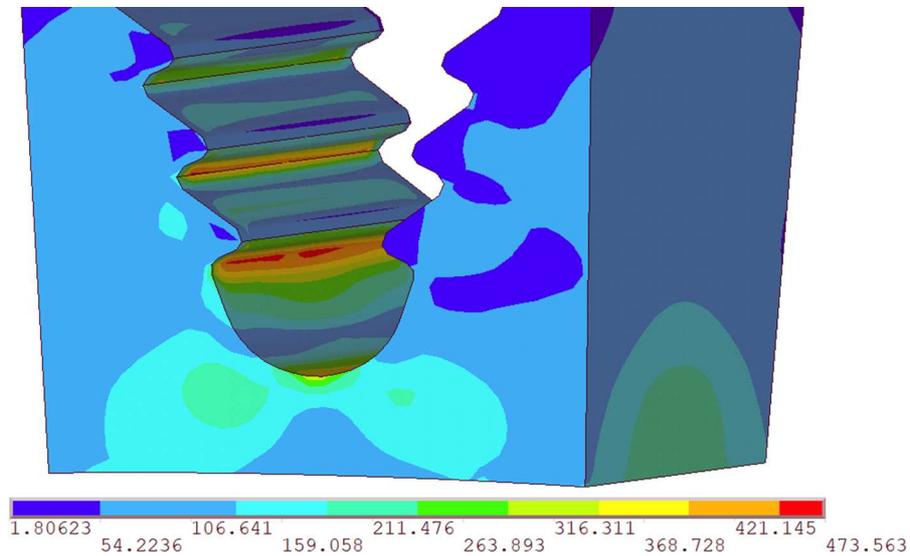
Fig. 1. Schematic of the firtree groove structure

Fig. 2. An example of von Mises stress distribution of the firtree groove structure (the values on the scale are von Mises stress values, unit: MPa)

Cracks may occur at the firtree groove structure (especially at the bottom) due to cyclic loading or initial defects. Considering the requirement of the turbine disk failure risk, it is necessary to develop an efficient and rapid method for calculation of the K for a large number of repeated crack growth evaluations.

The effective K-solution method proposed in this paper focuses on the calculation of the stress intensity factor for surface cracks in the firtree groove structure. As shown in Fig. 3,

a semi-elliptical surface crack is located at the transition arc between the fifth teeth and the bottom of the groove. It is assumed that the leading edge of the crack keeps an elliptical shape during the propagation process. Key dimensional parameters of the crack are $\{a, c, \xi\}$, where a and c mean the depth and half-length of the surface crack, respectively. The ξ denotes the position of point P on the edge of the crack, $\xi = |AC|/|AB|$, as shown in Fig. 4.
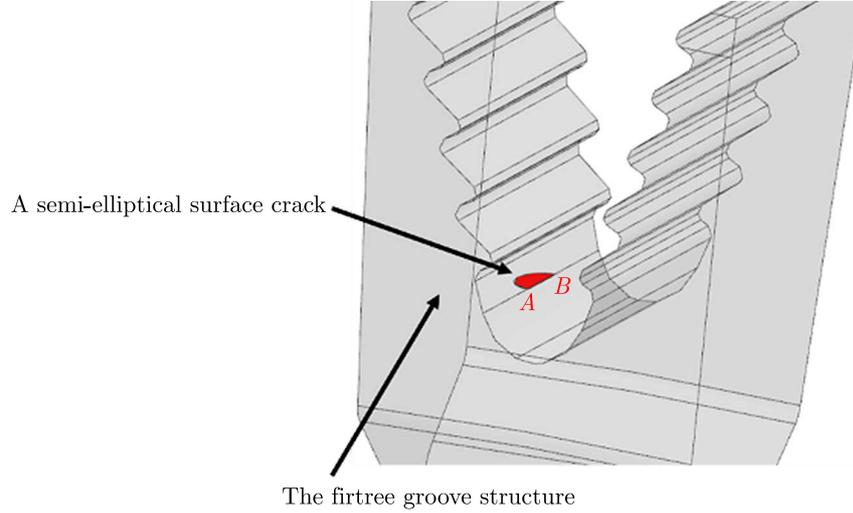

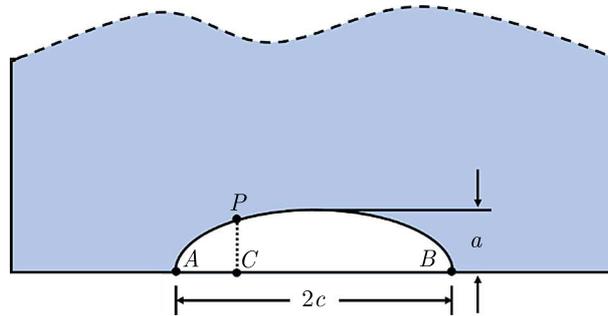Fig. 3. A semi-elliptical surface crack in the firtree groove structure of a turbine disk

Fig. 4. The geometry configuration of a semi-elliptical surface crack in the firtree groove structure of a turbine disk

For a particular type of the firtree groove structure, the stress distribution and shape of the crack are certain, but the initial crack size is random. That means the dimensional parameters $\{a, c\}$ are random variables. Meanwhile, in order to accurately evaluate the value of K, it is necessary to calculate K at any point (point P as shown in Fig. 4) along the crack front, i.e., $0 \leqslant \xi \leqslant 1$.

3. Pool-based Active learning with Gaussian Process Regression (PA-GPR)

The proposed PA-GPR method includes two critical parts, the GPR algorithm and the pool--based active learning strategy. The GPR algorithm is adopted for the prediction of K. The pool--based active learning strategy is used in the training progress of the GPR model, making the trained GPR model have a great generalization ability with only a few training samples required.

3.1. PA-GPR for calculation of K

The Gaussian process regression, belonging to supervised learning, is a non-parametric model for regression analysis of data using Gaussian processes (Rasmussen, 2003). It can be written as

$$f \sim GP(m(x), k(x, x')) \tag{3.1}$$

where $m(x)$ and $k(x, x')$ represent the mean and covariance function, respectively. Equation (3.1) means the function f is distributed as a Gaussian process with the mean function m and covariance function k. More details could be seen in the literature (Rasmussen, 2003).

The flow diagram of the proposed PA-GPR method is shown in Fig. 5, and it is based on the following steps:

(1) Generate a candidate sample pool $\mathbf{S}$ including N_S samples. In the calculation process of K, the input variables are parameters that affect the stress intensity factor results, such as crack depth and dimensions of the cracked structure. The candidate sample pool $\mathbf{S}$ is generated by random sampling based on the distribution or range of the input variables.

(2) $\mathbf{S} = \mathbf{L} \cup \mathbf{U}$, $\mathbf{L} \cap \mathbf{U} = \emptyset$. $\mathbf{L}$ denotes the labeled data set and $\mathbf{U}$ is the unlabeled data set. The labeled data set $\mathbf{L}$ consists of N samples picked from $\mathbf{S}$. $\mathbf{L} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$. Compared with the sample size of $\mathbf{S}$, N is typically several orders of magnitude smaller.

(3) After definition of the labeled data set $\mathbf{L}$, the corresponding response $\mathbf{Y}$ should be computed, i.e., $\mathbf{Y} = \{K_1, K_2, \dots, K_N\}$. Then a training data set $\mathbf{T}$ should be defined, $\mathbf{T} = \{\mathbf{L}, \mathbf{Y}\}$.

(4) Train the GPR model using $\mathbf{T}$.

(5) If the convergence criterion (mentioned in Section 3.3) is not satisfied, the training data set $\mathbf{T}$ should be enriched by the learning function (mentioned in Section 3.2). In detail, the new sample $\mathbf{X}_{new}$ could be selected from the unlabeled data set $\mathbf{U}$ using the learning function. And the corresponding response K_{new} should be calculated. Then the training data set $\mathbf{T}$ becomes the new training data set $\mathbf{T}_{new}$ by adding a new sample, i.e., $\mathbf{T}_{new} = \mathbf{T} \cup \{\mathbf{X}_{new}, K_{new}\}$.

(6) Retrain the GPR model using $\mathbf{T}_{new}$. Repeat steps (4) to (5) until the convergence criterion is satisfied.

3.2. Learning function

Different sample points in the same sample pool have different degrees of contribution to training of the GPR model. In the traditional machine learning method for regression, there are often spatial overlaps or approximate overlaps among the sample points in the training data set, which makes these samples a poor contribution to training of the GPR model. On the contrary, if the selected samples can avoid spatial overlaps or approximate overlaps situations, these samples will make a high contribution to training of the GPR model. A more detailed explanation could be seen in Section 5. Consequently, the GPR model could have a great generalization ability with only a few samples required, as long as samples with great contributions are selected by the proposed learning function from the unlabeled data set $\mathbf{U}$.

Inspired by the greedy sampling (GS) strategy, this paper proposes a GS-based learning function to pick key samples with great contributions to training of the GPR model (Wu *et al.*, 2019; Yu and Kim, 2010). In different application situations, two different learning functions, i.e., GS on the inputs and GS on the outputs are proposed.

3.2.1. GS on the inputs

To achieve diversity in the input space, GS on the inputs could be elaborated as follows.
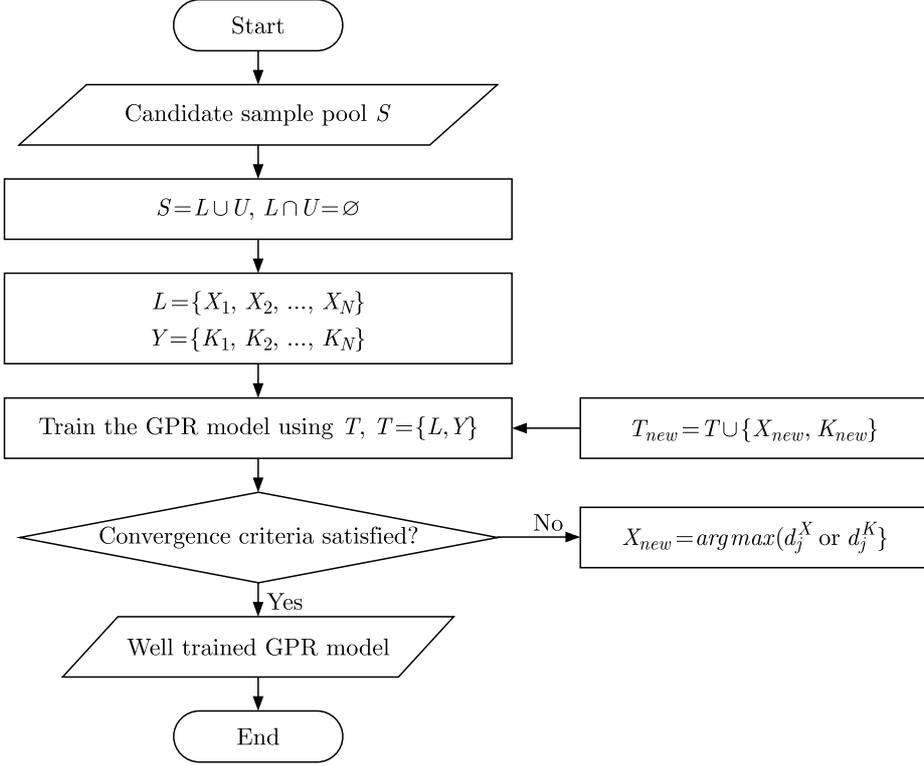
Fig. 5. Flowchart of the proposed PA-GPR

It is assumed that the labeled data set $\mathbf{L}$ consists of N samples $\mathbf{X}_i$ $(i = 1, 2, \ldots, N)$ picked from $\mathbf{S}$. For each of the remaining $N_S - N$ samples $\mathbf{X}_j$ $(j = N+1, N+2, \ldots, N_S)$, its distance to each of the labeled data set $\mathbf{L}$ could be calculated as

$$d_{ij}^X = \|X_i - X_j\|_2 \tag{3.2}$$

where $i = 1, 2, \ldots, N$, $j = N+1, N+2, \ldots, N_S$. Then the shortest distance from the j-th sample $\mathbf{X}_j$ to N labeled samples, $\mathbf{d}_j$ could be written as

$$d_j^X = \min(d_{ij}^X) \tag{3.3}$$

where $j = N+1, N+2, \ldots, N_S$.

The learning function based on GS on the inputs is proposed as

$$X_{new} = \arg\max(d_j^X) \tag{3.4}$$

the new sample point ($\mathbf{X}_{new}$, to be added in $\mathbf{T}$) could be chosen from $\mathbf{U}$ by this learning function.

3.2.2. *GS on the outputs*

In some application situations, the learning function based on GS on the outputs is more effective than GS on the inputs. Similar to the learning function based on GS on the inputs, details of the learning function based on GS on the outputs could be explained as follows.

Similarly, it is assumed that the training data set $\mathbf{T}$ consists of N samples $\{\mathbf{X}_i, K_i\}$ $(i = 1, 2, \ldots, N)$. The trained GPR model could predict the value of the corresponding response for each of the remaining $N_S - N$ samples as

$$\widehat{K_j} = f(X_j) \tag{3.5}$$

where $j = N+1, N+2, \ldots, N_S$.

Then the distance on the outputs from each of the remaining $N_S - N$ samples $\mathbf{X}_j$ ($j = N + 1, N + 2, \ldots, N_S$) to each of the labeled data set $\mathbf{L}$ could be calculated by

$$d_{ij}^K = \|K_i - \widehat{K}_j\|_2 \tag{3.6}$$

where $i = 1, 2, \ldots, N$, $j = N + 1, N + 2, \ldots, N_S$. Then, the shortest distance on the outputs from the j-th sample X_j to N labeled samples, $\mathbf{d}_j$, could be written as

$$d_j^K = \min(d_{ij}^K) \tag{3.7}$$

where $j = N + 1, N + 2, \ldots, N_S$.

The learning function based on GS on the outputs is proposed as

$$X_{new} = \arg\max(d_j^K) \tag{3.8}$$

the new sample point ($\mathbf{X}_{new}$, to be added in $\mathbf{T}$) could be chosen from $\mathbf{U}$ by this learning function.

For GS on the inputs or outputs, it is difficult to determine which method is more effective. For validation cases in this research, two learning functions are both applied to find the learning function with better performance.

3.3. Convergence criterion

In this study, N_V labeled data (not belonging to the labeled data set $\mathbf{L}$) are randomly selected to make up the validation set. The GPR model has a good generalization ability when the mean relative validation error reaches convergence, which could be written as

$$\frac{1}{N_V} \sum_{p=1}^{N_V} \frac{|\widehat{K}_p - K_p|}{K_p} \leqslant \varepsilon \tag{3.9}$$

where ε means the convergence threshold.

4. K-solution using PA-GPR

In this Section, the PA-GPR method is applied to the calculation of K for a semi-elliptical surface crack in a turbine disk firtree groove structure. The prediction accuracy of the PA-GPR is firstly discussed. Then, two different K-solution methods, PA-GPR and the traditional ML method are compared in order to evaluate which one requires fewer samples. Here, the traditional ML model is also GPR, where the training data set is generated by random sampling from the candidate sample pool $\mathbf{S}$.

According to the surface crack in the firtree groove structure described in Section 2, the input variables are $\mathbf{X} = (a, c, \xi)$, and the output variable is the stress intensity factor K. Each of the input variable is defined as

$$c, a \in [0.1\,\mathrm{mm}, 5\,\mathrm{mm}] \qquad \frac{a}{c} \in [0.1, 1] \qquad \xi \in [0, 1] \tag{4.1}$$

$10^5 K$ samples for surface cracks in the firtree groove structure subjected to the centrifugal force and temperature are calculated using the professional fracture analysis software FRANC3D to make up the candidate sample pool $\mathbf{S}$. The FEM crack mesh generated by FRANC3D is shown in Fig. 6. In addition, $N_V = 200$, $\varepsilon = 1.62\%$. The mean relative validation error of this ML model is 1.62%. So, the convergence threshold ε value is set to 1.62%, to evaluate the effectiveness of the PA-GPR by comparing the sample size with the identical mean relative validation error. The initial labeled data set consists of 100 samples. After the application of two learning functions
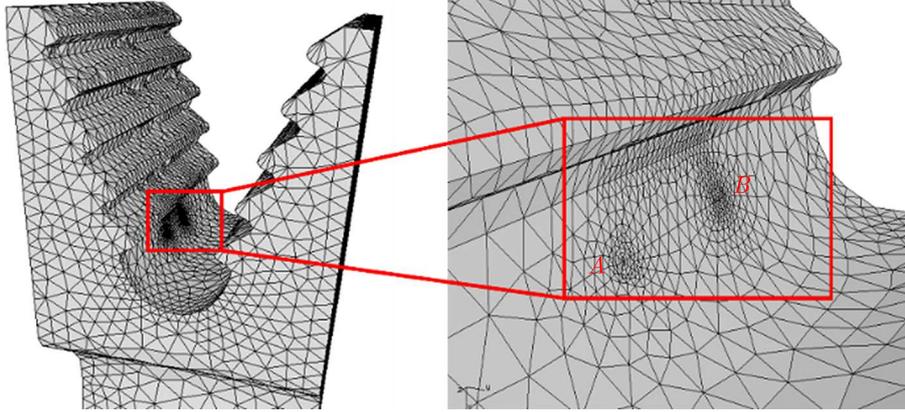
Fig. 6. The FEM crack mesh generated by FRANC3D

(GS on the outputs or inputs) in this validation case, it is found that the sample size is smaller when using the learning function based on GS on the inputs. Consequently, the learning function based on GS on the inputs is adopted.

The plot of the predicted K using PA-GPR vs. true K calculated by FRANC3D is shown in Fig. 7. It is shown that the 200 observation points are sited around the perfect prediction line. The K results with different dimensional parameters calculated by the proposed PA-GPR are compared with those calculated by FRANC3D (as shown in Fig. 8). It can be observed that the K result driven by PA-GPR is very close to the result calculated by FRANC3D.
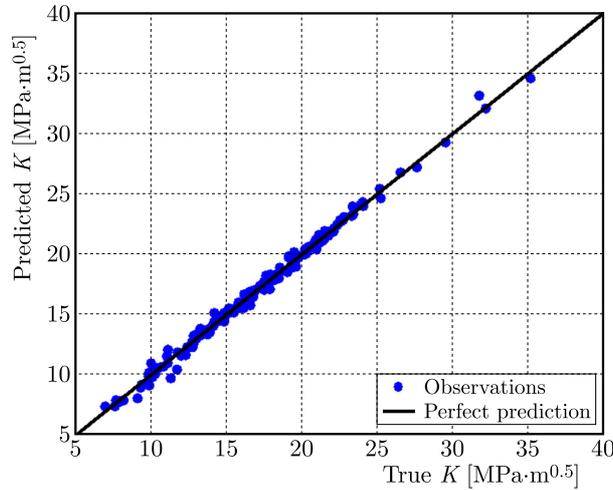

Fig. 7. Predicted K using PA-GPR vs. true K (calculated by FRANC3D) for a semi-elliptical surface crack in a turbine disk firtree groove

When the convergence is reached, the numbers of required samples of PA-GPR and the traditional ML method are shown in Table 1. That means, the proposed PA-GPR is two orders of magnitude more efficient than the traditional ML method.

It should be noted that the proposed PA-GPR adds time for sample selection. However, compared to the time to get the label of the reduced samples (calculating K by the finite element method or other K-solution method), the time for sample selection is negligible.
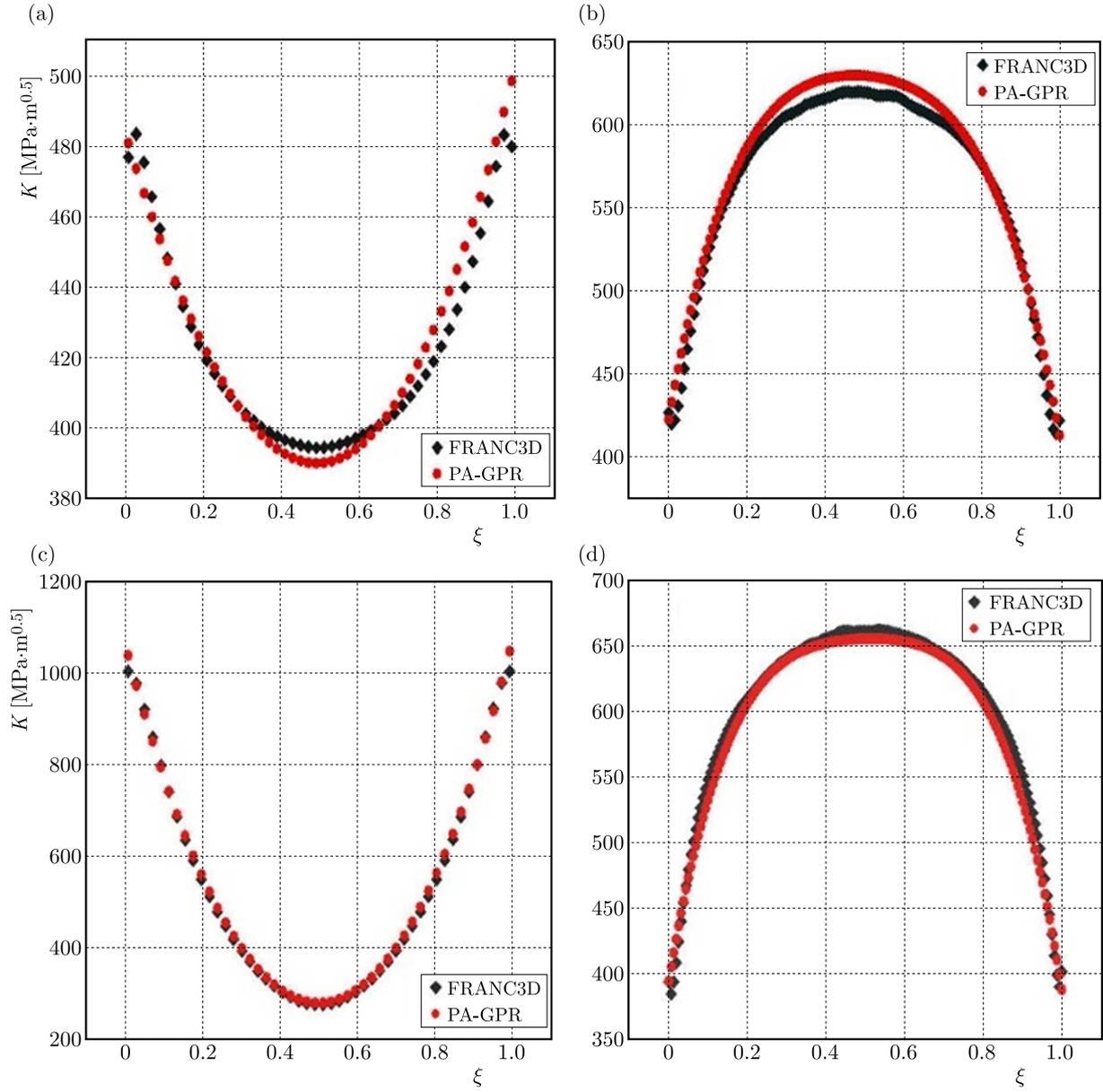
Fig. 8. *K* results with different dimensional parameters: (a) $a = 0.53$ mm, $c = 0.58$ mm, (b) $a = 0.65$ mm, $c = 2.49$ mm, (c) $a = 3.42$ mm, $c = 3.47$ mm and (d) $a = 0.65$ mm, $c = 3.31$ mm

Table 1. The sample size required for PA-GPR and the supervised ML method

	Traditional ML method	PA-GPR
The number of required samples	5E4	1135 (including initial 100 samples)

5. Discussion

The ML-based method becomes an effective stress intensity factor solution in the recent linear elastic fracture mechanics studies. However, for traditional ML regression problems, a good regression performance always means a lot of labeled training samples. That causes a time--consuming issue.

In traditional machine learning for regression, there are often spatial overlaps or approximate overlaps between the sample points in the training data set. For example, there are two sample points, $(a, c, \xi, K) = (3.3738, 1.5372, 0.1961, 21.1314)$ and $(3.4334, 1.3629, 0.8013, 21.1316)$ in the training data set for the traditional ML model in Section 4. These two sample points approximately overlap in Euclidean space, which makes the contribution of these two sample points to training of the model equal to the contribution of only one sample point. This phenomenon results in waste of sample points. Further, spatial overlaps or approximate overlaps of many sample points make the training of the regression model time-consuming.

As mentioned in this paper, the pool-based active learning strategy could solve this time-consuming problem by selecting specific samples using the learning function. The purpose of GS is to achieve diversity of sample space (inputs space or outputs space). Whether to choose GS on the inputs or outputs space depends on a specific application situation.

Diversity of the sample space could avoid the waste of sample points. This is also the main idea of the learning function based on GS. Specifically, the learning function based on GS could pick only one sample point among samples with spatial overlaps or approximate overlaps situations into the training data set. When the action of picking is done, sample points (picked by the learning function) of the training data set could cover the range of the independent (or dependent) variables. That achieves diversity of the sample space using a few sample points. Further, the regression model would have a great regression performance using a few sample points.
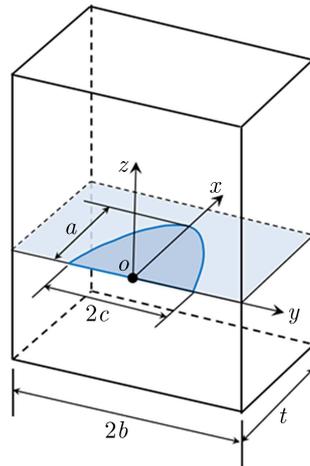

Fig. 9. The geometry configuration of a semi-elliptical surface crack in a finite plate

A calculation case is evaluated to visualize the diversity of the inputs space and the mechanism of the learning function based on GS. In this case, the calculation of K for a semi-elliptical surface crack in a finite plate is evaluated. The data for this validation case were obtained from the literature (Newman and Raju, 1981). The geometry configuration of a semi-elliptical surface crack in a finite plate is shown in Fig. 9, where the plate width is $2b = 50$ mm, and the plate thickness is $t = 10$ mm. The depth and half-length of the surface crack are represented by a and c, respectively. The plate is subjected to a remote uniform tension load with the stress ratio $R = 0$. This uniform tension load could be characterized by $\sigma = 800$ MPa. The input variables are $\mathbf{X} = (a, c)$, and the output variable is the stress intensity factor K. Each of the input variables is defined as

$$\ln a \simeq N(-1.346, 0.49) \qquad \frac{a}{c} \in [0.2, 1] \tag{5.1}$$

where the depth of the surface crack a is assumed to fit lognormal distribution (Liu and Mahadevan, 2009), and its mean value and standard deviation are 0.33 mm and 0.26 mm, respectively. 10^5 samples are randomly sampled to make up the candidate sample pool $\mathbf{S}$ based on the distri-

bution or range of input variables. as shown in Eq. (5.1). In detail, for a sample $\mathbf{X}_0 = (a_0, c_0)$, a_0 is sampled based on the lognormal distribution in Eq. (5.1) using the Box Muller method. Then, values of a_0/c_0 could be sampled by a uniform sampling method based on the distribution in Eq. (5.1).

100 samples are selected by the learning function based on GS on the inputs and randomly sampled, as shown in Fig. 10a and 10b, respectively. Meanwhile, a candidate sample pool including 10^4 samples is given in Fig. 10c, which shows the range of inputs space as a control.

As Fig. 10b shows, there are many sample points with spatial overlaps or approximate overlaps situations clustered in the dashed red line ellipse, while there are a few sample points in the dashed blue line circle. If the regression model is trained using these 100 samples, the imbalance situation will inevitably cause a bad regression performance.

Compared with Fig. 10b, Fig. 10a shows a good distribution of sample points in the sample space. Sample points, selected by the learning function, are evenly distributed in the sample space. That means the diversity of the sample space is achieved. If the regression model is trained using these 100 samples selected by the learning function, the diversity of the sample space would lead to a great regression performance using much fewer samples.
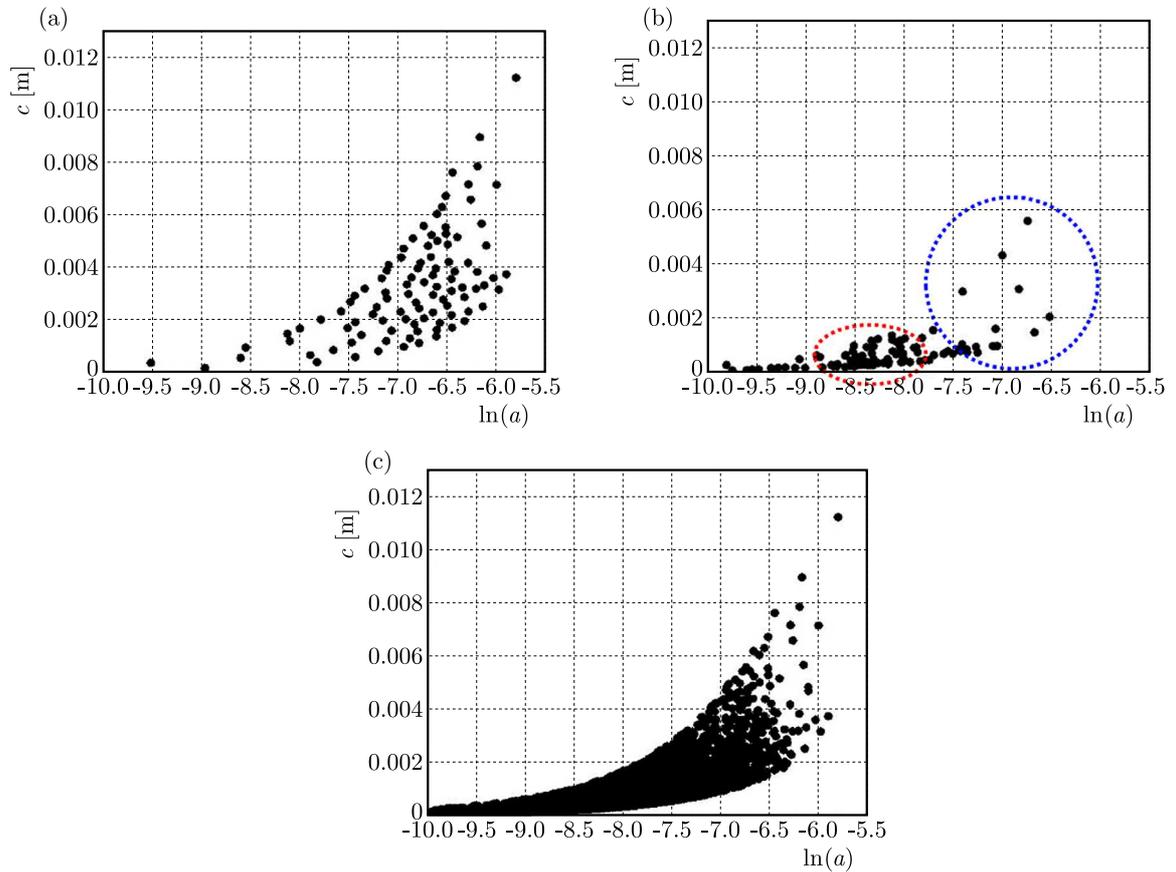

Fig. 10. Distribution of sample points in the inputs space: (a) samples selected by the learning function based on GS, (b) samples selected by randomly sampling and (c) range of the inputs space

6. Conclusion

Based on this study, three conclusions can be drawn:

- A novel K-solution method for surface cracks in the firtree groove structure of a turbine disk called PA-GPR (Pool-based Active learning with Gaussian Process Regression) is pro-

posed in this paper. Combining active learning strategy and Gaussian Process Regression (GPR), this novel method is accurate, time-saving and effective.

- Using the active learning strategy, the trained GPR model could have a great computational accuracy with only a few training samples required. For K calculation problems of surface cracks in the firtree groove structure of a turbine disk, the proposed PA-GPR is two orders of magnitude more efficient than the traditional ML method with almost identical K results.

- The learning function based on GS, the core of the proposed method, could select samples with a high contribution to the training of the GPR model by the achievement of diversity of the sample space. Focusing on spatial overlaps or approximate overlaps situations, the learning function picks only one sample point among samples with spatial overlaps or approximate overlaps situations into the training data set. That makes the GPR model have a great computational accuracy using a few samples only.

References

1. BASISTA M., WĘGLEWSKI W., 2006, Modelling of damage and fracture in ceramic matrix composites – an overview, *Journal of Theoretical and Applied Mechanics*, **44**, 3, 455-484

2. BOULENOUAR A., BENSEDDIQ N., MAZARI M., BENAMARA N., 2014, FE model for linear-elastic mixed mode loading: estimation of SIFs and crack propagation, *Journal of Theoretical and Applied Mechanics*, **52**, 2, 373-383

3. CUI W., WANG J., 2011, Probabilistic analysis of gas turbine disk multi-crack propagation, *Turbo Expo: Power for Land, Sea, and Air*, DOI: 10.1115/GT2011-45439

4. HUANG D., YAN X., LI P., QIN X., ZHANG X., QI M., LIU Z., 2018, Modeling of temperature influence on the fatigue crack growth behavior of superalloys, *International Journal of Fatigue*, **110**, 22-30

5. HUANG D., YAN X., QIN X., ZHANG X., QI M., LIU Z., TAO Z., 2019, Scatter in fatigue crack growth behavior of a Ni-base superalloy at high temperature, *International Journal of Fatigue*, **118**, 1-7

6. HUANG X., CHEN C., XUAN H., 2021, Experimental and analytical investigation for fatigue crack growth characteristics of an aero-engine fan disc, *International Journal of Fatigue*, **148**

7. KEPRATE A., CHANDIMA RATNAYAKE R.M., SANKARARAMAN S., 2017, Comparison of various surrogate models to predict stress intensity factor of a crack propagating in offshore piping, *Journal of Offshore Mechanics and Arctic Engineering*, **139**, 6

8. LI P., CHENG L., YAN X., HUANG D., QIN X., ZHANG X., 2018, A temperature-dependent model for predicting the fracture toughness of superalloys at elevated temperature, *Theoretical and Applied Fracture Mechanics*, **93**, 311-318

9. LI Y., WANG J., GUO W., GUO J., 2019, A modified model of residual strength prediction for metal plates with through-thickness cracks, *Journal of Theoretical and Applied Mechanics*, **57**, 3, 537-547

10. LIU X., ATHANASIOU C.E., PADTURE N.P., SHELDON B.W., GAO H., 2020, A machine learning approach to fracture mechanics problems, *Acta Materialia*, **190**, 105-112

11. LIU Y., MAHADEVAN S., 2009, Probabilistic fatigue life prediction using an equivalent initial flaw size distribution, *International Journal of Fatigue*, **31**, 3, 476-487

12. MEGUID S.A., KANTH P.S., CZEKANSKI A., 2000, Finite element analysis of fir-tree region in turbine discs, *Finite Elements in Analysis and Design*, **35**, 4, 305-317

13. MOUSTABCHIR H., ALAOUI M.A.H., BABAOUI A., DEARN K.D., PRUNCU C.I., AZARI Z., 2017, The influence of variations of geometrical parameters on the notching stress intensity factors of cylindrical shells, *Journal of Theoretical and Applied Mechanics*, **55**, 2, 559-569

14. MOUSTABCHIR H., ARBAOUI J., ZITOUNI A., HARIRI S., DMYTRAKH I., 2015, Numerical analysis of stress intensity factor and T-stress in pipeline of steel P264GH submitted to loading conditions, *Journal of Theoretical and Applied Mechanics*, **53**, 3, 665-672

15. MUÑOZ-ABELLA B., RUBIO L., RUBIO P., 2015, Stress intensity factor estimation for unbalanced rotating cracked shafts by artificial neural networks, *Fatigue and Fracture of Engineering Materials and Structures*, **38**, 3, 352-367

16. NEWMAN JR J., RAJU I., 1981, Stress-intensity factor equations for cracks in three-dimensional finite bodies, *ASTM National Symposium on Fracture Mechanics*, DOI: 10.1520/stp37074s

17. RASMUSSEN C.E., 2003, Gaussian processes in machine learning, *Summer School on Machine Learning*, 63-71, Springer, DOI: 10.1007/978-3-540-28650-9_4

18. RINALDI A., KRAJCINOVIC D., MASTILOVIC S., 2006, Statistical damage mechanics-constitutive relations, *Journal of Theoretical and Applied Mechanics*, **44**, 3, 585-602

19. SHLYANNIKOV V., ZAKHAROV A., YARULLIN R., 2016, Structural integrity assessment of turbine disk on a plastic stress intensity factor basis, *International Journal of Fatigue*, **92**, 234-245

20. WITEK L., 2012, Numerical simulation of fatigue fracture of the turbine disc, *Fatigue of Aircraft Structures*, **1**, 4, 114-122

21. WU D., LIN C.-T., HUANG J., 2019, Active learning for regression using greedy sampling, *Information Sciences*, **474**, 90-105

22. XU T., DING S., ZHOU H., LI G., 2021, Machine learning-based efficient stress intensity factor calculation for aeroengine disk probabilistic risk assessment under polynomial stress fields, *Fatigue and Fracture of Engineering Materials and Structures*, **45**, 2, 451-465

23. YANG F., PAN C., ZHANG D., TANG J, YAN J., 2017, Stress distribution and deformation analysis of gas turbine blades and disk with FEM method, *ASME Power Conference*, DOI: 10.1115/POWER-ICOPE2017-3409

24. YU H., KIM S., 2010, Passive sampling for regression, *2010 IEEE International Conference on Data Mining*, DOI: 10.1109/icdm.2010.9

25. YUAN R., LIAO D., ZHU S.P., YU Z.Y., CORREIA J., DE JESUS A., 2021, Contact stress analysis and fatigue life prediction of turbine disc-blade attachment with fir-tree tenon structure, *Fatigue and Fracture of Engineering Materials and Structures*, **44**, 4, 1014-1026