

ORIGINAL ARTICLE

Open Access



Crack Growth Rate Model Derived from Domain Knowledge-Guided Symbolic Regression

Shuwei Zhou, Bing Yang^{*} , Shoune Xiao, Guangwu Yang and Tao Zhu

Abstract

Machine learning (ML) has powerful nonlinear processing and multivariate learning capabilities, so it has been widely utilised in the fatigue field. However, most ML methods are inexplicable black-box models that are difficult to apply in engineering practice. Symbolic regression (SR) is an interpretable machine learning method for determining the optimal fitting equation for datasets. In this study, domain knowledge-guided SR was used to determine a new fatigue crack growth (FCG) rate model. Three terms of the variable subtree of ΔK , R -ratio, and ΔK_{th} were obtained by analysing eight traditional semi-empirical FCG rate models. Based on the FCG rate test data from other literature, the SR model was constructed using Al-7055-T7511. It was subsequently extended to other alloys (Ti-10V-2Fe-3Al, Ti-6Al-4V, Cr-Mo-V, LC9cs, Al-6013-T651, and Al-2324-T3) using multiple linear regression. Compared with the three semi-empirical FCG rate models, the SR model yielded higher prediction accuracy. This result demonstrates the potential of domain knowledge-guided SR for building the FCG rate model.

Keywords Fatigue crack growth rate, Stress intensity factor range, Threshold stress intensity factor range, R -ratio, Symbolic regression, Machine learning

1 Introduction

Fatigue failure is the most prevalent breakdown mechanism in engineering structures that are subjected to long-term load disturbances [1, 2]. The fatigue life has previously been characterised as two independent processes involving “crack initiation” and “crack growth” [3]. Fatigue crack growth (FCG) prediction is essential for the damage-tolerant design of engineering components [4, 5]. Stress intensity factor (SIF) is the main characteristic of crack growth [6, 7]. Based on the linear elastic fracture method (LEFM), the earliest model was proposed by Paris and Erdogan [8] and had a profound impact on the

field. However, Paris’ model fails to perform well in the threshold and fast-fracture regions.

Various studies have suggested improvements to the Paris’ model to handle new governing factors impacting the FCG rate, which can be separated into two types. First, an effective SIF (ΔK_{eff}) based on the crack closure phenomenon was proposed to characterise the effect of the stress ratio on the FCG [9–11]. The other methods established the FCG rate prediction model directly based on the numerical fitting method considering parameters such as R -ratio, threshold value (ΔK_{th}), and fracture toughness (K_C) [12–17]. Although several semi-empirical models have been developed, they have some application restrictions, which are discussed in the following section. FCG rate prediction is a nonlinear multivariable problem.

Machine learning (ML) has powerful nonlinear processing and multivariate learning capabilities. It has been widely used for crack growth to solve complex nonlinear prediction problems [18–25]. Indeed, the radial basis

*Correspondence:

Bing Yang

yb@swjtu.edu.cn

State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu 610031, China

function artificial neural network (RBF-ANN), backpropagation neural network (BPNN), extreme learning machine (ELM), fully connected neural network, random forest (RF), hidden Markov model (HMM), and long short-term memory (LSTM) all yield accurate life and crack growth predictions [26–31]. However, most current FCG rate models based on ML are inexplicable black-box models that are difficult to apply in engineering practice. Simultaneously, it is often desirable to find interpretable models for insight into the FCG models. Schmidt and Lipson [32] used symbolic regression (SR) to obtain interpretable formulas from pendulum motion test data. Even after removing the sin and cos components from the initial function, a Taylor series expansion expression with these two terms appeared in the final expression. In contrast to ordinary regression approaches, SR employs only one assumption: the expression for the relationship between the target and characteristic parameters can be obtained by combining various elementary functions using algebraic operations [33]. More importantly, the SR is a white-box model whose output comprises a defined initial function sign with constants and input variables [34]. At present, SR has been applied to the fields of astronomy, materials science, and engineering [35–38], demonstrating the prospect of discovering interpretable models from data-driven models.

SR is a violent search method based on genetic programming [39]. An infinite set of mathematical function symbols, input variables, and constants exists and the determination of an equation that is both simple and accurate for such a large set is time-consuming. To constrain the search space, it must rely on the domain knowledge and inspiration. In this study, eight existing semi-empirical FCG rate models were analysed, and the domain knowledge in the existing FCG rate model was inherited from the SR. The equation was derived from the FCG test data of Al-7055-T7511, which includes three variables, ΔK , R , and ΔK_{th} through the SR. The SR model is straightforward and suitable for engineering applications. This work also provides a fitting method for the SR model, which is used to fit the crack growth rate test results of six metal materials. The accuracy and universality of the SR model were evaluated and compared with traditional FCG rate model prediction results.

2 Crack Growth Models

The generally accepted FCG theory, based on the relation between the FCG rates da/dN and ΔK , was proposed by Paris [8], and is given by Eq. (1).

$$\frac{da}{dN} = C \cdot (\Delta K)^m, \tag{1}$$

where a and N represent the crack length and number of loading cycles, respectively, while C and m are material

constants in the Paris' model. The range of the SIF is determined by $\Delta K = K_{max} - K_{min}$, where K_{max} and K_{min} are the maximum and minimum stress intensity factors, respectively. In Figure 1, the FCG rates of identical materials at various R -ratios are shown as a series of essentially parallel curves associated with ΔK [14]. The FCG rate curve shifts to the left overall as the R -ratio increases. The R -ratio is defined by $R = \sigma_{min}/\sigma_{max}$, where σ_{max} and σ_{min} denote the maximum and the minimum loading amplitude. It is difficult to reflect the effect of the R -ratio on the FCG rate of the materials using only ΔK . To overcome this issue, some well-known FCG rate models incorporate the contribution of R -ratios, which are briefly listed in the following section. Furthermore, it is generally assumed that the crack does not grow when $\Delta K < \Delta K_{th}$. Therefore, ΔK_{th} in the FCG rate model is particularly important in practical engineering applications.

2.1 Forman's Model

Forman et al. [12] modified Paris' model to integrate the Paris region and fast-fracture region FCG behaviour, as shown in Eq. (2).

$$\frac{da}{dN} = \frac{C(\Delta K)^m}{(1-R)(K_C - K_{max})}, \tag{2}$$

where K_{max} denotes the maximum SIF, and K_C is fracture toughness that denotes the transition to unstable crack growth. As K_{max} approaches K_C , the denominator approaches zero and the FCG rate increases significantly. Thus, the model matches the fast-fracture region FCG. However, Forman's model is inadequate for forecasting FCG behaviour in the threshold region.

2.2 Elber's Model

Elber [9] introduced the notion of crack closure and proposed ΔK_{eff} as a driving force to characterise the influence of the R -ratio on the FCG rate. This model is expressed by Eqs. (3) and (4).

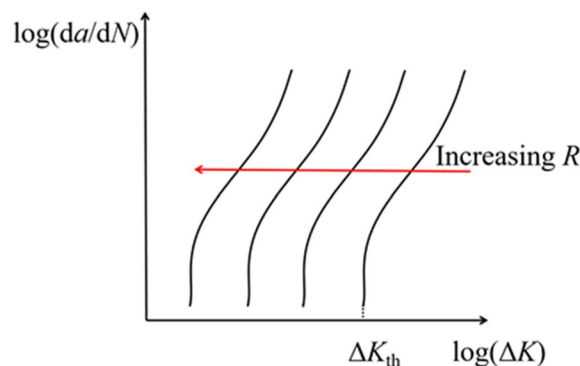


Figure 1 Crack growth rate curves under different R -ratios

$$\frac{da}{dN} = C \cdot (\Delta K_{\text{eff}})^m, \tag{3}$$

$$\Delta K_{\text{eff}} = K_{\text{max}} - K_{\text{op}}, \tag{4}$$

where K_{op} is the SIF corresponding to the crack opening stress σ_{op} , which is determined by the load associated with a 2% deviation in the slope of the load–displacement curve [40]. The crack closure phenomenon has a more significant impact at a low R -ratio. With the crack opening K_{op} approaching the minimum SIF, the crack closure phenomenon can be ignored at a high R -ratio.

2.3 Kujawski’s Model

Kujawski [13] established a new driving force model for predicting the FCG rate of aluminium alloys. This model is not based on the crack closure effect, but selects the positive value of the SIF range (ΔK^+) and K_{max} as the driving force to explain the R -ratio effect on crack growth. This model can be described using Eqs. (5) and (6).

$$\frac{da}{dN} = C \cdot (\Delta K^*)^m, \tag{5}$$

$$\Delta K^* = (K_{\text{max}} \Delta K^+)^{0.5}, \tag{6}$$

where $\Delta K^+ = \Delta K$ when $R \geq 0$, and $\Delta K^+ = K_{\text{max}}$ for $R < 0$. Subsequently, Eq. (6) is expanded into Eq. (7) based on the crack growth test data of the other metals [41, 42].

$$\Delta K^* = (K_{\text{max}})^{\alpha_K} (\Delta K^+)^{1-\alpha_K}, \tag{7}$$

where $0 \leq \alpha_K \leq 1$ is a fitting parameter determined by the material and geometry. Kujawski’s model is based on the assumption that when $R < 0$, negative stress does not contribute to crack growth. The impact of compressive loads on crack growth cannot be overlooked, following the observations of subsequent studies [43, 44].

2.4 Huang’s Model

To overcome the overestimated results in Kujawski’s model at high R -ratios, Huang et al. [14] developed an improved FCG model by introducing a correction factor M in the form of a piecewise function, as shown in Eqs. (8) and (9).

$$\frac{da}{dN} = C(M \Delta K)^m, \tag{8}$$

$$M = \begin{cases} (1 - R)^{-\beta_1} (-5 \leq R < 0), \\ (1 - R)^{-\beta} (0 \leq R < 0.5), \\ (1.05 - 1.4R + 0.6R^2)^{-\beta} (0.5 \leq R < 1). \end{cases} \tag{9}$$

where β and β_1 are constants that depend on the material properties and environmental conditions, and $\beta_1 = 1.2 \times \beta$. For the aluminum alloy and steel, the parameter β is approximately set to 0.7, while for Ti-6Al-4V, it is set to 0.5. Compared with Kujawski’s model, this model not only considers the contribution of the compression load, but also provides more accurate results under $R > 0.5$. More accurate parameters, β and β_1 can be acquired using numerous sets of test data. However, the FCG rate data with varying R -ratios are sometimes inadequate for engineering materials.

2.5 Zhan’s Model

Because the $da/dN - \Delta K$ curve resembles the parallel line under different R -ratios, Zhan et al. [15] proposed a simplified FCG rate prediction model, as shown in Eqs. (10) and (11).

$$\frac{da}{dN} = C_0 \cdot (\Phi \Delta K)^{m_0}, \tag{10}$$

$$\Phi = \exp(\alpha_Z R), \quad (-1 \leq R < 1), \tag{11}$$

where C_0 and m_0 are the material constants corresponding to $R = 0$ in Huang’s model, and α_Z is the correction factor of the R -ratio. It can be solved by Eq. (12).

$$\log \left(\frac{da}{dN} \right) \Big|_{R \neq 0} - \log \left(\frac{da}{dN} \right) \Big|_{R=0} = \Delta \log \left(\frac{da}{dN} \right) = m \cdot \alpha_Z \cdot R, \tag{12}$$

The solution of Zhan’s model is straightforward, which compensates for the limitation that Huang’s model is unsuitable for high-strength alloy steel. For most low-strength alloys, α_Z is set to 0.65. For some high-strength alloys, such as titanium alloys, α_Z is set to 0.75. Zhan suggested that the FCG rate curve under an arbitrary constant R -ratio could be used as the basic curve in the scarcity of test data with $R=0$. This model is limited to Paris’ region, and the choice of a basic curve with varying R -ratios results in erroneous solutions of C_0 , m_0 , and α_Z .

2.6 NASGRO’s Model

Based on the crack closure theory, Mettu et al. [10] proposed the NASGRO model of crack growth rate suitable for the entire process of the threshold, Paris’, and fast-fracture regions. Its form is shown in Eq. (13).

$$\frac{da}{dN} = C(\Delta K_{\text{eff}})^m \frac{(1 - \Delta K_{\text{th}}/\Delta K)^p}{(1 - K_{\text{max}}/K_C)^q}, \quad (13)$$

where p and q describe the contribution parameters of the threshold and fast-fracture regions, respectively. Because only a small fraction of the fatigue life is spent in the fast-fracture region, Zhang et al. [45] simplified the NASGRO model using Eq. (14).

$$\frac{da}{dN} = C(\Delta K_{\text{eff}})^m \left(1 - \frac{\Delta K_{\text{eff,th}}}{\Delta K_{\text{eff}}}\right)^p, \quad (14)$$

where ΔK_{eff} and $\Delta K_{\text{eff,th}}$ denote the effective SIF and its threshold value, respectively. This can be expressed by Eqs. (15) and (16).

$$\Delta K_{\text{eff}} = U \Delta K, \quad (15)$$

$$\Delta K_{\text{eff,th}} = U \Delta K_{\text{th}}, \quad (16)$$

where U is expressed in terms of Newman's crack closure function (f), and R in the form of Eq. (17) [11].

$$U = \left(\frac{1-f}{1-R}\right), \quad (17)$$

f is given by Eq. (18) as:

$$f = \begin{cases} \max(R, A_0 + A_1R + A_2R^2 + A_3R^3) & (R \geq 0), \\ A_0 + A_1R & (-2 \leq R < 0), \end{cases} \quad (18)$$

while Newman's crack closure estimations are expressed by Eqs. (19)–(22).

$$A_0 = \left(0.8613 - 0.3387\alpha_N + 0.0465\alpha_N^2\right) \left[\cos\left(\frac{(\pi/2)S_{\text{max}}}{\sigma_0}\right)\right]^{\frac{1}{\alpha_N}}, \quad (19)$$

$$A_1 = (1.047 - 0.233\alpha_N) \frac{S_{\text{max}}}{\sigma_0}, \quad (20)$$

$$A_2 = 1 - A_0 - A_1 - A_3, \quad (21)$$

$$A_3 = 2A_0 + A_1 - 1, \quad (22)$$

ere, S_{max} represents the maximum stress during the load cycle. The flow stress σ_0 was calculated as the average between the uniaxial yield and ultimate tensile strength [46]. α_N is the constraint value used to account for the three-dimensional stress state [47], depending on the material. The NASGRO model is highly accurate and is applicable to all the three regions of the FCG process. However, it must confirm various parameters, and a large amount of test data is required to support its establishment while considering crack closures. The various phenomena that cause crack closure include the crack

plasticity region, crack surface roughness, fluid inside the crack, and corrosion deposits near the crack tip [48]. The crack closure phenomenon cannot be precisely represented quantitatively, because it is heavily influenced by slight variations in the crack path, ambient factors, loading conditions, and test methods [49]. The model must be simplified to enhance its applicability to engineering problems, considering time and test costs.

3 Symbolic Regression

3.1 High-Performance Symbolic Regression in Python

As previously mentioned, the existing FCG models are either solely applicable to the Paris region or excessively complex for considering the crack closure phenomenon, and require a large amount of test data for support. Recently, various scholars have used ML regression models to predict FCG rates. Both the ML and traditional regression models were based on specific model parameters. For example, the standard linear regression model is based on the linear relationship between the dependent and independent variables. As a nonlinear technique, artificial neural networks (ANN) also depend on the definition of the activation or transfer functions.

SR does not require assuming a specific form of the function between independent and dependent variables in advance. Instead, it only requires that the connection between the independent and dependent variables be described using function expressions. Simultaneously, the special mathematical operators, constants, and analytical functions are introduced to search for the optimal solution using these unique module combinations. Any equation in the SR can be expressed in the form of a binary tree. The SR tree structure comprises the terminal and internal nodes, which represents the constant and variable, and the function and operation symbols, respectively. The SR tree representation equation is presented in Figure 2. The model becomes increasingly sophisticated as the length and depth of the tree increase. To avoid excessive tree development, the length and depth of the trees must be limited. SR is essentially genetic programming, which resembles Darwin's natural selection principle. In genetic programming, the initial population is randomly generated according to the defined function, and individuals are screened according to their fitness. It is easier for an individual possessing higher fitness to appear in the next generation of individuals. Individuals with higher fitness evolve through crossover and mutation. As shown in Figure 2, crossover refers to the random exchange of subtrees between the equations of the previous generation, which produces new offspring individuals. Mutation implies that a node or multiple nodes in the equation are randomly adjusted to

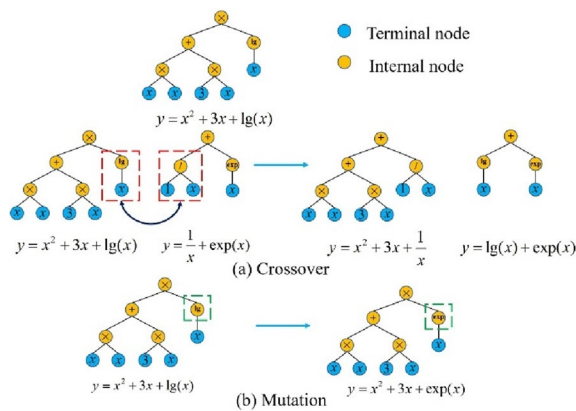


Figure 2 Structure trees of SR

ensure population variety to explore better data-fitting equations. Fitness evaluation, screening, crossover, and mutation occur in each generation to produce new individuals, and this process repeats until it reaches the number of iterations or is artificially terminated. In this study, high-performance symbolic regression in Python (PySR) is used for the symbolic regression [50].

The procedure for the domain knowledge-guided SR is shown in Figure 3. In PySR, each node has a Complexity of one and the Complexity increases with the number of nodes. Each variable, constant, or operator in the equation increases Complexity by one. The equation not only requires considering higher accuracy, but also its Complexity [51]. The SCORE is defined by Eq. (23).

$$SCORE = \frac{-[\ln(\text{curMSE}) - \ln(\text{lastMSE})]}{\text{curComplexity} - \text{lastComplexity}}, \quad (23)$$

where curMSE and lasMSE are the mean square error (MSE) of the contemporary and last individuals, respectively, while curComplexity and lastComplexity represent

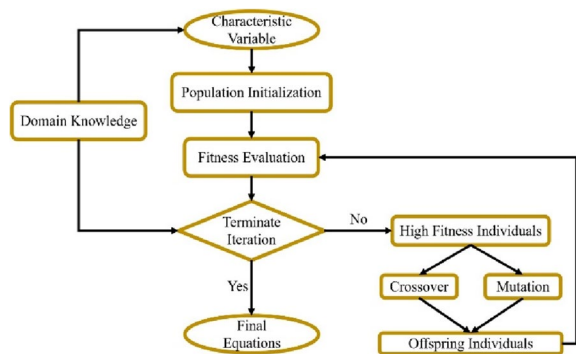


Figure 3 Procedure of domain knowledge-guided SR

the complexities of the contemporary and last individuals, respectively.

The main advantage of SR is that its output results are visual formulae with interpretable findings. However, genetic programming based on SR is essentially a random search process, with an almost limitless search space, and brute-force search without any preconditions may consume a lot of time and memory. Therefore, it is necessary to study the existing semi-empirical FCG rate model to identify specific conditions that may be used to develop an FCG rate model based on symbolic regression.

3.2 Domain Knowledge-Guided

The existing method for solving for the parameters of the FCG rate model usually uses the logarithm of the left and right sides of the equation. The corresponding results for the eight FCG rate models are shown in Table 1. The model is represented in SR tree form for the convenience of analysis, as shown in Figure 4. Figure 4 demonstrates that regardless of the type of FCG rate model, the compensation function of the constant term $\ln C$ is included, and values of the different models vary so that the crack growth rate equation can be expressed as $\ln(da/dN) = g(\bullet)$. The variables that affect $g(\bullet)$, both of which affect the crack growth rate of the material, include ΔK , K_{max} , K_{op} , f , ΔK^+ , R , ΔK_{th} and K_C . K_{max} and ΔK^+ can be represented individually by

Table 1 Eight equations of traditional FCG rates prediction models

Model name	Equation
Paris' model	$\ln C + m \cdot \ln \Delta K$
Forman's model	$\ln C + m \cdot \ln \Delta K - \ln [(1 - R) \cdot K_C - \Delta K]$
Elber's model	$\ln C + m \cdot \ln \left(\frac{\Delta K}{1-R} - K_{op} \right)$
Kujawski's model	$\ln C + m \cdot \alpha_K \cdot \ln \frac{\Delta K}{1-R} + (1 - \alpha_K) m \cdot \ln \Delta K^+$
Huang's model	$\ln C + m \cdot \ln \Delta K + m \cdot \ln M$
Zhan's model	$\ln C + m \cdot \ln \Delta K + m \cdot \alpha_Z \cdot R$
NASGRO model	$\ln C + m \cdot \ln \Delta K + m \cdot \ln \frac{1-f}{1-R} + p \cdot \ln \left(1 - \frac{\Delta K_{th}}{\Delta K} \right) - q \cdot \ln \left(1 - \frac{K_{max}}{K_C} \right)$
Simplified NASGRO model	$\ln C + m \cdot \ln \Delta K + m \cdot \ln \frac{1-f}{1-R} + p \cdot \ln \left(1 - \frac{\Delta K_{th}}{\Delta K} \right)$

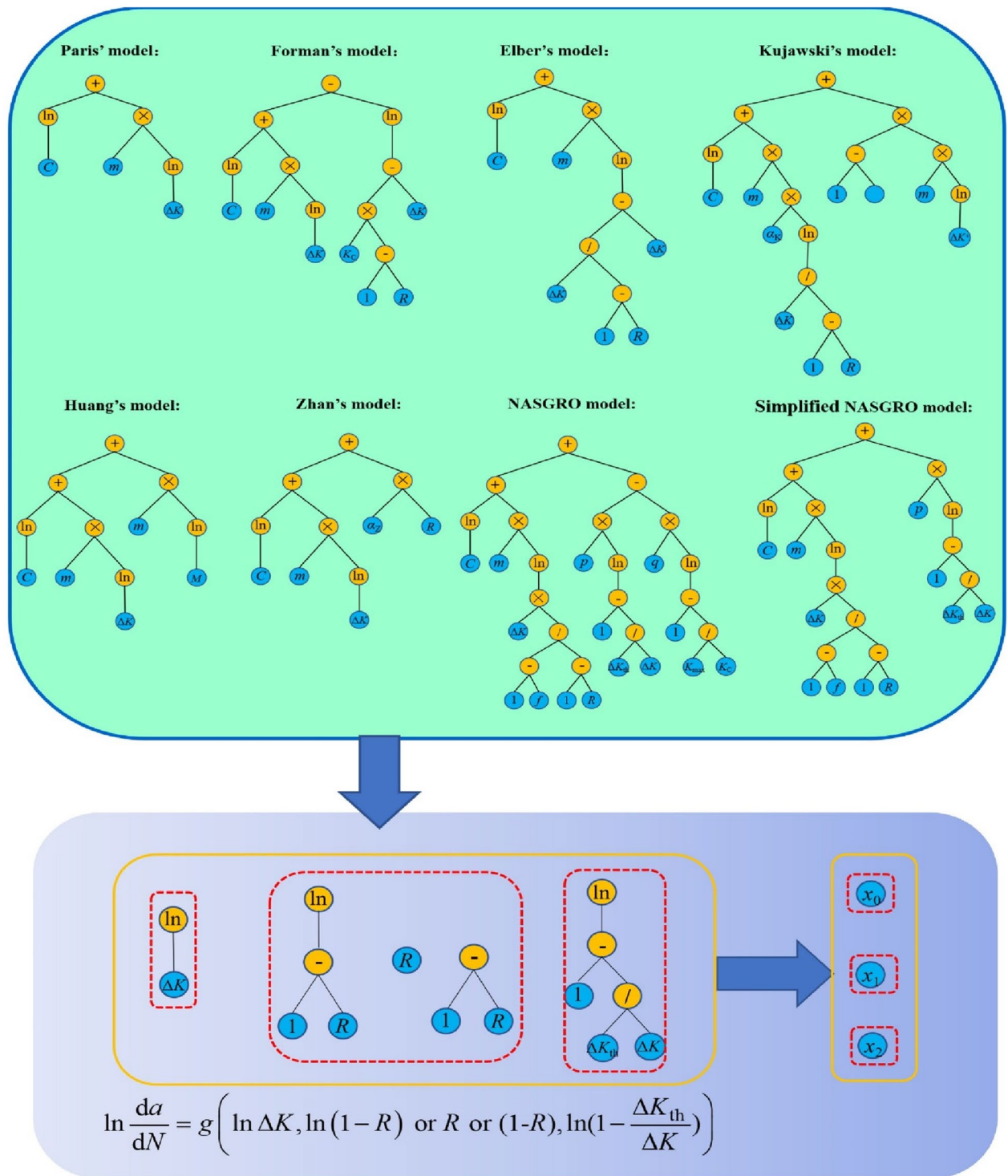


Figure 4 SR trees of eight traditional models equation for FCG rates prediction and formula structures

a function that includes ΔK and R . K_{op} and f are crack closure parameters that must be determined by the FCG test data. As described above, a precise quantitative description of crack closure was not feasible [52].

The threshold and Paris regions dominated the crack growth process most of the time. The fast-fracture region accounted for only a minor portion of the time during crack growth. Therefore, the five parameters

K_{max} , ΔK^+ , K_{op} , f and K_C were excluded from the present model. In the above eight FCG rate models, the ΔK and ΔK_{th} terms usually appeared in the form of $\ln\Delta K$ and $\ln(1-\Delta K_{th}/\Delta K)$. The R term may appear in the form of $\ln(1-R)$, R or $(1-R)$; therefore, three forms of R terms were imported to analyse R -ratio contribution to the establishment of the FCG rate model. The remaining $\ln C$, m , α_z , p and q are constant terms that can be generated by SR. Moreover, traditional semi-empirical FCG rate models contain Paris' term, $\ln C + m \times \ln\Delta K$. Therefore, the equation generated by SR should include Paris' term.

Thus, the FCG rate model established by the domain knowledge-guided SR is mainly related to ΔK , ΔK_{th} and R , which can be inferred from Eq. (24).

$$\ln \frac{da}{dN} = g \left(\ln \Delta K, \ln(1-R) \text{ or } R \text{ or } (1-R), \ln \left(1 - \frac{\Delta K_{th}}{\Delta K} \right) \right), \quad (24)$$

The FCG model established in the present work primarily included three variable parameters, which were represented by the SR tree, as shown in Figure 4. Previously, researchers proposed a connection between the three parameters based on experience or test data. The artificial introduction of the relationship between the three parameters may affect the accuracy of model establishment. In this study, PySR was used to explore the relationships between the three variable parameters. Here, x_0 , x_1 , and x_2 represent $\ln\Delta K$, $\ln(1-R)$ or R or $(1-R)$, and $\ln(1-\Delta K_{th}/\Delta K)$, respectively.

In the present study, the SR model was established using PySR, and the model parameters are listed in Table 2. The data used were the Al-7055-T7511 [14] FCG test data obtained from other studies. Niteration means the number of iteration, which was set to a large value (2000) here. PySR can yield optimal offspring individuals in real-time. The training process was manually terminated after identifying an interpretable and highly accurate individual. The operators '+', '-', '*', '/', 'ln_abs()' were used.

Table 2 Basic settings for PySR

Parameters	Value
Material	Al-7055-T7511
Niterations	2000
Complexity	20
Operators	'+', '-', '*', '/', 'ln_abs()'
Constraints	'ln_abs':(1)
Loss	$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$
Input characters	$\ln \Delta K, \ln(1-R)$ or $(1-R)$ or $R, \ln \left(1 - \frac{\Delta K_{th}}{\Delta K} \right)$
Output character	$\ln \frac{da}{dN}$

were used. To avoid the overcomplex of individuals generated by SR, the Complexity is set to 20, which means that the total number of operators, constants, and variables in equations can not beyond 20. A high logarithmic term such as $\ln(\ln(\bullet)+\bullet)$ did not appear in the semi-empirical FCG rate model; to obtain an interpretable solution, there was at most a variable in the constraint $\ln_abs()$ function. The MSE was used as the loss function of the SR to judge fitness. The Pearson correlation coefficient (r) was adopted to evaluate the fitting degree between the predicted and test values [53, 54], which is defined by Eq. (25):

$$r = \frac{\sum (y_{pre} - \bar{y}_{pre})(y_{test} - \bar{y}_{test})}{\sqrt{\sum (y_{pre} - \bar{y}_{pre})^2} \sqrt{\sum (y_{test} - \bar{y}_{test})^2}}, \quad (25)$$

where y_{pre} and y_{test} denote the predicted and test values of the output separately, and \bar{y}_{pre} and \bar{y}_{test} denote the mean of two variables.

The input characteristic X and output characteristic Y were $\ln\Delta K$, $\ln(1-R)$ or R or $(1-R)$, $\ln(1-\Delta K_{th}/\Delta K)$, and $\ln(da/dN)$, respectively. Moreover, the output feature Y of PySR was the logarithmic form of the FCG rates, which is $\ln(da/dN)$; therefore, the quantitative evaluation parameters used in the present work were all based on the value of feature Y .

Moreover, integrating other FCG rate models provided more information for the SR process. The eight collected FCG rate models were suitable for a wide range of R-ratios and exhibited a high prediction ability for a variety of materials. Therefore, the eight FCG rate models were considered reliable references for establishing the SR model.

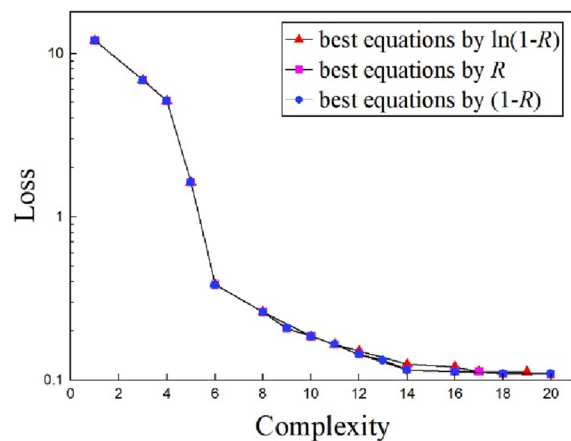


Figure 5 Pareto front of best equations showing Complexity and Loss

4 Results and Extensions

4.1 Symbolic Regression Results and Analysis

In this section, the FCG test data of the Al-7055-T7511 are used for the SR. The Pareto front in Figure 5 illustrates the trade-off between the equation complexity, as defined by the number of nodes in the SR tree, and MSE. To make the results more intuitive, we considered the logarithmic coordinates of the loss. Figure 5 shows that the loss decreases with an increase in the equation complexity, which represents an improvement in the accuracy of the regression. As shown in Figure 5, the loss of the equation generated by the three approaches is nearly identical before the complexity becomes less than 11.

When the complexity exceeds 12, the loss of the equation obtained by $\ln(1-R)$ exceeds that of the other two methods, indicating that the result obtained by R or $(1-R)$ is more accurate than that obtained by $\ln(1-R)$. The loss obtained using R or $(1-R)$ then decreases only marginally with more complex equations after the equation complexity reaches 14.

Tables 3, 4, 5 show the detailed equations obtained by the SR. Combined with Table 3, 4, 5, the equations with a complexity of less than 11 obtained by the other three methods have the same form, except for the equation of complexity 9. The equations of complexities 9 and 12 obtained by R or $(1-R)$ can take the same form after

Table 3 Best equation obtained by PySR using $\ln(1-R)$, where x_0 , x_1 , and x_2 represent $\ln\Delta K$, $\ln(1-R)$, and $\ln(1-\Delta K_{th}/\Delta K)$, respectively

Complexity	Loss	SCORE	Equation
1	11.95728320	0	-16.64
3	6.85792887	0.506034	$x_0 - 18.65$
4	5.12093165	1.015250	$-\ln(x_2) - 17.52$
5	1.62683224	5.138852	$-0.79/x_2 - 20.06$
6	0.38504615	7.903765	$-2.78 \times \ln(x_2) - 19.08$
8	0.26074687	1.355014	$-3.66 \times \ln(x_2) - x_2 - 20.76$
10	0.18593335	1.515445	$-3.33 \times \ln(x_2) - 0.62 \times x_2 - 20.12$
11	0.16588967	1.196780	$\ln(x_2) \times ((-0.31 \times \ln(x_2)) + 2.38) - 19.44$
12	0.15085690	1.091710	$-4.43 / (x_0 - x_2) - 2.79 \times \ln(x_2) - 17.41$
14	0.12528518	1.204919	$(-6.19 - x_1) / (x_0 - x_2) - 2.80 \times \ln(x_2) - 16.90$
16	0.12092982	0.264973	$6.19 / (x_0 - x_2) - 0.44 \times x_1 - 2.80 \times \ln(x_2) - 16.90$
17	0.11267048	1.166901	$-0.02 / (x_1 + 0.58) \times (0.31 \times \ln(x_2) - 2.38) + \ln(x_2) - 19.39$
19	0.11246359	0.016524	$-0.02 / (x_1 + 0.59) \times (0.31 \times \ln(x_2) - 2.38) - 0.02 + \ln(x_2) - 19.39$

Table 4 Best equation obtained by PySR using R , where x_0 , x_1 , and x_2 represent $\ln\Delta K$, R , and $\ln(1-\Delta K_{th}/\Delta K)$, respectively

Complexity	Loss	SCORE	Equation
1	11.95728320	0	-16.64
3	6.85792887	0.506034	$x_0 - 18.65$
4	5.12093165	1.015250	$-\ln(x_2) - 17.52$
5	1.62683224	5.138852	$-0.79/x_2 - 20.06$
6	0.38504615	7.903765	$-2.78 \times \ln(x_2) - 19.08$
8	0.26074687	1.355014	$-3.66 \times \ln(x_2) - x_2 - 20.76$
9	0.20797211	1.920022	$3.97 \times x_0 + 2.99 \times x_1 - 24.82$
10	0.18593337	1.063165	$-3.33 \times \ln(x_2) - 0.62 \times x_2 - 20.12$
11	0.16588967	1.196781	$\ln(x_2) \times ((-0.31 \times \ln(x_2)) + 2.38) - 19.44$
12	0.14459242	1.579150	$1.65 \times (x_0 - \ln(x_2)) + 1.12 \times x_1 - 18.17$
14	0.11595452	1.431851	$2.59 \times x_0 + 1.87 \times x_1 - 1.03 \times \ln(x_2) - 22.85$
16	0.11349541	0.160530	$2.61 \times x_0 + x_1 + x_1 - 0.02/x_1 - \ln(x_2) - 22.85$
17	0.11347807	0.002519	$2.62 \times x_0 + x_1 + x_1 + 0.10 \times \ln(x_1) - \ln(x_2) - 22.83$
18	0.10986783	0.565648	$2.67 \times x_0 + 0.16/x_0 + 1.88 \times x_1 - \ln(x_2) + 0.44 - 22.69$
20	0.10930385	0.048846	$2.69 \times x_0 + 0.13/x_0 + x_1 + x_1 - 0.02/x_1 - \ln(x_2) - 23.11$

Table 5 Best equation obtained by PySR using (1-R), where $x_0, x_1,$ and x_2 represent $\ln\Delta K, (1-R),$ and $\ln(1-\Delta K_{th}/\Delta K),$ respectively

Complexity	Loss	SCORE	Equation
1	11.95728320	0	-16.64
3	6.85792887	0.506034	$x_0 - 18.65$
4	5.12093165	1.015250	$-\ln(x_2) - 17.52$
5	1.62683224	5.138852	$-0.79/x_2 - 20.06$
6	0.38504615	7.903765	$-2.78 \times \ln(x_2) - 19.08$
8	0.26074687	1.355014	$-3.66 \times \ln(x_2) - x_2 - 20.76$
9	0.20797211	1.920022	$3.97 \times x_0 - 2.99 \times x_1 - 21.83$
10	0.18593335	1.063166	$-3.33 \times \ln(x_2) - 0.62 \times x_2 - 20.12$
11	0.16588967	1.196780	$\ln(x_2) \times ((-0.31 \times \ln(x_2)) + 2.38) - 19.44$
12	0.14459242	1.579150	$1.65 \times (x_0 - \ln(x_2)) + 1.12 \times x_1 - 20.32$
13	0.13260567	1.081161	$3.65 \times x_0 - 2.72 \times x_1 + 0.30 \times x_2 - 21.17$
14	0.11579778	1.828878	$2.55 \times x_0 - (1.79 \times x_1 + \ln(x_2)) / 0.97 - 20.95$
16	0.11292975	0.187817	$2.55 \times x_0 - (\ln(x_2) + x_1 \times (1.79 - 0.05 \times x_2)) - 20.95$
18	0.11002615	0.221151	$2.64 \times x_0 - 0.04 / x_0 / x_1 + 1.84 \times x_1 - \ln(x_2) + 0.44 - 21.17$
20	0.11002614	0.000001	$2.64 \times (x_0 - 0.07) - 0.04 / x_0 / x_1 + 1.84 \times x_1 - \ln(x_2) + 0.44 - 20.99$

numerical transformation. The complexity 6 equation has the highest evaluation SCORE of 7.903765, indicating that it possesses the optimal value of the improved precision-to-complexity ratio when compared to other equations. From the standpoint of model efficiency, the complexity 6 equation guarantees a lower complexity while considering the accuracy and is the optimum solution for SR. However, the equation contained only one characteristic variable x_2 , without x_0 and x_1 . This study aimed at building an FCG rate model with three characteristic variables: $\Delta K, R$ and ΔK_{th} . Furthermore, we discovered that this equation and the complexity 9 equation share some similarities with Zhan’s model, which confirms the reliability of the SR results. The form of the equation derived by the three approaches was dissimilar when the complexity exceeds 14. Further, the equations derived from the three subtrees are separately analysed.

4.1.1 Symbolic Regression Results by $\ln(1-R)$

Table 3 lists the detailed equations for $\ln(1-R)$. When the complexity is less than 11, the characteristic variables of each equation appear alone, and there is no x_1 term related to R . Those equations are inconsistent with the purpose of this study. Other equations lack Paris’ term which is necessary for traditional semi-empirical equations. Thus, the equations obtained using $\ln(1-R)$ do not contain the objective equations of this study.

4.1.2 Symbolic Regression Results by R

As shown in Table 4, equations with a complexity of less than 11 are consistent with those obtained by $\ln(1-R)$ and

are not analysed in this section. Only the fitting coefficient values differed between the equations of complexities 12 and 14, where the x_0 and $\ln(|x_2|)$ constant coefficients of the complexity 12 equation are the same, changing them to alternative coefficients reduces the loss and has a SCORE of 1.431851. Thus, the complexity 14 equation was chosen over the complexity 12 equation. Equations with complexity greater than 16 contain the Paris’ term necessary for the FCG rate model. In contrast, the x_1 term, conversely, takes the form $\ln(|x_1|)$ or $1/x_1$, which is singular at $R=0$. Therefore, only the complexity 14 model can be regarded as an SR-undetermined model.

4.1.3 Symbolic Regression Results by (1-R)

As shown in Table 5, the equations whose complexity is less than 12 are consistent with those obtained using R . Note that following a numerical operation, the equations of complexities 12 and 14 given by R or $(1-R)$ can assume the same form. As in Section 4.1.2, the complexity 14 model can be regarded as an SR-undetermined model. In addition, the complexity 13 equation is supplemented by the threshold subtree of the constant terms multiplier x_2 compared with the complexity 9 equation, and the constant multipliers x_0 and x_1 are adjusted. Therefore, the complexity 13 equation can be chosen as an SR-undetermined model. Moreover, the complexity 20 equation has the highest accuracy in this round of SR. Nonetheless, obtaining a hint of $x_0 \times x_1$ from the standard semi-empirical FCG rate model is problematic. The complexity 18 equation also lacks interpretability for the term. The complexity 16 equation also lacks an explanation for $x_1 \times x_2$ from the traditional semi-empirical FCG

rate model. Thus, the equations of complexity 13 and 14 models can be regarded as SR-undetermined models.

4.2 Equation Selection and Extension

Following the description provided in Section 4.1, there are currently three SR-undetermined models. The best equation is selected as the final model in this section. Because the equations of complexity 14 by R and $(1-R)$ could be converted to each other numerically, they have been treated as complexity 14 equations. The complexity 14 equations had a higher SCORE than the complexity 13 equations, indicating that the former conserved more equation space. Moreover, the complexity 14 equation was more precise than the complexity 13 equation. The two equations primarily differed in the processing of threshold terms. In the threshold term of the complexity 13 equation, the constant term multiplied by x_2 was added as compensation, followed by adding the constant term multiplied by $\ln(x_2)$.

Figure 6 shows the effect of the x_2 coefficient on the FCG process in the two SR-undetermined models. The influence of the threshold value on FCG is known to

growth. As illustrated in Figure 6a, the x_2 correlation factor of complexity equation 13 approaches the zero baseline as ΔK increases. However, the x_2 correlation factor of the complexity 14 equation maintains an increasing trend. After removing the x_2 term of the two SR-undetermined models, the complexity 13 equation in Figure 6b still reflects a good correlation with the test data in the Paris region whereas the complexity 14 equation results deviate significantly from the test data. Thus, considering the equation fitness, complexity, and interpretability, the complexity 13 equation was selected as the final model derived from the domain knowledge-guided SR, defined as the SR model.

The complexity 13 equation maintained the foundation of Zhan’s model and adds a parameter with a threshold value. The threshold value parameter resembled that of the NASGRO model and could be regarded as compensation for the threshold value of Zhan’s model. By replacing the constants in the model by constant coefficients, the SR model is expressed as Eq. (26).

$$\begin{aligned} \ln\left(\frac{da}{dN}\right) &= \ln C + m \ln(\Delta K) + q \ln\left(1 - \frac{\Delta K_{th}}{\Delta K}\right) + \alpha(1 - R) \\ &= \ln\left[C \cdot e^{\alpha(1-R)} \cdot (\Delta K)^m \cdot \left(1 - \frac{\Delta K_{th}}{\Delta K}\right)^q\right], \end{aligned} \tag{26}$$

be mainly concentrated in the threshold region. The impact of the threshold value rapidly diminishes after the FCG enters the Paris region. Therefore, the x_2 term should approach zero in the second half of crack

Exponentiating both sides of the equation, we obtain Eq. (27).

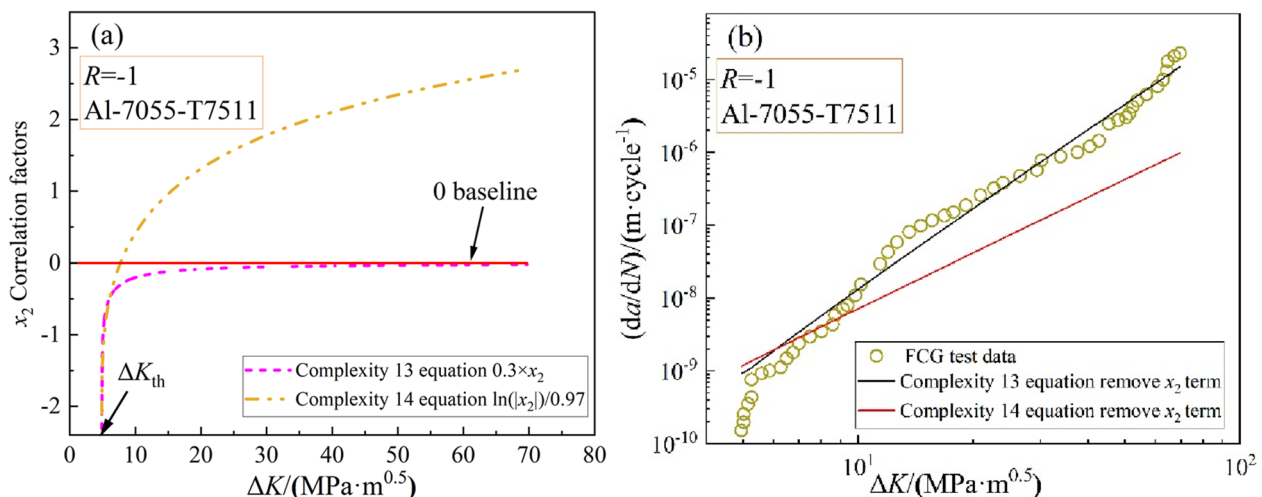


Figure 6 Effect of the x_2 term coefficient on the Al-7055-T7511 ($R = -1$) FCG process in two SR-undetermined models, **a** the x_2 term correction factor versus ΔK , **b** predicting curves and test data after removing the x_2 term

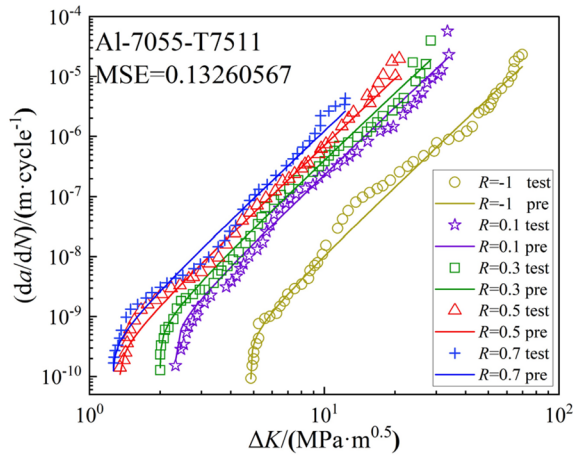


Figure 7 SR model predicting curves and test data about Al-7055-T7511

$$\frac{da}{dN} = C \cdot e^{\alpha(1-R)} \cdot (\Delta K)^m \cdot \left(1 - \frac{\Delta K_{th}}{\Delta K}\right)^q, \quad (27)$$

Defined $\alpha' = \alpha/m$, the SR model was simplified manually to Eq. (28).

$$\frac{da}{dN} = C \cdot (e^{\alpha'(1-R)} \Delta K)^m \cdot \left(1 - \frac{\Delta K_{th}}{\Delta K}\right)^q, \quad (28)$$

Therefore, an FCG rate model considering ΔK , R and ΔK_{th} was obtained using domain knowledge-guided SR. The various colour label points in Figure 7 represent the FCG test rates of the Al-7055-T7511 under different R -ratios, and the corresponding colour curves represent the predictions of the SR model. Figure 7 shows that the test results are consistent with those predicted by the SR model, and the MSE corresponding to $\ln(da/dN)_{pre}$ and $\ln(da/dN)_{test}$ is 0.13260567.

Observing Eq. (28), the explanatory variables $\ln(da/dN)$ and $(\ln \Delta K, 1-R, \ln(1-\Delta K_{th}/\Delta K))$ were regarded as a linear relationship, implying that there was a multivariate linear function between the explanatory variables. The SR model had four undetermined parameters, which were divided into the partial regression coefficient $K=(K_1, K_2, K_3)$ and constant term B . Hence, multiple linear regression (MLR), as shown in Eqs. (29) and (30), can be used to obtain the fitting parameters.

$$Y = B + KX, \quad (29)$$

$$\ln\left(\frac{da}{dN}\right) = \ln C + m \ln(\Delta K) + \alpha(1-R) + q \ln\left(1 - \frac{\Delta K_{th}}{\Delta K}\right), \quad (30)$$

Table 6 SR model fitting parameters for various materials

Materials	lnC	m	α	q
Ti-10V-2Fe-3Al	-23.08	3.29	-2.40	0.47
Ti-6Al-4V	-24.90	3.61	-2.07	0.38
Cr-Mo-V steel	-24.55	3.16	-2.76	0.26
LC9cs	-20.28	3.23	-2.29	0.49
Al-2324-T3	-22.85	4.78	-4.85	0.55
Al-6013-T651	-22.93	3.76	-2.63	0.36

which partial regression coefficient $K=(K_1, K_2, K_3)=(m, \alpha, q)$ and constant term $B=\ln C$. The MLR method can extend the SR model to other materials, and the next section demonstrates the application of the SR model to other materials and evaluates it in comparison with the semi-empirical models.

4.3 Performance Evaluation and Model Comparison

To demonstrate the effectiveness of the SR model, the following examples used test data for the FCG rate from the literature. The material fitting coefficient was obtained using the MLR. This section contains test data for the titanium alloys Ti-10V-2Fe-3Al [55], Ti-6Al-4V [56], Cr-Mo-V steel [57], aluminum alloys LC9cs [58], Al-2324-T3 [43] and Al-6013-T651 [59]. In this section, the minimum FCG rate corresponding to an order of $10^{-10} \text{ m} \cdot \text{cycle}^{-1}$ was considered as ΔK_{th} . Because the minimum crack growth rate of some test data was lower than $10^{-10} \text{ m} \cdot \text{cycle}^{-1}$, ΔK , corresponding to the minimum FCG rate, was defined as ΔK_{th} to ensure the integrity of the test data. The fitting parameters and correlation coefficients of the different materials obtained by the MLR are listed in Table 6. The value of α in Table 6 is the weight term of the R -ratio effect. In Zhan’s model, the value of α_Z is closer to 0.75 for some high-strength metallic materials such as titanium alloys, and for other metallic materials, α_Z is set to 0.65[15]. However, the value of α did not have a fixed value in the SR model. Zhan’s model ignored the threshold value and arbitrarily selected the test data under a constant R -ratio as the basic crack growth rate curve to solve, whereas the MLR method adopted by the SR model more comprehensively considers the influence of various variables.

The various colour label points in Figure 8 represent the FCG test rates with different R -ratios, and the corresponding colour curves represent the prediction of the SR model. According to Figure 8, the majority of the test data are condensed to the SR model’s predicting curves, and the MSE of the prediction and test is less than 0.1. Most of the test data points for titanium alloys Ti-6Al-4V,

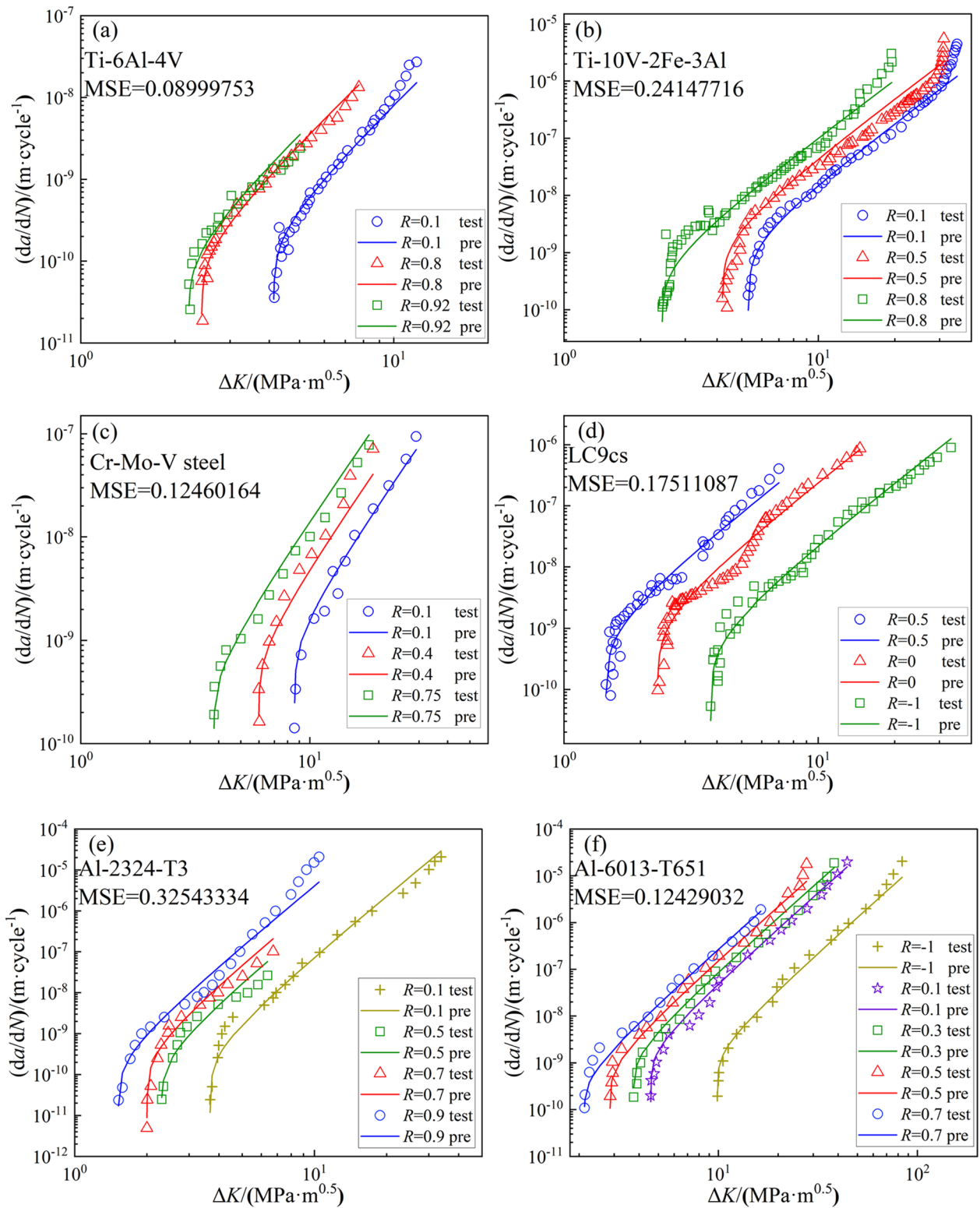


Figure 8 SR model predicting curves and test data about six materials: **a** Ti-6Al-4V, **b** Ti-10V-2Fe-3Al, **c** Cr-Mo-V steel, **d** LC9cs, **e** Al-2324-T3, **f** Al-6013-T651

Ti-10V-2Fe-3Al are consistent with the predictive curves, as shown in Figure 8a and b, indicating that the SR model is suitable for titanium alloys. The matching effect between the test data points and predicted

curves was relatively unsatisfactory when the FCG process approached the fast-fracture region because the SR model considered only the threshold and Paris regions and ignored the fast-fracture region of crack growth.

Table 7 MSE of the four models with different materials

	SR model	Kujawski's model	Huang's model	Zhan's model
Ti-6Al-4V	0.08999753	0.34794424	1.13046593	1.03957789
Ti-10V-2Fe-3Al	0.24147716	0.99731395	0.38633674	0.61096132
Cr-Mo-V	0.12460164	0.39392591	0.33399462	0.28708033
LC9cs	0.17511087	0.32597147	0.35360496	0.35360907
Al-2324-T3	0.32543334	1.26745959	0.94963109	0.82653289
Al-6013-T651	0.12429032	0.29828217	0.36357324	0.42400604
Al-7055-T7511	0.13260567	0.21224501	0.26963431	0.27983963
Average	0.17155109	0.51196363	0.46937770	0.50304010

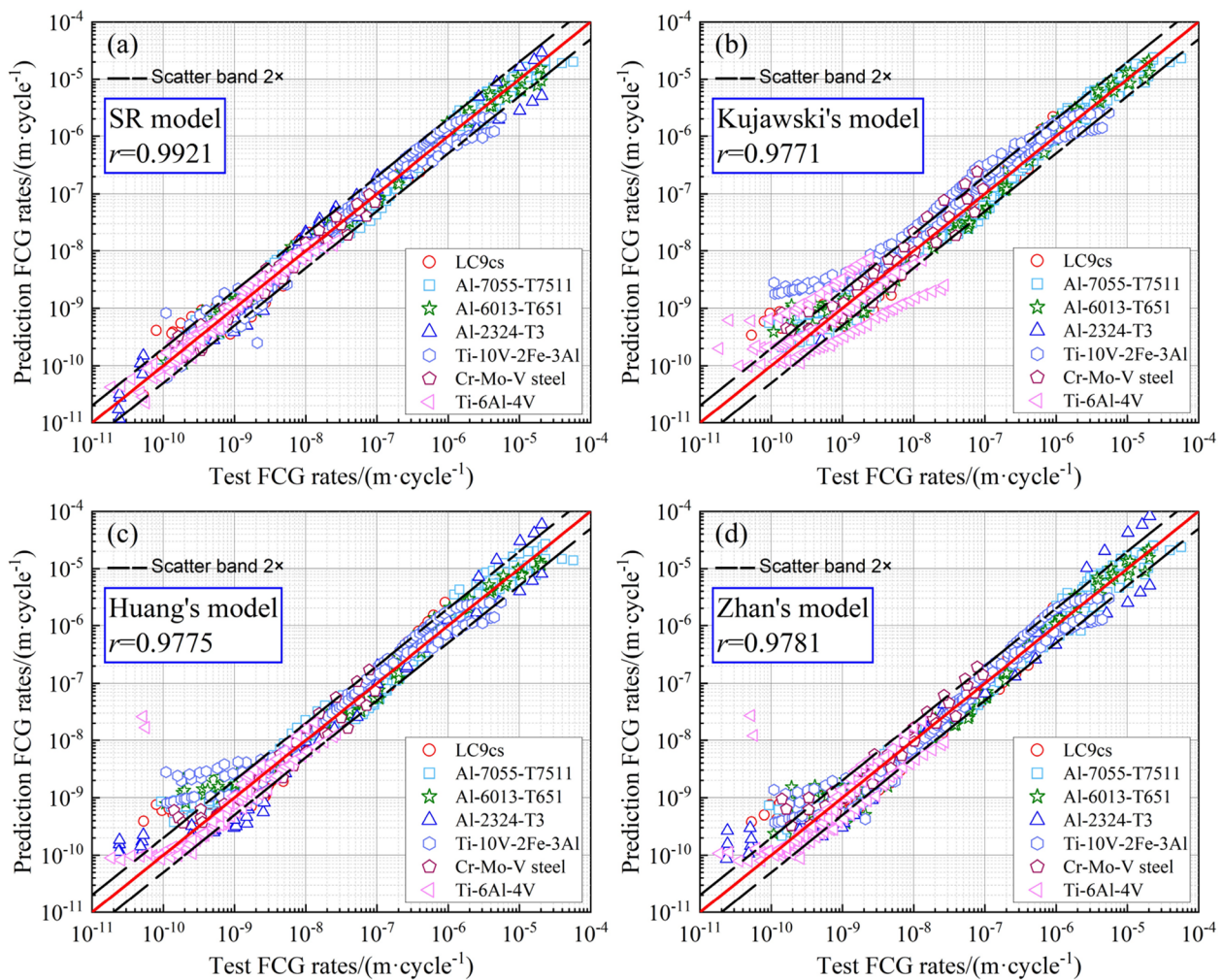


Figure 9 FCG rates based on four models between the SR model and test data: **a** SR model, **b** Kujawski's model, **c** Huang's model, **d** Zhan's model

Figure 8c demonstrates that the predicted curve has a good coincidence relationship with the test data for Cr-Mo-V steel which indicates that the SR model is appropriate for steel materials. According to Figure 8d, e, and f, the predicted curves for the aluminium alloys LC9cs, Al-2324-T3 and Al-6013-T651 have a good coincidence relationship with the test data, indicating the suitability of the SR model for aluminium alloys.

The crack growth rates predicted by the SR model were satisfactory for all the above materials and cases. Nevertheless, the R -ratios of the FCG test data used for prediction were between $-1 \leq R < 1$ in the present research work, so it is considered that the model showed a good prediction effect when the R -ratios were between $-1 \leq R < 1$. However, the prediction effect of the crack growth rate for $R < -1$ requires further verification.

Furthermore, the three FCG models, namely, Kujawski's model, Huang's model, and Zhan's model, were chosen for comparison with the SR model. Owing to the insufficient crack closure and K_C test data and Paris' model cannot predict the crack growth rate with different R -ratios, other FCG models introduced above were not evaluated in the present work. Table 7 summarises the MSE values of the various models for various materials and R -ratios, and Figure 9 compares the test and predicted values of the four FCG rate prediction models.

As shown in Table 7, the MSE of the prediction results of the SR model for various materials is smaller than those of the other three models, which shows the accuracy of the SR model in FCG rate prediction. Figure 9 shows that the other three models predict well in the Paris region but not in the threshold region. Furthermore, the SR model can predict the FCG rates in the threshold region well. In general, as the r approaches 1, the global model exhibits better global prediction performance. For 850 groups of FCG test data with different R -ratios, the evaluation parameters r of the four prediction models are 0.9921 (SR model), 0.9771 (Kujawski's model), 0.9775 (Huang's model), and 0.9781 (Zhan's model). The SR model continued to exhibit the highest global prediction precision.

5 Discussion

The evaluation and comparison of previous models indicate that the proposed FCG rate prediction model based on domain knowledge-guided symbolic regression is suitable for predicting the threshold and Paris' regions with different R -ratios. The SR model does not condense the FCG test data to a constant R -ratio in the narrow band of the crack-growth rate curve. Instead, the FCG rate prediction model was built directly based on the test data. As previously demonstrated, the SR model provides a more accurate prediction in the threshold region than the three traditional semi-empirical FCG rate models.

In addition, the domain knowledge-guided symbolic regression proposed in this study can serve as a general model construction method in research on crack growth prediction. The FCG rate-prediction model has the advantage of involving fewer subjective factors. Unlike the semi-empirical FCG rate prediction model developed by researchers for the test phenomenon, the SR model is primarily driven by test data under domain knowledge guidance. This reduces human influence in the SR model and ensures the interpretability of the model within the framework of the traditional semi-empirical FCG rate model. The successful implementation of the SR model demonstrated the feasibility of domain knowledge-guided SR in the construction of FCG rate models. Owing to data-driven adaptability, domain knowledge-guided SR can develop more accurate models than the traditional FCG rate-modelling strategy based on experience and inspiration. Furthermore, because it considers the substructure of the traditional semi-empirical model, domain knowledge-guided SR can establish more interpretable models than the pure numerical regression modelling technique. In contrast, traditional ML methods are not only less explanatory, but also the training results can only be applied to the data space of the training set. In this study, the SR model was constructed by training with the Al-7055-T7511 FCG test data, and MLR was used to extend the SR model to other materials with accuracy. The FCG test data R -ratios used to train and evaluate the performance of the SR model were between $-1 \leq R < 1$, and its prediction performance for $R < -1$ requires further investigation.

Note that despite being built on the domain knowledge, the SR model is still data-driven. Therefore, more research into the physical meaning of each subtree structure is required for a better understanding of the crack growth process. Furthermore, although the present study only considers ΔK , R , and ΔK_{th} , other useful domain knowledge, such as the crack closure factor f may provide guidance for extending the application scope of the SR model. Other methods, such as XGboost, can be used to examine and select the importance of features as the number of characteristics increases, thereby reducing the spatial dimensions of the model.

6 Conclusions

(1) The proposed domain knowledge-guided SR obtained the variable subtree required for SR construction by analysing traditional semi-empirical FCG rate models. SR based on the variable subtree could balance the accuracy and interpretability of the data-driven model. This method provides a new direction for research on FCG.

(2) The SR model established in this work considered the comprehensive relationship between ΔK , R , and

ΔK_{th} , and the prediction equation had a concise mathematical structure. The model was acquired based on the AI-7055-T7511 FCG test data and could be extended to other materials using MLR. The prediction curve of the SR model had a good correlation with the test results.

(3) In comparison to the other three traditional semi-empirical FCG models, the SR model exhibited a more accurate prediction performance in both the threshold and Paris' regions. Overall, to seven materials in the study, the average MSE of the three conventional models was about 0.5, while the average MSE of the SR model was only 0.171, a more than 60% reduction. These results highlight the reliability of the SR model for predicting the FCG rate.

Acknowledgements

Not applicable.

Author Contributions

SZ wrote the manuscript; BY was in charge of the whole trial; SZ and BY carried out the numerical simulations and analyses. SX, GY and TZ guided the numerical simulations. All authors read and approved the final manuscript.

Authors' Information

Shuwei Zhou, born in 1997, is currently a master candidate at State Key Laboratory of Traction Power, Southwest Jiaotong University, China.

Bing Yang, born in 1979, is currently a professor at State Key Laboratory of Traction Power, Southwest Jiaotong University, China. He received his Ph.D. degree from Southwest Jiaotong University, China, in 2011. His research interests include strength of vehicle structure and fatigue and fracture of materials.

Shoune Xiao, born in 1964, is currently a professor at State Key Laboratory of Traction Power, Southwest Jiaotong University, China. He received his master degree from Southwest Jiaotong University, China, in 1988. His research interests include vehicle dynamics, collision, structural strength and fatigue reliability.

Guangwu Yang, born in 1977, is currently a professor at State Key Laboratory of Traction Power, Southwest Jiaotong University, China. He received his Ph.D. degree from Southwest Jiaotong University, China, in 2005. His research interests include vehicle dynamics, vibration, and fatigue reliability.

Tao Zhu, born in 1984, is currently an associate professor at State Key Laboratory of Traction Power, Southwest Jiaotong University, China. He received his Ph.D. degree from Southwest Jiaotong University, China, in 2012. His research interests include, vehicle dynamics, collision.

Funding

Supported by Sichuan Provincial Science and Technology Program (Grant No. 2022YFH0075), Opening Project of State Key Laboratory of Performance Monitoring and Protecting of Rail Transit Infrastructure (Grant No. HJGZ2021113), Independent Research Project of State Key Laboratory of Traction Power (Grant No. 2022TPL_T03).

Competing Interests

The authors declare no competing financial interests.

Received: 5 November 2022 Revised: 30 January 2023 Accepted: 9 March 2023

Published online: 27 March 2023

References

- [1] K Hectors, W Waele. An X-FEM based framework for 3D fatigue crack growth using a B-spline crack geometry description. *Engineering Fracture Mechanics*, 2022, 261: 108238.
- [2] M Koyama, Z Zhang, M Wang, et al. Bone-like crack resistance in hierarchical metastable nanolaminate steels. *Science*, 2017, 355(6329): 1055-1057.
- [3] D Bang, A Ince, M Noban. Modeling approach for a unified crack growth model in short and long fatigue crack regimes. *International Journal of Fatigue*, 2019, 128: 105182.
- [4] H Liu, X Yang, S Li, et al. Modeling fatigue crack growth for a through thickness crack: An out-of-plane constraint-based approach considering thickness effect. *International Journal of Mechanical Sciences*, 2020, 178: 105625.
- [5] L Xu, K Wang, X Yang, et al. Model-driven fatigue crack characterization and growth prediction: A two-step, 3-D fatigue damage modeling framework for structural health monitoring. *International Journal of Mechanical Sciences*, 2021, 195: 106226.
- [6] Z Jing, X Wu. Wide-range weight functions and stress intensity factors for arbitrarily shaped crack geometries using complex Taylor series expansion method. *Engineering Fracture Mechanics*, 2015, 138: 215-232.
- [7] A Fahem, A Kidane, M Sutton. Geometry factors for Mode I stress intensity factor of a cylindrical specimen with spiral crack subjected to torsion. *Engineering Fracture Mechanics*, 2019, 214: 79-94.
- [8] P Paris, F Erdogan. A critical analysis of crack propagation laws. *Journal of Basic Engineering*, 1963, 85(4): 528-533.
- [9] W Elber. *The significance of fatigue crack closure*. ASTM STP, 1971: 230-243.
- [10] S Mettu, V Shivakumar, J Beek, et al. NASGRO 3.0 - a software for analyzing aging aircraft. *The Second Joint NASA/FAA/DoD Conference on Aging Aircraft*, 1999: 792-801.
- [11] J Newman. A crack opening stress equation for fatigue crack growth. *International Journal of Fracture*, 1984, 24(4): R131-R135.
- [12] R Forman, V Kearney, R Engle. Numerical analysis of crack propagation in cyclic-loaded structures. *Journal of Basic Engineering*, 1967, 89(3): 459-463.
- [13] D Kujawski. A new $(\Delta K^+ K_{max})^{0.5}$ driving force parameter for crack growth in aluminum alloys. *International Journal of Fatigue*, 2001, 23(8): 733-740.
- [14] X Huang, T Moan. Improved modeling of the effect of R-ratio on crack growth rate. *International Journal of Fatigue*, 2007, 29(4): 591-602.
- [15] W Zhan, N Lu, C Zhang. A new approximate model for the R-ratio effect on fatigue crack growth rate. *Engineering Fracture Mechanics*, 2014, 119: 85-96.
- [16] S Kwofie, K Mensah. Equivalent crack growth model for correlation and prediction of fatigue crack growth under different stress ratios. *International Journal of Fatigue*, 2022, 163: 107106.
- [17] H Li, S Yang, P Zhang, et al. Material-independent stress ratio effect on the fatigue crack growth behavior. *Engineering Fracture Mechanics*, 2022, 259: 108116.
- [18] H Younis, K Kamal, M Sheikh, et al. Prediction of fatigue crack growth rate in aircraft aluminum alloys using optimized neural networks. *Theoretical and Applied Fracture Mechanics*, 2022, 117: 103196.
- [19] L Zhang, X Wei. Prediction of fatigue crack growth under variable amplitude loading by artificial neural network-based Lagrange interpolation. *Mechanics of Materials*, 2022, 171: 104309.
- [20] W Zhang, Z Bao, S Jiang, et al. An artificial neural network-based algorithm for evaluation of fatigue crack propagation considering nonlinear damage accumulation. *Materials*, 2016, 9(6): 483.
- [21] Z Lian, M Li, W Lu. Fatigue life prediction of aluminum alloy via knowledge-based machine learning. *International Journal of Fatigue*, 2022, 157: 106716.
- [22] M Bartošák. Using machine learning to predict lifetime under isothermal low-cycle fatigue and thermo-mechanical fatigue loading. *International Journal of Fatigue*, 2022, 163: 107067.
- [23] M Gorji, A Pannemaecker, S Spevack. Machine learning predicts fretting and fatigue key mechanical properties. *International Journal of Mechanical Sciences*, 2022, 215: 106949.
- [24] B Zheng, T Li, H Qi, et al. Physics-informed machine learning model for computational fracture of quasi-brittle materials without labelled data. *International Journal of Mechanical Sciences*, 2022, 223: 107282.
- [25] H Wang, B Li, F Xuan. Fatigue-life prediction of additively manufactured metals by continuous damage mechanics (CDM)-informed machine learning with sensitive features. *International Journal of Fatigue*, 2022, 164: 107147.
- [26] A Raja, S T Chukka, R Jayaganthan. Prediction of fatigue crack growth behaviour in ultrafine grained Al 2014 Alloy using machine learning. *Metals*, 2020, 10(10): 1349.

- [27] H Wang, W Zhang, F Sun, et al. A comparison study of machine learning based algorithms for fatigue crack growth calculation. *Materials*, 2017, 10(5): 543.
- [28] D Nguyen-Le, Q B Tao, V Nguyen, et al. A data-driven approach based on long short-term memory and hidden Markov model for crack propagation prediction. *Engineering Fracture Mechanics*, 2020, 235: 107085.
- [29] X Ma, X He, Z C Tu. Prediction of fatigue–crack growth with neural network-based increment learning scheme. *Engineering Fracture Mechanics*, 2021, 241: 107402.
- [30] S Mortazavi, A Ince. An artificial neural network modeling approach for short and long fatigue crack propagation. *Computational Materials Science*, 2020, 185: 109962.
- [31] X Peng, S Wu, W Qian, et al. The potency of defects on fatigue of additively manufactured metals. *International Journal of Mechanical Sciences*, 2022, 221: 107185.
- [32] M Schmidt, H Lipson. Distilling free-form natural laws from experimental data. *Science*, 2009, 324(5923): 81–85.
- [33] A Singh, Z Gu, X Hou, et al. Design optimisation of braided composite beams for lightweight rail structures using machine learning methods. *Composite Structures*, 2022, 282: 115107.
- [34] S Udrescu, M Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 2020, 6(16): y2631.
- [35] L Gan, H Wu, Z Zhong. Integration of symbolic regression and domain knowledge for interpretable modeling of remaining fatigue life under multistep loading. *International Journal of Fatigue*, 2022, 161: 106889.
- [36] H Shao, F Villaescusa-Navarro, S Genel, et al. Finding universal relations in subhalo properties with artificial intelligence. *The Astrophysical Journal*, 2022, 927(1): 1–19.
- [37] M Ziatdinov, Y Liu, A Morozovska, et al. Hypothesis learning in automated experiment: Application to combinatorial materials libraries. *Advanced Materials*, 2022, 34(20): 2201345.
- [38] K Matchev, K Matcheva, A Roman. Analytical modeling of exoplanet transit spectroscopy with dimensional analysis and symbolic regression. *The Astrophysical Journal*, 2022, 930(1): 1–13.
- [39] B Weng, Z Song, R Zhu, et al. Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts. *Nature Communications*, 2020, 11(1): 1–8.
- [40] J Song, J Kang, J Koo. Proposal of modified (normalized) ASTM offset method for determination of fatigue crack opening load. *International Journal of Fatigue*, 2005, 27(3): 293–303.
- [41] S Dinda, D Kujawski. Correlation and prediction of fatigue crack growth for different R -ratios using K_{max} and ΔK^+ parameters. *Engineering Fracture Mechanics*, 2004, 71(12): 1779–1790.
- [42] D Kujawski. A fatigue crack driving force parameter with load ratio effects. *International Journal of Fatigue*, 2001, 23: 239–246.
- [43] A Noroozi, G Glinka, S Lambert. A study of the stress ratio effects on fatigue crack growth using the unified two-parameter fatigue crack growth driving force. *International Journal of Fatigue*, 2007, 29(9–11): 1616–1633.
- [44] C Chen, D Ye, L Zhang, et al. Effects of tensile/compressive overloads on fatigue crack growth behavior of an extra-low-interstitial titanium alloy. *International Journal of Mechanical Sciences*, 2016, 118: 55–66.
- [45] W Zhang, Q Wang, X Li, et al. A simple fatigue life prediction algorithm using the modified NASGRO equation. *Mathematical Problems in Engineering*, 2016: 1–8.
- [46] J Newman. Fatigue-life prediction methodology using a crack-closure. *Journal of Engineering Materials and Technology*, 1995, 117(4): 433–439.
- [47] J Newman, E Phillips, M Swain. Fatigue-life prediction methodology using small-crack theory. *International Journal of Fatigue*, 1999, 21(2): 109–119.
- [48] R Ritchie. Mechanisms of fatigue-crack propagation in ductile and brittle solids. *International Journal of Fracture*, 1999, 100: 55–83.
- [49] M Meyers, K Chawla. *Mechanical behavior of materials*. Cambridge, England: Cambridge University Press, 2008.
- [50] M Cranmer, S Alvaro. Discovering symbolic models from deep learning with inductive biases. *ARXIV preprint ARXIV*. 2020, 2006.11287.
- [51] J Craven, V Jejjala, A Kar. Disentangling a deep learned volume formula. *Journal of High Energy Physics*, 2021, 2021(6): 1–39.
- [52] M Zhu, F Xuan, S Tu. Effect of load ratio on fatigue crack growth in the near-threshold regime: A literature review, and a combined crack closure and driving force approach. *Engineering Fracture Mechanics*, 2015, 141: 57–77.
- [53] B Qiu, M Zhang, X Li, et al. Unknown impact force localisation and reconstruction in experimental plate structure using time-series analysis and pattern recognition. *International Journal of Mechanical Sciences*, 2020, 166: 105231.
- [54] Y Huang, X Ye, B Hu, et al. Equivalent crack size model for pre-corrosion fatigue life prediction of aluminum alloy 7075-T6. *International Journal of Fatigue*, 2016, 88: 217–226.
- [55] S Jha, K Ravichandran, S Univ. Effect of mean stress (stress ratio) and aging on fatigue-crack growth in a metastable beta titanium alloy, Ti-10V-2Fe-3Al. *Metallurgical and Materials Transactions A*, 2000, 31(3): 703–714.
- [56] R Ritchie, B L Boyce, J P Campbell, et al. Thresholds for high-cycle fatigue in a turbine engine Ti–6Al–4V alloy. *International Journal of Fatigue*, 1999, 21(7): 653–662.
- [57] J Bulloch. Near threshold fatigue crack propagation behaviour of CrMoV turbine steel. *Theoretical and Applied Fracture Mechanics*, 1995, 23(1): 89–101.
- [58] X Wu, J Newman, W Zhao, et al. Small crack growth and fatigue life predictions for high-strength aluminium alloys: Part I—experimental and fracture mechanics analysis. *Fatigue & Fracture of Engineering Materials & Structures*, 1998, 21(11): 1289–1306.
- [59] P Paris, H Tada, J Donald. Service load fatigue damage—a historical perspective. *International Journal of Fatigue*, 1999, 21: 35–46.

Submit your manuscript to a SpringerOpen[®] 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)