Physics-Informed Surrogate Modeling for Hydro-Fracture Geometry Prediction Based on Deep Learning

Yutian Lu[a,*], Bo Wang[b,*], Yingying Zhao[a,**], Xiaochen Yang[c], Lizhe Li[b], Mingzhi Dong[a], Qin Lv[d], Fujian Zhou[b], Ning Gu[a] and Li Shang[a]

[a]School of Computer Science, Fudan University, Shanghai, China, Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China, and Shanghai Institute of Intelligent Electronics & Systems, Shanghai, China.

[b]Petroleum College, China University of Petroleum-Beijing at Karamay, Karamay, Xinjiang, China.

[c]School of Mathematics and Statistics, University of Glasgow, Glasgow, UK.

[d]University of Colorado Boulder, Boulder, CO, USA.

ARTICLE INFO

Keywords:
hydro-fracture geometry
physics-informed
deep learning
data efficiency
interpretability

ABSTRACT

Hydro-fracture geometry prediction is of great practical importance for optimizing construction parameters and evaluating stimulation effects. Existing physical simulation methods are computationally intensive. Deep learning-based methods offer fast model inference, yet typically require a large amount of field data for accurate model training and lack model interpretability in explaining the complex physical processes. This work presents a physics-informed surrogate modeling method for hydro-fracture geometry prediction. The proposed method encodes the hydro-fracture physical laws, in the form of partial differential equations, as a loss term to govern the training process of the surrogate model, aiming to alleviate the data requirement for model training. Experimental studies demonstrate that the proposed modeling method effectively reduces the training data requirement and improves model accuracy and interpretability.

1. Introduction

The unconventional hydrocarbon resources, such as shale oil/gas, tight oil/gas and coal bed gas, cannot naturally produce oil/gas due to low reservoir permeability. Hydraulic fracturing is an indispensable technique for economically exploiting these resources through creating hydro-fractures and increasing the contact area between the wellbore and the reservoir [20]. Hydraulic fracturing is a high-cost process during oil and gas well development, as nearly ten thousand cubes of liquid and one thousand cubes of proppant need to be injected into the reservoir to create hydraulic fractures in each horizontal well. Thus hydro-fracture geometry (HFG) prediction is of great practical significance for optimizing construction parameters and evaluating stimulation effects [1]. The injection process during hydraulic fracturing and the HFG are illustrated in Figure 1.

HFG refers to the length, width, and height of hydraulic fracture. The length of hydraulic fracture determines the stimulated volume after hydraulic fracturing. Longer fractures increase the oil/gas production rate. On the other hand, longer fractures increase the fluid volume and raise the fluid cost. Therefore, accurate prediction of fracture length is of great value for the optimization of injection fluid. The width of hydraulic fracture determines the difficulty in transporting proppant. For fractures with small widths, proppant tends to bridge and settle down, which largely increases the construction risk of high injection pressure and decreases the conductivity along with the entire hydraulic fracture.

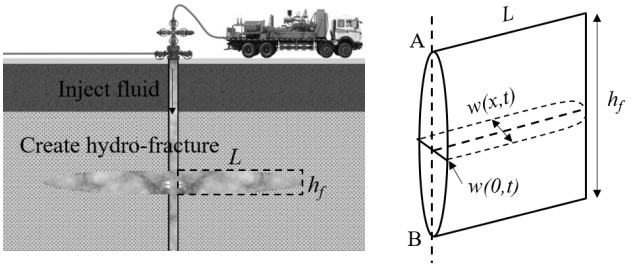


Figure 1: The *left* shows the fluid injection process during hydraulic fracturing in practical engineering, and the *right* shows the characteristics of a hydro-fracture. L denotes the fracture half-length, h_f denotes the constant height, and $w(x,t)$ denotes the width of the hydro-fracture corresponding to the distance along the fracture x at injection time t.

Thus, accurate prediction of fracture width is crucial to the optimization of proppant concentration. Effort on HFG prediction has been focusing on physical simulation and machine learning methods.

1.1. Physical Simulation Methods

Physical simulation methods for HFG prediction mainly include classical fracture models and modern numerical methods. Perkins and Kern [19] and Nordgren [17] were the first to propose the PKN model to predict the length and width of hydraulic fractures with injection time as a variable. Geertsma and De Klerk [8] proposed the KGD model to predict HFG at the early injection time. PKN and KGD establish the theoretical basis of the hydraulic fracturing process simulation and open the door to simulation methods. Chen et al. [6] applied the finite element method (FEM) to

*Equal contribution
**Corresponding author
yingyingzhao@fudan.edu.cn (Y. Zhao)
ORCID(s): 0000-0001-5902-1306 (Y. Zhao)

predict the HFG when a hydraulic fracture interacts with a natural fracture. FEM simulates hydro-fracture propagation through setting cohesive elements or re-meshing, thus the propagation path is pre-defined or the computational cost is prohibitively high. Wang et al. [28] applied the extended finite element method (XFEM) to predict the HFG when several hydraulic fractures propagate simultaneously. XFEM can simulate hydro-fracture propagation along arbitrary paths free from the mesh, but it cannot deal with complex fractures due to the computational cost. Wu and Olson [29] applied the boundary element method (BEM) to predict the complex HFG in naturally fractured reservoirs. When using BEM, only the fracture paths need to be discretized and the fracture element number determines the dimension of the solution equations. The drawback of BEM is that it can only deal with a homogeneous reservoir. Zou et al. [31] applied the discrete element method (DEM) to predict the HFG in the reservoirs containing weak bedding planes. Based on DEM, the formation is an assembly of bonded particles or deformable blocks which is interfaced by several sets of joints, and hydro-fracture growths are not allowed beyond the joint trajectories.

Existing physical simulation methods can predict HFG accurately. However, they suffer from high computation cost due to the large scale of the simulation problem as well as the complexity of the physical process, e.g., fluid flow, fluid partition, rock deformation, stress shadow effects, and fracture propagation. For example, the XFEM method can simulate hydro-fracture propagation along an arbitrary path free from the mesh, which requires several hours of simulation time to predict the geometry of a single 2D field-scale hydro-fracture. When more complex fracture structure is considered, the simulation time further increases by one or two orders of magnitude. Such high computation cost becomes even more unbearable for supporting infield hydraulic fracturing operation, where onsite engineers often need to adjust the construction parameters on the fly. In summary, how to reduce the computation cost is a fundamental challenge when predicting HFG.

1.2. Machine Learning Methods

The application of machine learning methods in hydraulic fracturing has drawn significant attention in the recent past. Deep neural network (DNN) is known for its universal approximation and fast inference capabilities, offering a promising alternative to support efficient hydraulic fracturing modeling, analysis and prediction [12, 11, 7, 4, 2]. Tamez et al. [26] utilized a DNN to timely predict the problem of a screen out and assist in making intelligent decisions throughout each stage. Their work is based on large sets of data as each monitored variable on a well is measured every second. Ben et al. [5] adopted the idea of continual learning to provide real-time prediction of wellhead pressure during hydraulic fracturing, which assists engineers in monitoring and optimizing the pumping schedule. For each hydraulic fracturing stage, the first model is trained on the data from the first couple of minutes and used to predict the wellhead pressure for the next several minutes; the second model is then trained after adding the data from the next couple of minutes and again used to predict the pressure for the next several minutes; the process continues until there is no more data. The best model among them is selected for prediction. Lapin et al. [11] used a DNN to predict the fracture aspect ratio (the ratio of the larger half-axis of the fracture to the smaller one) during hydraulic fracturing. The network is trained on a large amount of training samples generated from numerical simulation. Kim et al. [13] used a DNN to determine completion methods and fracturing fluids in shale gas reservoirs. The network architecture is designed by employing three types of training algorithms and adjusting the number of hidden layers and hidden layer neurons. Wang et al. [27] predicted the first-year well production after hydraulic fracturing by using four supervised learning approaches, namely random forest, adaptive boosting, SVM, and DNN. He et al. [10] built a spatial-temporal database from field data including information on reservoir characteristics, completion, stimulation, and production, and used four DNNs to model different hydraulic fracturing design scenarios and predict well production. One common drawback of machine learning methods is the lack of interpretability. It is difficult to analyze how the output of these methods changes when varying the physical parameters of the hydraulic fracturing process, which consequently hamper their applicability for onsite construction.

Hybrid modeling improves interpretability by integrating physical models and machine learning methods. Bangi et al. [3] proposed a deep hybrid model. Specifically, a DNN is used to predict the unknown fluid leak-off rate and the predicted parameter is then used in three physical models describing three subprocesses of the hydraulic fracturing process. While the method is interpretable, it requires a large amount of process data to train an accurate DNN.

In general, another drawback of existing machine learning methods is that these surrogate models are all completely data-driven and require a large number of training samples. However, the field data of hydraulic fracturing is expensive and even scarce for a given oil field. New methods should be proposed to train a reliable and robust surrogate model from a small data set.

1.3. Physics-Informed Neural Network Methods

In summary, accurate prediction of HFG is of great value for hydraulic fracturing optimization and until now it mainly relies on time-consuming physical simulation methods. Recently, machine learning methods demonstrate great potential, but most of them, in particular DNNs, require a large amount of field data and lack interpretability.

Recent advances in physics-informed neural networks offer the capability to embed domain knowledge, in the form of a set of PDEs, into the loss of the neural network using automatic differentiation. They potentially help reduce the required amount of training data and improve model interpretability. Raissi et al. [22] encoded the physical information in the form of PDEs as a loss term into the DNN

and demonstrated through experiments that the proposed method, termed PINN, can reduce the data required for training. Meanwhile, PINN is interpretable owing to the inclusion of physical information. Since then, PINN has been adopted to solve different types of PDEs. Guo et al. [18] extended PINNs to solve space-time fractional advection-diffusion equations. Lu et al. [16] improved the computational efficiency of PINNs and developed a Python library for PINNs applicable to forward and inverse problems in computational science and engineering. However, the library can only solve simple physical problems associated with PDEs. Recent researches on PINNs are exploring applications in the complex engineering field. Kissas et al. [14] employed PINN for cardiovascular flows modeling and constrained the output of DNN to satisfy the physical conservation laws by using one-dimensional models of pulsatile blood flow. Zhang et al. [30] applied PINNs for solving inverse identification problems of nonhomogeneous materials in elasticity imaging. Two DNNs are used for solving the forward problem and inferring the distribution of material parameters, respectively. In this work, we propose a PINN-based method to tackle the HFG prediction problem. Specifically, the proposed model encodes the physical principles of the hydraulic fracturing process described in the form of PDEs into DNN.

This work presents a physics-informed surrogate modeling method to support HFG modeling, analysis and prediction. The proposed work aims to address the aforementioned modeling challenges, i.e., the high computation cost of physical simulation, the need of a large amount of training data, and the lack of interpretability of DNN-based surrogate modeling methods. The proposed work is motivated by recent research on physics-informed neural network (PINN) [22, 25, 21]. Recent studies have shown that PINN can learn and encode underlying physical laws described by a set of partial differential equations (PDEs), thereby reducing the data requirement for model training [22, 25]. In this work, we devise a multilayer fully-connected neural network to approximate the dynamic process of hydraulic fracturing as a function of the reservoir and construction parameters. We introduce PKN, a PDE-based classical hydraulic fracturing physical model, as a loss term of the physics-informed surrogate model, driving network parameter learning towards satisfying the PDE conditions during model training. Experimental studies demonstrate that introducing the physical laws into the surrogate model effectively reduces the requirement of pre-labeling data for training, and improves the model accuracy and interpretability.

In summary, this work makes the following contributions:

1. To the best of our knowledge, we are the first to present a physics-informed surrogate modeling method based on physics-informed neural network for HFG modeling, analysis, and prediction. Compared with existing computation-intensive physical simulation methods, the proposed surrogate modeling method significantly reduces the computation cost. More importantly, the proposed method encodes domain knowledge as the PDE loss to facilitate model learning. The PDE loss regularizes the model learning towards satisfying the underlying physical knowledge during training. Thus it is more than existing DNN data-efficiency without PDE loss.
2. Experimental results show the proposed surrogate modeling method achieves the best performance among the baseline methods when using the same training samples. Also, the proposed method is more robust to the variation of the training sample size. Specifically, when the training sample size is 50, the scaled mean squared error ($sMSE$) of the proposed method is 83.0% lower than support vector machine (SVM) and 24.6% lower than DNN; when the sample size decreases to 5, the performance gain further increases to 92.3% for SVM and 46.3% for DNN.
3. This work pays special attention to model interpretability by introducing key HFG parameters as the input of the surrogate model, aiming to support onsite experts to conduct efficient and accurate infield exploration and optimization of construction during hydraulic fracturing.

The remainder of this paper are organized as follows. Section 2 covers the data generation method used in this study, and the dataset accuracy validation. Section 3 presents our proposed method for HFG prediction. Section 4 presents the experiments and results. Finally, section 5 concludes this work.

2. PKN and Accuracy Validation

This work applies the PKN model, a classical fracture model, to generate the ground truth data samples due to its low computation cost. And a numerical model with FEM is established to verify the accuracy of the PKN model and the reliability of the data set.

2.1. PKN Model

The PKN model [17] is a classical 2D model, founded on physical principles, and considers core components of hydraulic fracturing processes, such as the fluid flow within the hydro-fracture, the rock deformation, and the hydro-fracture propagation. Without loss of generality, the following discussion focuses on the fracture width, a key attribute of HFG.

Generally, there is an inter layer above and below the target fracturing layer. The height of hydraulic fracture nearly equals the thickness of the target layer. The PKN model is built on the assumption that a vertical hydraulic fracture propagates in a lateral direction with a constant height h_f, as shown in Figure 1. According to the model, at the injection time t, the width of the hydro-fracture w(x,t) corresponding to the distance along the fracture x can be defined by Eq. (1):

$$w(x,t) = w(0,t)\left\{ \frac{x}{L(t)} \sin^{-1} \frac{x}{L(t)} + \left[1 - \left(\frac{x}{L(t)}\right)^2\right]^{\frac{1}{2}} - \frac{\pi}{2}\frac{x}{L(t)} \right\}^{\frac{1}{4}}, \tag{1}$$

where $w(0,t)$ denotes the width of the hydro-fracture when x equals to 0 and is defined by Eq. (2):

$$w(0,t) = 4\left(\frac{\mu q^2}{\pi^3 E' C_L h_f}\right)^{\frac{1}{4}} t^{\frac{1}{8}}; \tag{2}$$

$L(t)$ denotes the fracture half-length and is a function of t given by

$$L(t) = \frac{q t^{\frac{1}{2}}}{2\pi C_L h_f}; \tag{3}$$

μ is fluid viscosity, q is the flow rate, C_L is the leakoff coefficient, and $E^{'}$ is the plane strain modulus defined by Eq. (4), where E is Young's modulus and υ is the Poisson's ratio:

$$E' = \frac{E}{1-\upsilon^2}. \tag{4}$$

Eq. (5) defines the PDE that represents the dynamic variation of the hydro-fracture width:

$$f(x,t) := \frac{\partial w}{\partial t} - \lambda_1 \frac{\partial^2 w^4}{\partial x^2} + \lambda_2, \tag{5}$$

where $\lambda_1 = \frac{E'}{128\mu h_f}$ and $\lambda_2 = \frac{8C_L}{\pi\sqrt{t}} \cos\left(\frac{x}{L(t)}\right)$.

The initial and boundary conditions of the PDE are defined by Eq. (6) and Eq. (7), respectively:

$$w(x,0) = 0, \tag{6}$$

$$\begin{cases} w(x,t) = 0, \text{ for } x > L(t), \\ \left.\frac{\partial w^4}{\partial x}\right|_{x=0} = \frac{128(1-\upsilon)q}{\pi G}, \end{cases} \tag{7}$$

where G is the shear modulus and its relationship to $E^{'}$ and υ is defined by Eq. (8):

$$G = \frac{E'(1-\upsilon)}{2}. \tag{8}$$

2.2. Dataset Accuracy Validation

Now, we verify how accurate the PKN model is in predicting HFG. This work uses the FEM method to prove the accuracy of the surrogate modeling because FEM is one of the modern numerical methods commonly used to predict the ideal and indoor HFG. Compared with other modern numerical methods, including DEM, BEM, and XFEM, FEM is the most accurate method when predicting HFG because of its high capabilities in solving non-linear and fluid-solid coupling problems. However, the FEM method cannot be directly used to generating the ground truth data samples when predicting the fracture width under varying fracture length conditions. Establishing these FEM hydro-fracture models is technically demanding because simulation results are sensitive to the element size and element type. Also, the astringency of the model is highly influenced by the boundary and initial conditions of the model. Therefore, we use the FEM model to predict the fracture width when varying the fracture length and observe its predicted results. If the PKN model predicts similar results, then it can be regarded as an accurate model.

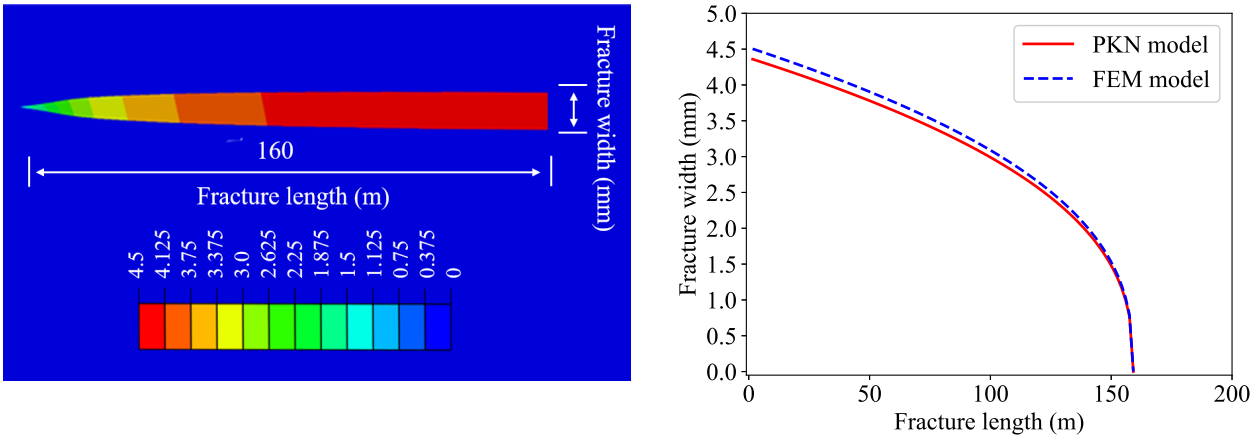
For this purpose, a FEM model with a single hydraulic fracture is established. In general, the reservoir properties can be characterized clearly through well logging technology. The target layer is rich in oil/gas resources. Field engineers fractured the target layer and developed the oil/gas resources. During the process of formation deposition, the interlayers were formed, which can protect the oil/gas from dissipating. Therefore, hydro-fracture was vertically restricted by two interlayers during lateral propagation. This work further considers the following hypothesis or scenario setting: the target layer is between a top layer and a bottom layer; the height of the hydro-fracture equals the thickness of the target layer; the rock of the target layer generates pure elastic deformation; total injection rate and fluid leak-off rate remain constant; the injection time should be long enough to generate a hydro-fracture with a large length. Figure 2a shows the prediction results from the FEM model. When the fracture length increases from $0m$ to $160m$, the fracture width decreases from $4.5mm$ to $0mm$; the value of fracture width is indicated by the color grid at the bottom of the figure. Figure 2b presents the prediction results from both models. We can see that the general trend of the PKN model is consistent with the FEM model. Also, the difference between the two models is small. For example, when the fracture length is $100m$, the FEM predicts the fracture width to be $3.08mm$ and the PKN model predicts the value to be $2.99mm$. This suggests that the PKN model is accurate in predicting the HFG and simulating from it can generate high-quality data.

3. Methodology

This section presents the proposed method for HFG prediction. It first introduces the overall architecture of the proposed method. Then, it presents a general framework for integrating the physics laws into a DNN-based surrogate model. Next, it presents a concrete physics-informed surrogate model design for predicting a specific HFG attribute. Finally, it details the network architecture and training process of the proposed model.

3.1. Architecture Overview

Figure 3 shows the overall architecture of the proposed DNN-based physics-informed surrogate modeling method for HFG prediction. It consists of a DNN-based surrogate



(a) FEM model prediction results

(b) Variation of fracture width along fracture length

Figure 2: (a) Graphical representation of prediction results from the FEM model; (b) Comparison of prediction results between the FEM model and the PKN model.

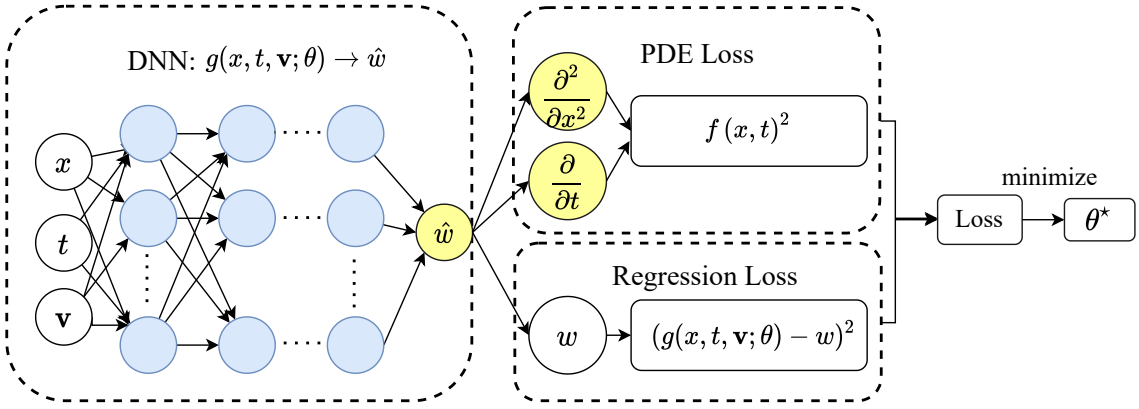


Figure 3: Architecture of the proposed model.

model to approximate the functional relationship between input and output of the simulated data. The surrogate model is enforced to satisfy the governing physical laws of the hydraulic fracturing process, which is achieved by constructing a loss term relating to the PDEs of the HFG problem. The network training process is driven by a ground truth dataset collected from highly accurate yet time-consuming physical model-based numerical solvers collocated with sampling methods, and aims to minimize the loss function on the approximation error and PDE residuals.

In summary, the proposed surrogate model for predicting HFG consists of the following steps:

1. Collect the ground truth data using highly accurate yet time-consuming physical model-based numerical solvers, and collocation data points using Latin hypercube sampling [24].
2. Build a surrogate model based on DNN and enforce it to satisfy the physical laws of the hydraulic fracture process described in the form of PDEs.
3. Train the surrogate model.

The collocation data generated in the first step are used to compute the PDE loss, which will be explained in Section 3.3. The adopted Latin hypercube sampling technique [24] is a sampling method which can be used to generate input values in the sample space and estimate expectations of functions of output variables.

3.2. DNN-based Physics-Informed Surrogate Model

We formulate HFG physical laws using nonlinear PDEs $f(x,t)$ parametrized by $\mathbf{v}$ with the latent solution $w(x,t)$ as in Eq. (9) [22]:

$$f(x,t) := \frac{\partial w}{\partial t} + \mathcal{D}(w;\mathbf{v}), \tag{9}$$

where w is an attribute of HFG to be predicted, which depends on the distance along the fracture x, the injection time t, and the reservoir and construction parameters $\mathbf{v}$; $\mathcal{D}(w;\mathbf{v})$ denotes differential operator consisting of the potential solutions $w(x,t)$ and parameters $\mathbf{v}$. As w can be represented by x, t and its partial derivatives, Eq. (9) can be

equivalently written as follows [22]:

$$f(x,t;\frac{\partial w}{\partial x},\frac{\partial w}{\partial t},\frac{\partial^2 w}{\partial x^2},\cdots,\mathbf{v})=0 \tag{10}$$

with boundary conditions $\mathcal{B}(w,x,t)=0$ and initial conditions $\mathcal{I}(w,x,0)=0$.

Since there is no analytical solution for w and numerical solvers are time-consuming, we propose to employ DNN to solve w for its fast inference capabilities. We adopt a fully-connected multi-layer neural network as the surrogate model to approximate the functional relationship between w and x, t, $\mathbf{v}$:

$$\hat{w}=g(x,t,\mathbf{v};\theta), \tag{11}$$

where g denotes the DNN, and θ denotes parameters of the network.

3.3. HFG Prediction via Physics-Informed Surrogate Model

We now give a specific example of HFG prediction by using the DNN-based physics-informed surrogate model. Sp-ecifically, the fracture width w is the attribute of HFG to be predicted, which, as explained before, depends on the distance along the fracture x, the injection time t, and reservior and construction parameters $\mathbf{v}$. Ideally, a well-trained network can accurately model how the fracture width w varies with the geometric and physical parameters $x,t,\mathbf{v}$.

To learn θ, we propose to minimize the following loss function:

$$Loss=\frac{1}{N_w}\sum_{i=1}^{N_w}(g(x_w^i,t_w^i,v_w^i;\theta)-w^i)^2+\frac{1}{N_f}\sum_{i=1}^{N_f}f(x_f^i,t_f^i)^2, \tag{12}$$

where $\{x_w^i,t_w^i,v_w^i\}_{i=1}^{N_w}$ denotes the initial and boundary data generated from the HFG numerical solver [17], and $\{x_f^i,t_f^i\}_{i=1}^{N_f}$ denotes the collocations data points generated using Latin hypercube sampling. The first term of Eq. (12) is the commonly used mean square error (MSE) for encouraging a good data fit. The second term is a newly introduced MSE derived from the PDE $f(x,t)$ (Eq. (10)). As $f(x,t)$ describes the physical processes of HFG, minimizing this loss will drive the model toward learning parameters that adhere to the encoded physical laws. In other words, we use prior domain knowledge to regularize the training of DNN. Consequently, the network could be learned from few training samples with good generalization ability. Furthermore, we take the reservior and construction parameters $\mathbf{v}$ as additional input to the DNN, which allows domain experts to examine the results of HFG prediction in practical engineering.

3.4. Network Architecture and Training Process

In this work, the DNN g is chosen as a k-layer fully-connected neural network architecture with n neurons per hidden layer ($k=6$, $n=300$ in the experimental settings). The L-BFGS [15] algorithm is adopted for optimizing the parameters θ.

Algorithm 1 details the training process of the proposed physics-informed surrogate model.

Algorithm 1 Physics-Informed Surrogate Model Training

Require: the maximum number of iterations K, threshold τ

```
Generate N_w ground truth data using Eq. (1)
Generate N_f data points using Eq. (5)
Construct a fully-connected multi-layer neural network g(x, t, v; θ) with initialized parameters θ
iter = 0
while iter < K and Loss > τ do
    Train g(x, t, v; θ) by minimizing Loss in Eq. (12)
    iter = iter + 1
end while
return g(x, t, v; θ*) with minimized Loss.
```

4. Experiments and Results

This section describes the experimental evaluation of the proposed method.

4.1. Dataset Generation

In this work, data were generated based on the real hydraulic fracturing scenario. The values of the key parameters (i.e. the injection rate, the injection time, the fluid viscosity et, al.) were specified among the common ranges. The dynamic change of the fracture width was captured by varying the injection time while fixing the fracture position. Specially, using the PKN model, we generate the ground truth data samples. Fig. 4 shows the selected training samples and the ground truth solution $w(x,t)$ over the range of x and t. x denotes the fracture length and t represents the injection time. In this work, $t\in[3000,7200]$ and $x\in[0,L(3000)]$. The injection time of a common field fracturing ranges from 3600 seconds to 5400 seconds. We enlarge the range to $[3000,7200]$ thus fully covering all of the field cases; $L(3000)$ is the initial fracture length, and we can capture the dynamic change of the fracture width along with the initial fracture length $L(3000)$ during the following fracturing fluid injection. We randomly select N_w samples from the boundary or initial data as the training samples, marked using a cross sign; $N_w=50$ unless stated otherwise. We also generate $N_f=10,000$ collocation points using the Latin hypercube sampling technique [24] to regularize the DNN training. For the test set, we generate $21,100$ samples by meshing x and t, where $x\in[0,L(3000)]$ with an interval of $0.01L(3000)$ and $t\in[3000,7200]$ with an interval of 20.

The accuracy of the PKN model can be guaranteed when the length $L(t)$ is much greater than the height h_f. Generally, in order to obtain a hydro-fracture with a large length, the injection time t of one fracturing stage is between 3600 and 5400 s. In this work, to expand the sample space, the range of the injection time is extended to 3000 to 7200 s, i.e., $t\in$

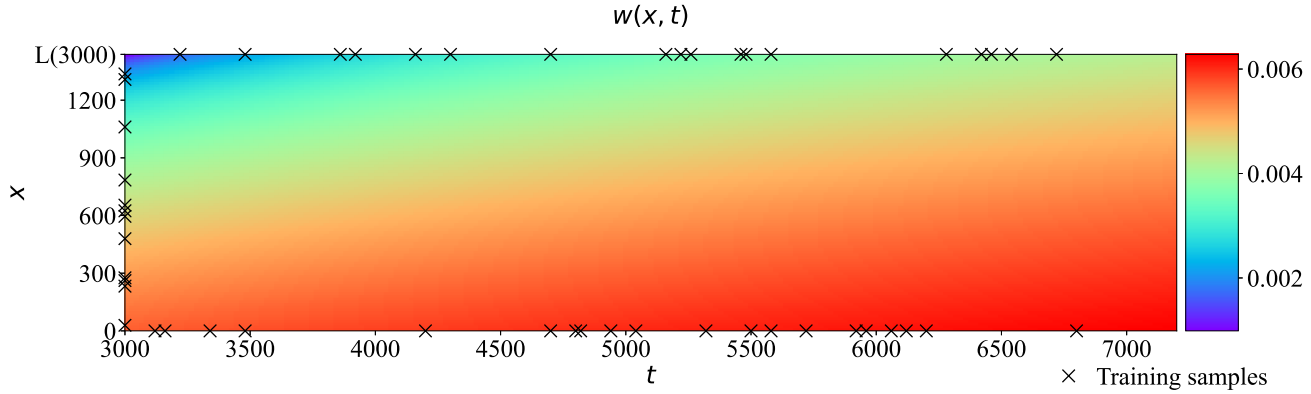


Figure 4: Visualization of training samples and solution $w(x,t)$ with varying x and t. Training samples are indicated by the cross sign, and value of $w(x,t)$ is indicated by color.

$[3000, 7200]$. Other parameters are set as: $x \in [0, L(3000)]$, $C_L = 1 \times 10^{-5}$ (with unit $m/s^{0.5}$), $E = 2 \times 10^{10}$ (with unit Pa), $\mu = 0.01$ (with unit $Pa \cdot s$), $q = 1/30$ (with unit m^3/s), $v = 0.2$, and $h_f = 20$ (with unit m).

4.2. Experiment Setting

4.2.1. Evaluation Metrics

We compare the generalization performance of different methods by using the scaled mean squared error ($sMSE$) as the evaluation criterion, which is defined as follows:

$$sMSE = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \frac{(w_i - \hat{w}_i)^2}{w_i^2}, \tag{13}$$

where w_i denotes the actual value, $\hat{w}_i$ denotes the predicted value, and N_{test} is the number of test samples. A smaller value of $sMSE$ indicates better generalization ability.

This work also investigates the robustness of the proposed surrogate model when learning from nosiy training data. The relative absolute error (RAE), defined in Eq. (14), is used as the evaluation criterion:

$$RAE = \frac{|sMSE^* - sMSE|}{sMSE}, \tag{14}$$

where $sMSE^*$ denotes the $sMSE$ obtained by training on the noise-contaminated data. RAE reflects the relative difference of the prediction accuracy when using noisy training data and the training data without noise. The smaller the RAE, the more robust the model's prediction capability. In other words, a small value of RAE indicates that the method does not deteriorate much under noise conditions and is robust.

4.2.2. Baselines

We compare the proposed method with two baseline methods: (1) a DNN-based method without the PDE-associated loss. The DNN architecture is same as that of the proposed surrogate model; and (2) SVM method. SVM [23] is a supervised learning method which has been applied to various tasks, e.g., regression and classification, and is one of the widely used methods for HFG prediction [9].

4.3. HFG Prediction Evaluation

4.3.1. Overall Performance

Fig. 5 shows the HFG prediction accuracy of the two baseline methods and the proposed surrogate modeling method under the varying number of training samples. We can see that the proposed method achieves higher prediction accuracy, i.e., lower $sMSE$, than the baseline methods for all training sample sizes. Despite the fluctuations, the relative performance gain further increases when the training sample size decreases. For instance, when the sample size decreases to 5, the proposed method makes the maximum performance gain, which improves SVM by 92.3% and DNN by 46.3%. In addition, the proposed method is more robust to the variation of the training sample number. The vertical bars in Fig. 5 indicate the standard deviation of $sMSE$ estimated from five different models trained on the randomly selected data from the ground truth data samples. The proposed method has less variation than the baselines methods for the same training sample size. Also, the proposed method is less impacted by the training sample number. For example, when the number of training samples is 10, the standard deviation of the proposed method is 0.008. In contrast, it is 0.024 and 0.055 for DNN and SVM, respectively.

The following investigation further explains why the proposed method outperforms the DNN-based method. Figures 6a, 6b, 6c, and 6d show the prediction accuracy of the proposed method and the DNN-based method when $t = 3000$, $t = 4000$, $t = 5000$, and $t = 6000$, respectively. We can see that $sMSE$ of the proposed method is lower than the baseline for most values of x. This suggests that incorporating the physical laws into DNN regularizes the network and leads to more accurate prediction.

4.3.2. Robustness Analysis

The proposed method is more robust against noisy data than the two baseline methods. To better understand the robustness of the proposed method, we evaluate the prediction accuracy of the three methods when learning from noisy training data. More specifically, we first add noise to the training data w. The noise value equals the product of the

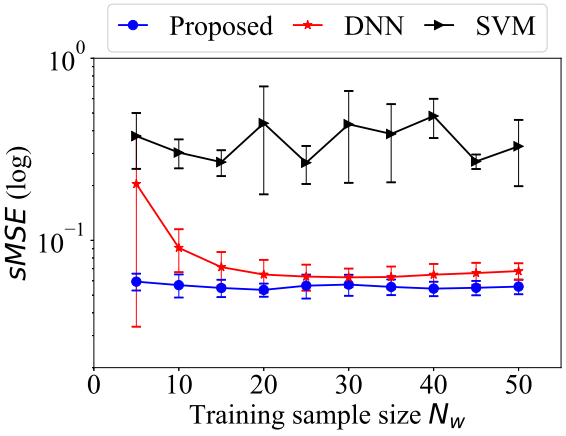


Figure 5: Variation of $sMSE$ with training sample size N_w. The vertical bars indicate the standard deviation of $sMSE$ estimated from five different trained models.

noise factor α and a random value sampled from the Gaussian distribution $\mathcal{N}(0, s^2)$, where $\alpha \in \{0.025, 0.05, \dots, 0.2\}$ and s^2 is the sample standard deviation estimated from the training samples. Then, we evaluate the robustness of the three methods regarding the above-defined metric RAE.

Fig. 7 shows how RAE varies with the noise factor α. We can see that he proposed method outperforms the baselines over the range of α, and its improvement is more pronounced when the noise factor increases. More specifically, when $\alpha = 0.2$, the RAE of the proposed method is 0.005, compared with 0.052 for DNN and 0.349 for SVM, respectively. This positive result matches the expectation as the proposed model is trained to meet the underlying physical laws described by PDEs during the training process, which makes it more robust to noise compared with the baselines trained only based on data.

4.3.3. *Interpretability Analysis*

To evaluate the interpretability, we investigate how the model performance varies with one of the most important construction parameters – the flow rate q. According to domain experts, variation in q will significantly affect the fracture width w in practical engineering. Therefore, domain experts desire a surrogate model that could accurately predict HFG under varying q. Fig 8 shows how the prediction accuracy varies with q. Given that $q \in [0.5/60, 4/60]$, we first generate the data set by varying q from $0.5/60$ to $4/60$ with an interval of $0.5/60$. Then, we select N_w samples from the boundary and midpoint of q to form the training set, i.e., $q \in \{0.5/60, 2/60, 2.5/60, 4/60\}$. The remaining values of q are used to generate the test set. The proposed method consistently outperforms the baseline DNN and SVM in terms of $sMSE$ when q changes. For example, when $q = 1/60$, i.e. 0.067, the $sMSE$ of the proposed method is 0.135. In contrast, it is 0.263 for DNN and 0.556 for SVM, respectively.

4.3.4. *Sensitivity Analysis*

To further analyze the performance of the proposed method, we conduct the following sensitivity analysis to quantify HFG prediction accuracy for different neural network architectures. Table 1 shows $sMSE$ for different number of hidden layers and different number of neurons per layer. We can observe that $sMSE$ tends to decrease as the number of hidden layers and neurons increases.

4.3.5. *Efficiency Analysis*

The proposed surrogate model is implemented on a 3.60GHz 16-core computer. The computation time is in the range of seconds. Compared with physical simulation methods that typically take hours to days, the proposed method is much more computationally efficient, making onsite hydraulic fracturing support feasible.

5. Conclusion

In this paper, we present an interpretable and data-efficient surrogate model for HFG prediction. We propose a physics informed deep neural network architecture, leveraging the physical laws in the form of PDEs to drive the model training. Experiments demonstrate that the surrogate model: obtains lower scaled mean squared error than that of DNN without the PDE-associated loss when changing construction parameters; and achieves higher prediction accuracy given fewer training samples. This work demonstrates the benefit of physics informed surrogate modeling method on data efficiency, model accuracy, and model interpretability. Future work includes the investigation of applying the physics-informed surrogate model to support hydraulic fracturing field work.

References

[1] Adachi, J., Siebrits, E., Peirce, A., Desroches, J., 2007. Computer simulation of hydraulic fractures. International Journal of Rock Mechanics and Mining Sciences 44, 739–757.

[2] Ali, A., 2021. Data-driven based machine learning models for predicting the deliverability of underground natural gas storage in salt caverns. Energy 229, 120648. URL: https://doi.org/10.1016/j.energy.2021.120648, doi:10.1016/j.energy.2021.120648.

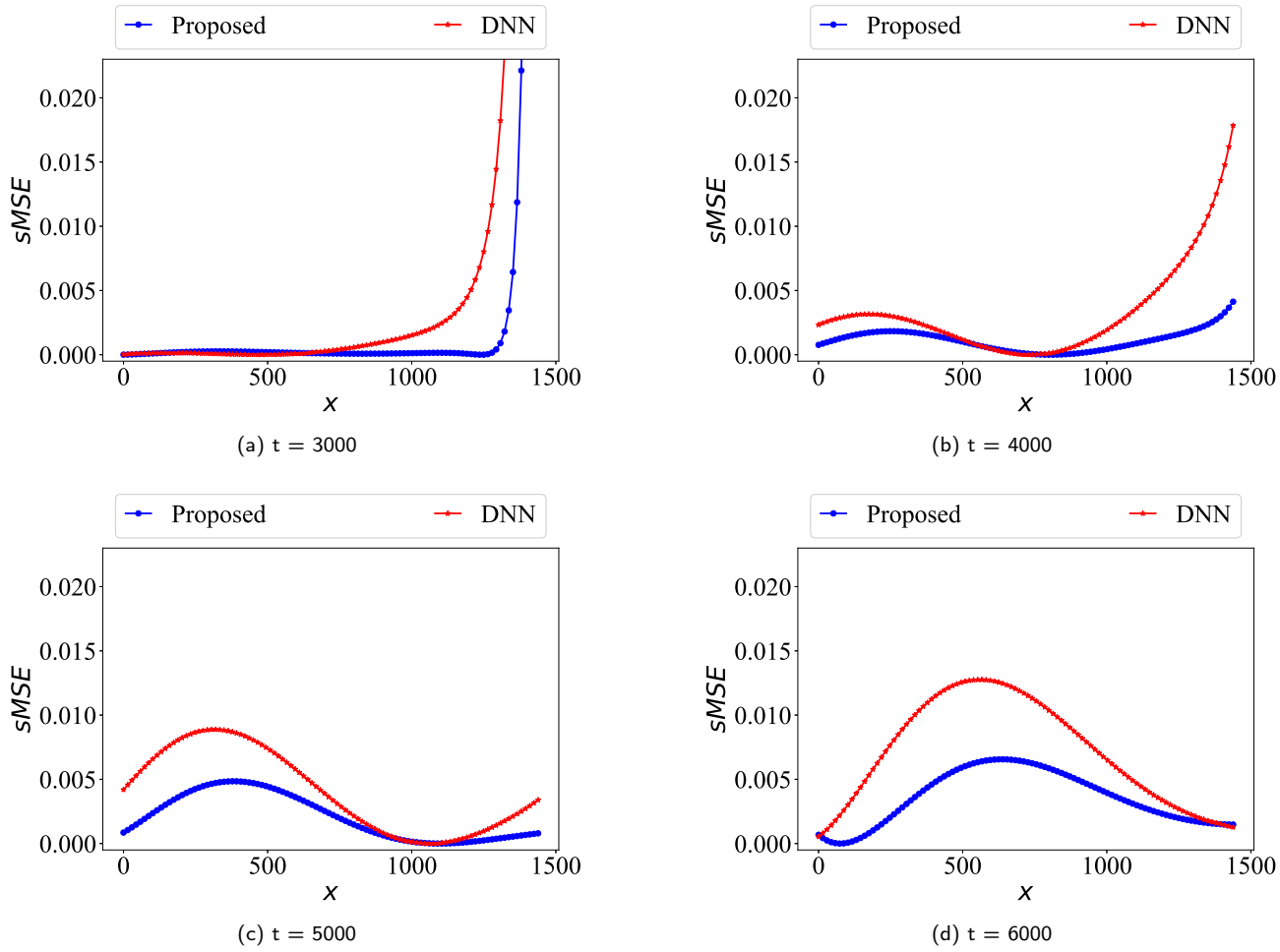
[3] Bangi, M.S.F., Kwon, J.S.I., 2020. Deep hybrid modeling of chemical process: Application to hydraulic fracturing. Computers & Chemical Engineering 134, 106696.

[4] Ben, Y., Perrotte, M., Ezzatabadipour, M., Ali, I., Sankaran, S., Harlin, C., Cao, D., et al., 2020a. Real time hydraulic fracturing pressure prediction with machine learning, in: SPE Hydraulic Fracturing Technology Conference and Exhibition, Society of Petroleum Engineers.

[5] Ben, Y., Perrotte, M., Ezzatabadipour, M., Ali, I., Sankaran, S., Harlin, C., Cao, D., et al., 2020b. Real time hydraulic fracturing pressure prediction with machine learning, in: SPE Hydraulic Fracturing Technology Conference and Exhibition, Society of Petroleum Engineers.

[6] Chen, Z., Jeffrey, R.G., Zhang, X., Kear, J., et al., 2017. Finite-element simulation of a hydraulic fracture interacting with a natural fracture. SPE Journal 22, 219–234.

[7] Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., Sun, Z., 2021. Well production forecasting based on ARIMA-LSTM model considering



(a) t = 3000 (b) t = 4000

(c) t = 5000 (d) t = 6000

Figure 6: Comparison of prediction accuracy of the proposed method (with PDE loss) and DNN (without PDE loss) at three injection time.

Table 1
Comparison of $sMSE$ under different number of hidden layers and different number of neurons per layer.

Hidden layers \ Neurons	50	100	150	200	250	300
2	2.26E-01	9.45E-02	1.72E-01	7.32E-02	8.74E-02	6.49E-02
4	1.16E-01	1.64E-01	1.11E-01	8.34E-02	6.40E-02	**4.78E-02**
6	7.93E-02	1.42E-01	7.42E-02	7.02E-02	6.22E-02	5.28E-02

manual operations. Energy 220, 119708. URL: https://doi.org/10.1016/j.energy.2020.119708, doi:10.1016/j.energy.2020.119708.

[8] Geertsma, J., de Klerk, F., 1969. A rapid method of predicting width and extent of hydraulically induced fractures. Journal of petroleum technology 21, 1571–1581.

[9] Gupta, I., Rai, C., Devegowda, D., Sondergeld, C., et al., 2018. Use of data analytics to optimize hydraulic fracture locations along borehole. Petrophysics 59, 811–825.

[10] He, Q., Zhong, Z., Alabboodi, M., Wang, G., et al., 2019. Artificial intelligence assisted hydraulic fracturing design in shale gas reservoir, in: SPE Eastern Regional Meeting, Society of Petroleum Engineers.

[11] Kalyuzhnyuk, A.V., Lapin, R.L., Murachev, A., Osokina, A., Sevostianov, A., Tsvetkov, D.V., 2019. Neural networks and data-driven surrogate models for simulation of steady-state fracture growth. Materials Physics and Mechanics 42, 351–358.

[12] Kim, C., Kim, Y., Shin, C., Lee, J., 2017a. A comprehensive approach to select completion and fracturing fluid in shale gas reservoirs using the artificial neural network. Environmental Earth Sciences 76. doi:10.1007/s12665-017-7028-4.

[13] Kim, C., Kim, Y., Shin, C., Lee, J., 2017b. A comprehensive approach to select completion and fracturing fluid in shale gas reservoirs using the artificial neural network. Environmental Earth Sciences 76, 671.

[14] Kissas, G., Yang, Y., Hwuang, E., Witschey, W.R., Detre, J.A., Perdikaris, P., 2020. Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4D flow MRI data using physics-informed neural networks. Computer Methods in Applied Mechanics and Engineering 358, 112623.

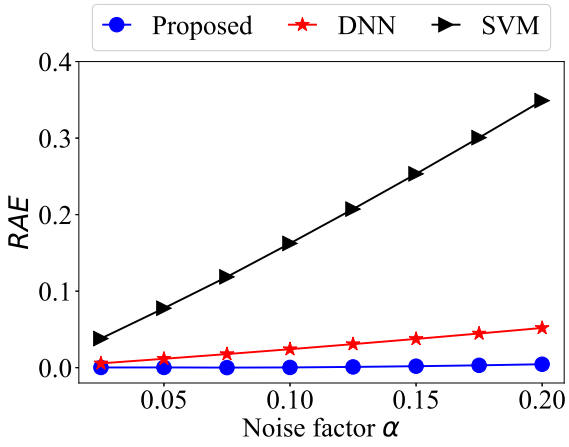


Figure 7: Variation of RAE with noise factor α.

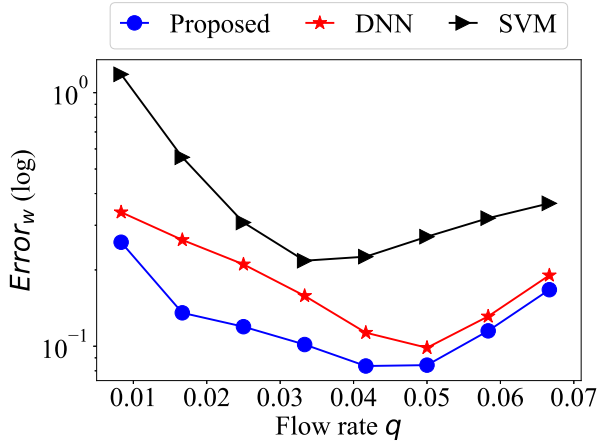


Figure 8: Variation of $sMSE$ with flow rate q.

[15] Liu, D.C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. Mathematical programming 45, 503–528.

[16] Lu, L., Meng, X., Mao, Z., Karniadakis, G.E., 2021. DeepXDE: A deep learning library for solving differential equations. SIAM Review 63, 208–228.

[17] Nordgren, R., et al., 1972. Propagation of a vertical hydraulic fracture. Society of Petroleum Engineers Journal 12, 306–314.

[18] Pang, G., Lu, L., Karniadakis, G.E., 2019. fPINNs: Fractional physics-informed neural networks. SIAM Journal on Scientific Computing 41, A2603–A2626.

[19] Perkins, T., Kern, L.R., et al., 1961. Widths of hydraulic fractures. Journal of Petroleum Technology 13, 937–949.

[20] Rahm, D., 2011. Regulating hydraulic fracturing in shale gas plays: The case of texas. Energy Policy 39, 2974–2981.

[21] Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017. Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations. CoRR abs/1711.10561. URL: http://arxiv.org/abs/1711.10561, arXiv:1711.10561.

[22] Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378, 686–707.

[23] Sain, S.R., 1996. The nature of statistical learning theory.

[24] Stein, M., 1987. Large sample properties of simulations using Latin hypercube sampling. Technometrics 29, 143–151.

[25] Sun, L., Gao, H., Pan, S., Wang, J.X., 2020. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. Computer Methods in Applied Mechanics and Engineering 361, 112732. doi:https://doi.org/10.1016/j.cma.2019.112732.

[26] Tamez, G., Castillo, D., Colmenero, A., Ayala, J.A., Bailey, C.P., 2019. Machine learning application to hydraulic fracturing, in: Big Data: Learning, Analytics, and Applications, International Society for Optics and Photonics. p. 109890A.

[27] Wang, S., Chen, S., 2019. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. Journal of Petroleum Science and Engineering 174, 682–695.

[28] Wang, X., Shi, F., Liu, C., Lu, D., Liu, H., Wu, H., 2018. Extended finite element simulation of fracture network propagation in formation containing frictional and cemented natural fractures. Journal of Natural Gas Science and Engineering 50, 309–324.

[29] Wu, K., Olson, J.E., et al., 2016. Numerical investigation of complex hydraulic-fracture development in naturally fractured reservoirs. SPE production & operations 31, 300–309.

[30] Zhang, E., Yin, M., Karniadakis, G.E., 2020. Physics-informed neural networks for nonhomogeneous material identification in elasticity imaging. arXiv preprint arXiv:2009.04525 .

[31] Zou, Y., Zhang, S., Ma, X., Zhou, T., Zeng, B., 2016. Numerical investigation of hydraulic fracture network propagation in naturally fractured shale formations. Journal of Structural Geology 84, 1–13.