

D. Suleimanov *, D. Stolyarov , Y. Tolegenuly 

Al-Farabi Kazakh National University, Almaty, Kazakhstan

*e-mail: damirsuleimanov1008w@gmail.com

SURROGATE MODELING OF TEMPORAL EVOLUTION FOR THE BUCKLEY–LEVERETT RIEMANN PROBLEM USING FOURIER NEURAL OPERATOR

Abstract. The Buckley–Leverett equation is a fundamental model of two-phase flow in porous media, widely used in petroleum engineering to describe waterflooding. The analytical solution to this nonlinear hyperbolic problem contains discontinuous shock waves and requires Oleinik's entropy condition to select a physically admissible weak solution. This study investigates whether the Fourier Neural Operator (FNO) can learn the time-evolution map of a fixed Riemann problem solution for the Buckley–Leverett equation at mobility ratio $M = 1$. Unlike operator learning over a family of initial conditions, we consider surrogate modeling where the input is a fixed initial saturation profile together with a time instant, and the output is the saturation profile at that instant. The training dataset comprises 1000 temporal slices of a single analytical solution obtained by the method of characteristics with the Rankine–Hugoniot condition satisfied, for randomly selected time instants $T \in [0.01, 1.0]$. The trained FNO model demonstrated high approximation accuracy: relative L_2 error of 0.0191 ± 0.0253 , coefficient of determination $R^2 = 0.9970 \pm 0.0127$, and mean absolute error $MAE = 0.0023 \pm 0.0016$. The shock front position was predicted with a displacement of less than one spatial grid step (4×10^{-4}) across all time quartiles. The main limitation was the Gibbs oscillations near the discontinuity, manifested as local oscillations. The results show that the FNO architecture can effectively approximate the structure of the shock-wave solution to a fixed Riemann problem, including the rarefaction zone and the saturation front position.

Keywords: Fourier Neural Operator, Buckley–Leverett equation, surrogate model, two-phase flow, porous media, shock wave, Riemann problem.

1. Introduction

Modeling filtration processes in porous media is a crucial task in fluid dynamics and petroleum engineering. One of the foundational mathematical models describing the displacement of one fluid by another is the Buckley–Leverett equation, proposed in their classical work [1]. This equation is a nonlinear hyperbolic partial differential equation whose solutions can exhibit discontinuities, which significantly complicates its numerical solution and analysis.

Classical numerical methods are widely applied to solve hyperbolic partial differential equations [2]–[4]. However, in problems with discontinuous solutions, such as the Buckley–Leverett equation, standard schemes can lead to incorrect results, namely the emergence of spurious oscillations near saturation fronts and convergence to non-physical solutions.

To correctly resolve discontinuities, a scheme must satisfy the conservation property and entropy conditions for selecting a physically admissible

weak solution. One of the fundamental approaches is based on the exact or approximate solution of the Riemann problem at each time step [5]. A further development in this direction led to high-order schemes with TVD (Total Variation Diminishing) limiters [4], [6], which ensure solution monotonicity and suppress non-physical oscillations.

Despite the maturity of traditional numerical methods, their application to the Buckley–Leverett equation requires careful tuning of the scheme and becomes computationally expensive when solving the problem multiple times with varying initial conditions—a scenario typical for optimization and parameter uncertainty analysis in petroleum engineering.

The desire to overcome these computational limitations has stimulated interest in machine learning methods for solving differential equations. One of the most prominent approaches is Physics-Informed Neural Networks (PINNs), proposed by Raissi et al. [7], which embed physical laws into the neural network training process by incorporating differential equations into the loss

function. Despite their versatility, such methods can encounter training difficulties when applied to problems with pronounced nonlinearities and discontinuous solutions, which are characteristic of the Buckley–Leverett equation.

An alternative direction is the training of neural operators, which approximate mappings between functional spaces [8]. The Fourier Neural Operator (FNO) method was proposed by Li et al. [9] as a neural operator architecture that parameterizes the integral kernel directly in the Fourier space. In the original work, FNO demonstrated high accuracy in solving a series of parametric problems in mathematical physics, including the Burgers equation, the Darcy flow problem, and the Navier–Stokes equations. An important property of the architecture is its discretization invariance: the model is trained on one grid resolution and can be applied to an arbitrary one, which fundamentally distinguishes it from standard convolutional networks [10].

Since its publication, FNO has found application in several computational physics problems and has received several architectural extensions. Geo-FNO extended the standard architecture to problems with irregular geometries via a diffeomorphic mapping of the physical domain into a computational one, which is highly relevant for realistic geological reservoir geometries [11]. In the field of seismology, neural operators based on the FNO architecture have been used to model elastic wave propagation: the trained model reproduces the full wave field nearly two orders of magnitude faster than traditional numerical methods while maintaining accuracy when generalizing to velocity models not included in the training set [12], [13].

The application of FNO to hyperbolic equations with discontinuous solutions deserves special attention. A method for approximating numerical fluxes using FNO for hyperbolic conservation laws was proposed, replacing traditional numerical fluxes in conservative schemes with a neural operator, which improved the model's generalization ability when solving problems with discontinuities [14]. Nevertheless, the question regarding the applicability of the standard FNO architecture to problems with shock waves and saturation fronts characteristic of the Buckley–Leverett equation remains open.

Thus, the literature analysis shows that despite the active development of neural operator methods

and their success in computational physics, their capability to approximate discontinuous solutions of hyperbolic equations remains less studied. In particular, the Buckley–Leverett equation, which describes two-phase flow in porous media, yields solutions with shock waves and sharp saturation fronts, presenting a fundamental challenge for architectures based on low-frequency parameterization in Fourier space. Existing works on applying neural operators to filtration problems [15] utilize a physics-informed approach (PI-DeepONet) that requires the explicit inclusion of the equation in the loss function, whereas the capabilities of a purely data-driven FNO-based approach for this class of problems have not been systematically investigated.

Consequently, the objective of the present work is to evaluate the ability of the Fourier Neural Operator architecture to approximate the temporal evolution and shock front structure of the analytical solution to a fixed Riemann problem for the Buckley–Leverett equation. Unlike the operator learning framework over a family of initial conditions, this work considers surrogate modeling, where the input is a fixed initial saturation profile along with a time instant, and the output is the saturation profile at that instant.

While exact analytical solutions to the 1D Buckley–Leverett Riemann problem can be derived using Welge's construction and the Rankine–Hugoniot condition, obtaining the saturation profile at an arbitrary time t requires iterative numerical root-finding to locate the shock front and resolve the shock-rarefaction wave structure. The proposed time-parameterized Fourier Neural Operator surrogate eliminates these iterative root-finding steps, executing an instantaneous forward pass $t \rightarrow S_w(\cdot, t)$ without relying on step-by-step time integration or non-linear flux solvers. Furthermore, although the initial saturation profile is kept fixed in the present dataset, the three-channel input structure (S_0, x, t) is deliberately designed to allow seamless extension to multi-initial-condition operator learning in future work.

2. Materials and methods

2.1. Problem Statement

This study addresses a one-dimensional two-phase flow problem for immiscible fluids in a porous medium, governed by the Buckley–Leverett equation. The model is obtained from the mass

conservation law for the wetting phase, subject to the assumptions of incompressibility of both fluids and the porous matrix, negligible capillary and gravitational forces, and a constant total flow rate.

This paper uses the dimensionless form of the equation [1]:

$$\frac{\partial S_w}{\partial t} + \frac{\partial f(S_w)}{\partial x} = 0, \quad (1)$$

where $S_w(x, t) \in [0, 1]$ – is the water saturation, $f(S_w)$ – is the fractional flow function.

The fractional flow function is determined in terms of the relative permeabilities and viscosities of the phases and is defined in this paper as:

$$f(S_w) = \frac{\lambda_w(S_w)}{\lambda_w(S_w) + \lambda_o(S_w)},$$

where $\lambda_w = \frac{k_{rw}(S_w)}{\mu_w}$, $\lambda_o = \frac{k_{ro}(S_w)}{\mu_o}$. (2)

A standard quadratic approximation of relative permeabilities is used:

$$k_{rw} = S_w^2, k_{ro} = (1 - S_w)^2$$

which leads to the expression:

$$f(S_w) = \frac{S_w^2}{S_w^2 + \frac{(1-S_w)^2}{M}}, \quad (3)$$

$$M = \frac{\mu_o}{\mu_w},$$

where M - is the mobility ratio.

Fractional flow derivative:

$$f'(S_w) = \frac{2S_w(1-S_w)}{M\left(S_w^2 + \frac{(1-S_w)^2}{M}\right)^2}. \quad (4)$$

The saturation at the S^* front is determined by the tangent condition (Welge's construction [16]):

$$\frac{f(S^*) - f(S_r)}{S^* - S_r} = f'(S^*), \quad (5)$$

where $S_r = 0$ – initial saturation ahead of the front.

Shock velocity:

$$v_{shock} = f'(S^*). \quad (6)$$

The Riemann problem is specified by the following conditions. The initial condition is a step function corresponding to the domain being completely filled with oil at $t = 0$:

$$S_w(x, 0) = \begin{cases} 1, & x = 0 \\ 0, & x > 0 \end{cases} \quad (7)$$

where $S_w = 0$ is the residual water saturation in the reservoir (oil is not displaced), and $S_w = 1$ is the full water saturation at the left boundary. The boundary condition at the left boundary fixes a constant water injection:

$$S_w(0, t) = 1, t > 0. \quad (8)$$

which corresponds to the instantaneous start of pure water injection at $t = 0$. In the numerical implementation, the initial discontinuity is approximated at the grid node $x_0 = 0$.

Consequently, the code utilizes the condition $S_w = 1$ for $x < 10^{-3}$ and $S_w = 0$ for $x \geq 10^{-3}$. This approach serves as a technical discretization technique and has no impact on the analytical solution for $t > 0$.

$$S_w(x, 0) = \begin{cases} 1, & x < 0.001 \\ 0, & x \geq 0.001 \end{cases} \quad (9)$$

The shock front reaches the right boundary $x = 1$ at the breakthrough time $t_{bt} = \frac{1}{v_{shock}}$. For the case $M = 1$, this gives $t_{bt} \approx 0.828$. For $t > t_{bt}$, the domain no longer contains a shock in its interior. The solution is entirely described by the rarefaction wave:

$$S_w(x, t) = \text{clip}(f'^{-1}(x/t), S^*, 1) \quad (10)$$

where f'^{-1} denotes the inverse of the flux derivative, and the clipping enforces the physical lower bound S^* on saturation throughout the

domain. The training dataset includes time instances up to $T = 1.0$, which extends beyond breakthrough, so the model is trained on both pre- and post-breakthrough solution profiles.

The training dataset is formed from temporal slices of a single analytical solution to the Riemann problem for the Buckley-Leverett equation, obtained by the method of characteristics [5]. The shock front position was determined from the Rankine-Hugoniot condition (5) and the selection of a physically admissible weak solution was ensured by the Oleinik entropy condition [17], [18].

It should be noted that in this work the initial condition is fixed and does not vary in the training dataset. Thus, the model is trained not on a general solution operator over the space of initial conditions, but on the mapping $T \mapsto S_w(\cdot, T)$ for a fixed Riemann problem. Extension to arbitrary initial conditions is a direction for future research.

Additionally, the random split of the dataset into training, validation, and testing sets primarily evaluates the model's capability to interpolate between previously observed time instants. Investigating the capacity of the FNO architecture for temporal extrapolation under a block-wise partitioning of the time interval is beyond the scope of this study and is identified as a direction for future research.

The spatial domain was discretized with a uniform grid of 1024 nodes over the spatial interval $x \in [0, 1]$, $x_i = i\Delta x$, $i = 0, \dots, 1023$ with mesh size $h = \frac{1}{1023}$. For training, 1000 spatial saturation profiles $S_w(x, T_k)$, were generated, corresponding to randomly selected time instants $T_k \sim \mathcal{U}(0.01, 1.0)$ for a fixed step-type initial condition. For brevity, we denote $u_i^{(k)} \equiv S_w(x_i, t_k)$, where $u_{true,i}^{(k)}$ is the analytical solution and $u_{pred,i}^{(k)}$ is the model prediction.

2.2. Metrics

The following metrics were used to evaluate model quality, computed for each temporal slice $k = 1, \dots, N_{test}$ and averaged over the test dataset:

$$Rel L_2^{(k)} = \frac{\|u_{pred}^{(k)} - u_{true}^{(k)}\|_2}{\|u_{true}^{(k)}\|_2}, \quad (11)$$

$$MAE^{(k)} = \frac{1}{N_x} \sum_{i=1}^{N_x} |u_{true,i}^{(k)} - u_{pred,i}^{(k)}|, \quad (12)$$

$$R^2(k) = 1 - \frac{\sum_{i=1}^{N_x} (u_{true,i}^{(k)} - u_{pred,i}^{(k)})^2}{\sum_{i=1}^{N_x} (u_{true,i}^{(k)} - \frac{1}{N} \sum_{i=1}^{N_x} u_{true,i}^{(k)})^2}, \quad (13)$$

$$L_\infty^{(k)} = \max_{i=1, \dots, N_x} |u_{true,i}^{(k)} - u_{pred,i}^{(k)}|. \quad (14)$$

To account for potential skewness and heavy-tailed behavior in the error distribution across temporal slices, non-parametric statistics are reported alongside parametric metrics. The median error is computed as $\text{Median}(E) = Q_2(E)$, where $Q_2(E)$ denotes the 50th percentile of the error metric E across all test time slices. Dispersion around the median is quantified by the Interquartile Range, defined as $\text{IQR} = [Q_1(E), Q_3(E)]$, where $Q_1(E)$ and $Q_3(E)$ represent the 25th and 75th percentiles, respectively.

To quantitatively evaluate the accuracy of reproducing the shock front, two additional metrics were utilized. The front position x_{front} was defined as the first intersection of the saturation profile with the level $\frac{S^*}{2}$ using the linear interpolation method:

$$x_{\text{front}} = x_j + \frac{S_w(x_j) - S^*/2}{S_w(x_j) - S_w(x_{j+1})} \cdot \Delta x, \quad (15)$$

where j is the node index such that $S_w(x_j) \geq \frac{S^*}{2} > S_w(x_{j+1})$, and $\Delta x = \frac{1}{1023}$.

The front displacement was calculated as:

$$\Delta x_{\text{front}} = |\hat{x}_{\text{front}} - x_{\text{front}}^{\text{true}}|, \quad (16)$$

where $x_{\text{front}}^{\text{true}} = v_{\text{shock}} \cdot t$ is the analytical front position.

The blurring width ω was defined within a window of ± 40 nodes around x_{front} as the distance between the levels $0.9S^*$ and $0.1S^*$:

$$\omega = x(S_w \leq 0.1 S^*) - x(S_w \geq 0.9 S^*), \quad (17)$$

where $x(S_w \geq 0.9 S^*)$ is the last node within the window with a saturation of at least $0.9S^*$, and $x(S_w \leq 0.1 S^*)$ is the first node to the right of the front with a saturation not exceeding $0.1S^*$.

The use of a local window eliminates the influence of Gibbs oscillations occurring far from the front [19]. For the analytical solution (an ideal shock), $\omega^{\text{true}} \leq \Delta x \approx 9.8 \times 10^{-4}$.

2.3. Fourier Neural Operator Architecture

This study employs a one-dimensional Fourier Neural Operator (FNO-1d) architecture [9], training a surrogate model that approximates the temporal evolution of the solution for a fixed initial condition.:

$$\mathcal{G}_\theta: (S_w(x, 0), x, t) \rightarrow S_w(x, t), \quad (18)$$

where θ – is the trainable model parameters.

The architecture consists of three sequential stages: lifting (dimensionality lifting), iterative application of spectral layers, and projection into the output space.

The model input is formed as a tensor of dimension $N_x \times C_{in}$, where $C_{in} = 3$. The first channel contains the fixed initial saturation profile $S_w(x, 0)$, the second – the spatial coordinate x , and the third – the time instant T , replicated across the entire spatial grid. The model output is a one-dimensional saturation profile, i.e., $C_{out} = 1$. In this study, the input channel containing the initial condition has a fixed value across all considered samples. Nevertheless, this channel is intentionally retained within the model architecture to ensure the extensibility of the method. This will allow future investigations to transition to solving more general problems without altering the structure of the neural network.

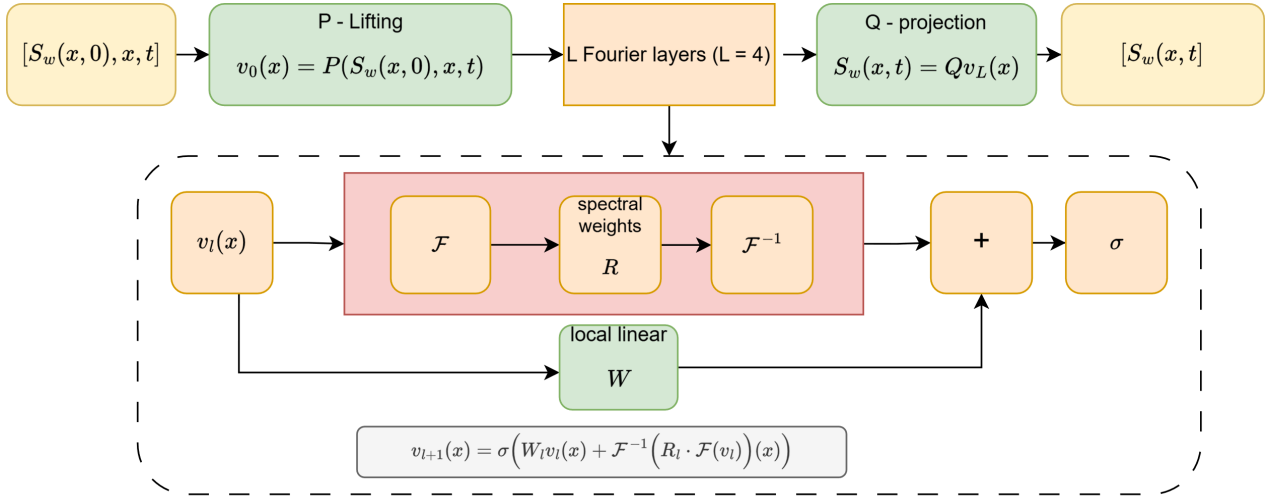


Figure 1. Fourier Neural Operator Architecture

The first stage is a lifting operation. The input tensor is projected into a hidden space of dimension via a learnable linear transformation P :

$$v_0(x) = P(S_w(x, 0), x, t). \quad (19)$$

This is necessary to transform the physical input channels into a sufficiently rich internal representation before applying the spectral layers.

The second stage comprises spectral layers. Four Fourier layers ($L = 4$) are applied sequentially. Each layer performs two parallel transformations and sums their results:

$$v_{l+1}(x) = \sigma(W_l v_l(x) + \mathcal{F}^{-1}(R_l \cdot \mathcal{F}(v_l))(x)), \quad (20)$$

where W_l – is the local linear transformation that captures local dependencies: R_l are the learnable complex weights in Fourier space, applied only to the first $k_{max} = 32$ modes; $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the forward and inverse Fourier transforms; σ – is the GELU nonlinear activation function.

The high-frequency modes are zeroed out, which is equivalent to low-pass filtering and allows the model to focus on the global structures of the solution. It is this spectral parameterization of the kernel that ensures discretization invariance.

After the last spectral layer, the hidden representation is projected back into the physical space via a two-layer network:

$$S_w(x, t) = Q v_L(x), \quad (21)$$

where $Q: \mathbb{R}^d \rightarrow \mathbb{R}^{C_{out}}$ – is the trainable linear transformation.

The model was trained using the Adam optimizer [20] with an initial learning rate of $\eta = 10^{-3}$ and a regularization coefficient of $\lambda = 10^{-4}$.

The Cosine Annealing scheduler [21] was employed for adaptive learning rate adjustment over 300 epochs. The mean squared error (MSE) was used as the loss function:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \| \hat{S}_w^{(i)} - S_w^{(i)} \|_2^2 \quad (22)$$

Table 1. Hyperparameter table.

Hyperparameter	Value
Spatial domain	$x \in [0,1]$
Temporal domain	$t \in [0.01,1.0]$
Number of grid nodes	$N_x = 1024(\text{mesh size } h = \frac{1}{1023})$
Mobility ratio	$M = 1$
Initial saturation	0
Injected saturation	1.0
Number of spectral layers L	4
Number of Fourier modes k_{max}	32
Hidden dimension d	64
Input channels C_{in}	3
Output channels C_{out}	1
Optimizer	Adam
Learning rate η	10^{-3}
Regularization λ	10^{-4}
Scheduler	Cosine Annealing
Scheduler parameters	$T_{max} = 300, \eta_{min} = 10^{-6}$
Model parameters	329,537
Epochs	300
Batch size	32
Train/validation/test split	60/20/20
Random seed	42
Loss function	MSE
Hardware	NVIDIA GeForce RTX 3060 Laptop GPU (6 GB)
Training time	~ 2 minutes

3. Results

The loss function curves for the training and validation sets during the training of the FNO model are shown in Figure 2. Both curves show a steady convergence over 300 epochs, decreasing

from an initial value of approximately 1.0 to a final value of approximately 0.02.

Figure 3 shows a comparison of the saturation profiles predicted by the FNO surrogate model with the analytical solution of the fixed Riemann problem at six time points ($t \in [0.1,0.6]$).

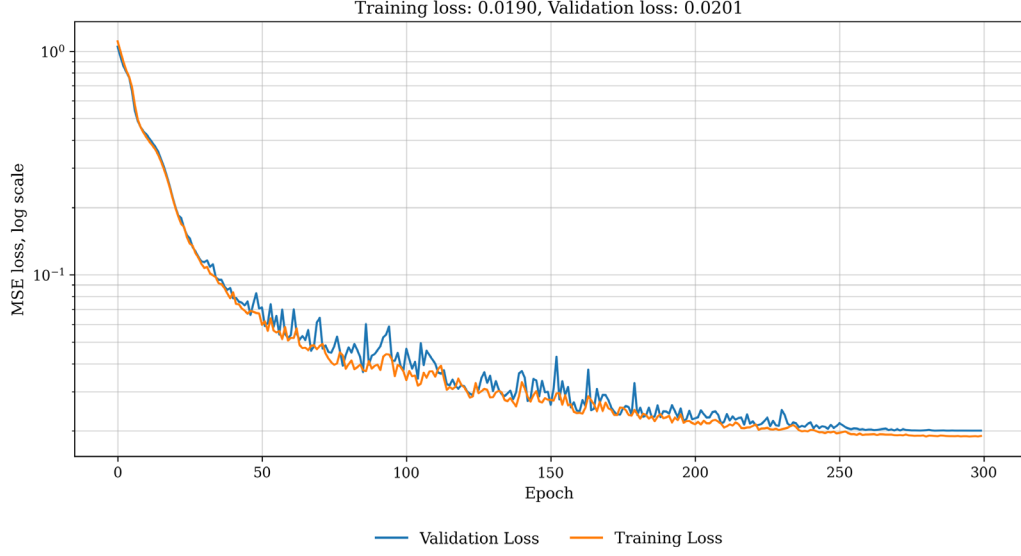


Figure 2. Loss function curves

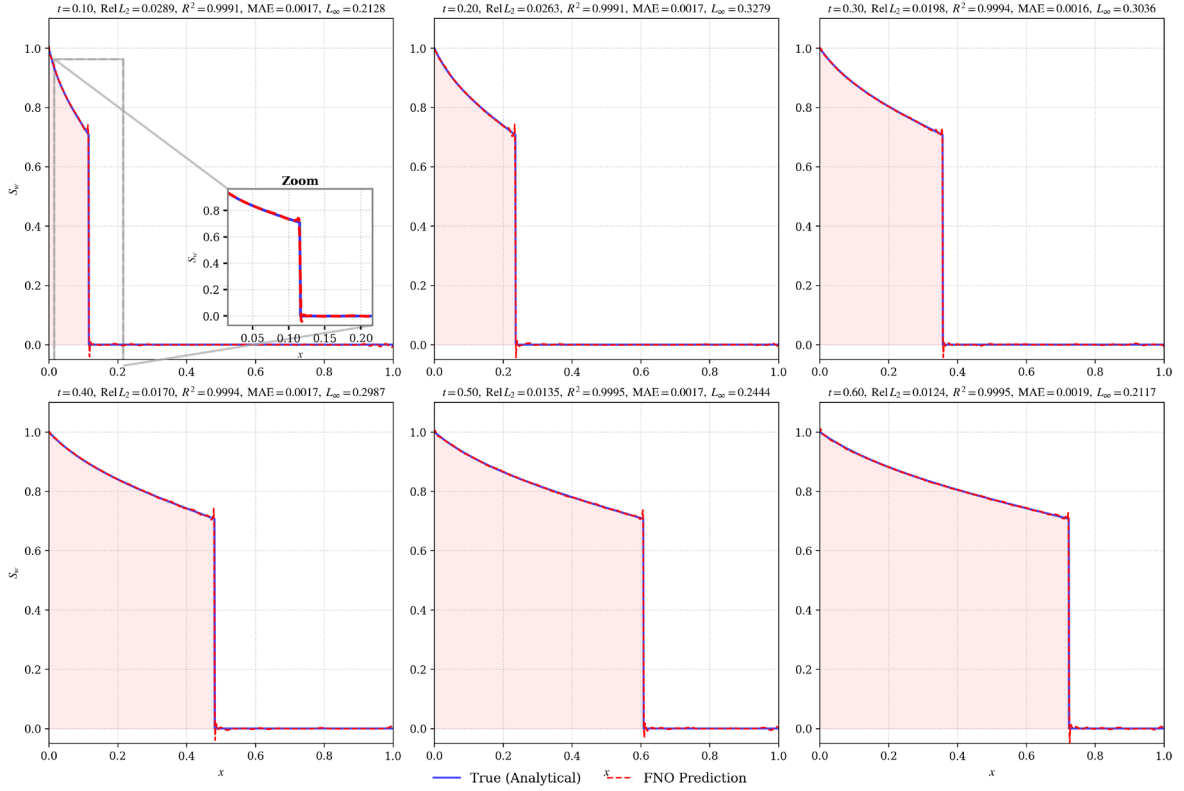


Figure 3. Comparison of saturation profiles: FNO model prediction versus the analytical solution at $M = 1$ for time instants $t = 0.1, t = 0.2, t = 0.3, t = 0.4, t = 0.5$, and $t = 0.6$. These instances illustrate model performance during the pre-breakthrough flow regime ($t < t_{\text{breakthrough}} \approx 0.828$), where errors are lower than the overall test-set mean due to the absence of boundary exit dynamics

To evaluate the internal physical correctness of the FNO surrogate model, a numerical analysis of mass conservation was conducted. On the test set, the mean absolute mass conservation error is 2.00×10^{-4} , and the mean relative balance error does not exceed 0.033%.

The spatial distribution of the pointwise absolute error $|\Delta S_w| = |S_w^{\text{pred}} - S_w^{\text{true}}|$ is shown in Figure 4.

The spatio-temporal map of the absolute error across the entire space-time domain is presented in Figure 5.

The overall quantitative evaluation on the test set is summarized in Table 2.

A quantitative analysis of the shock front position and its blurring width across time quartiles is presented in Table 3 and Figure 6

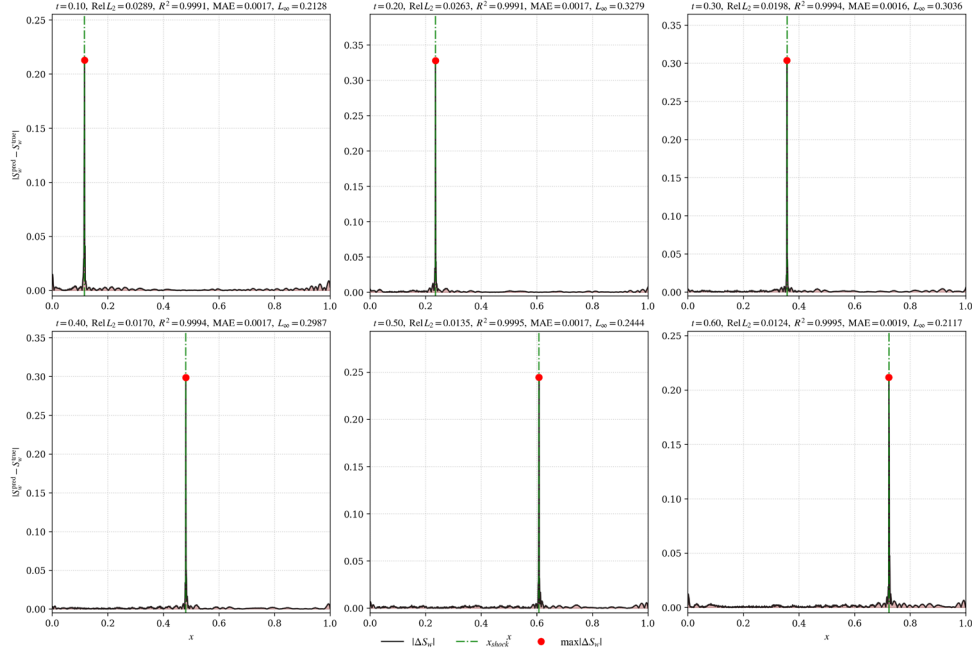


Figure 4. Spatial distribution of the pointwise absolute error $|\Delta S_w| = |S_w^{\text{pred}} - S_w^{\text{true}}|$ for the same pre-breakthrough time instants ($t \in [0.10, 0.60]$). The green dash-dotted vertical line indicates the exact analytical position of the shock front (x_{shock}), and the red dot marks the maximum error $\max |\Delta S_w|$

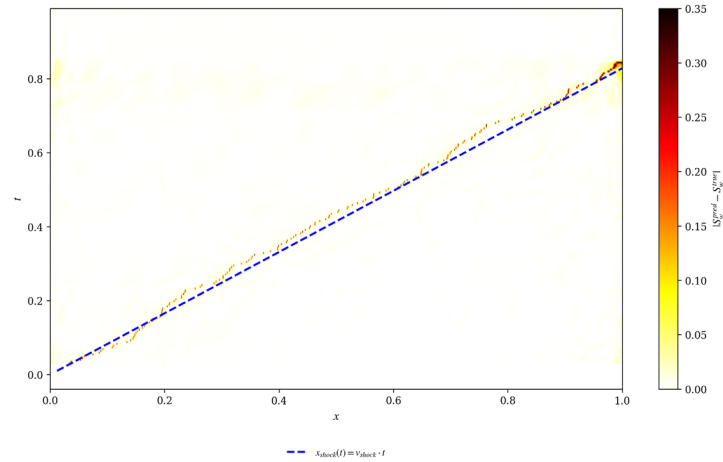


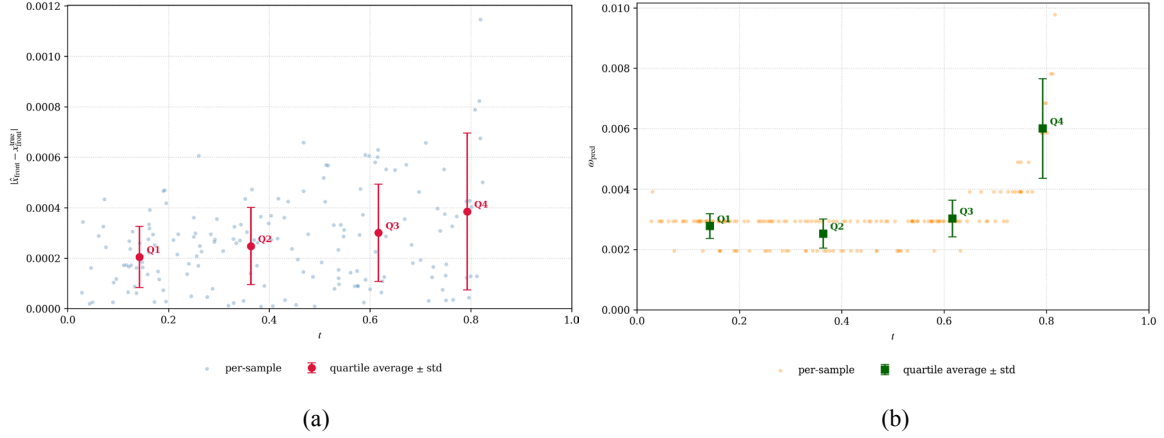
Figure 5. Spatio-temporal map of the model's absolute prediction error FNO $|S_w^{\text{pred}} - S_w^{\text{true}}|$ on the test sample

Table 2. FNO model quality metrics on the test set

Metric	Designation	Mean $\pm$ std	Median [IQR: Q1–Q3]
FNO Surrogate (Rel. L_2)	$Rel. L_2$	0.0191 ± 0.0253	0.0183[0.0140 – 0.0223]
Cubic Spline Baseline (Rel. L_2)	$Rel. L_2$	0.0533 ± 0.0533	0.0286 [0.0034 – 0.0471]
Coefficient of Determination	R^2	0.9970 ± 0.0127	0.9988[0.9984 – 0.9990]
Mean Absolute Error	MAE	0.0023 ± 0.0016	0.0020[0.0019 – 0.0022]
Maximum Absolute Error	L_∞	0.1585 ± 0.1185	0.1843[0.1478 – 0.2333]

Table 3. Quantitative analysis of the position of the shock front and the width of its blurring by time quartiles

Quartile	t (mean)	Front displacement Δx_{front} (mean $\pm$ std)	ω_{pred} (mean $\pm$ std)
Q1 (0.00-0.25)	0.143	0.00020 ± 0.00012	0.00277 ± 0.00041
Q2 (0.25-0.5)	0.364	0.00025 ± 0.00015	0.00252 ± 0.00048
Q3 (0.5-0.75)	0.616	0.00030 ± 0.00019	0.00302 ± 0.00060
Q4 (0.75-1.00)	0.793	0.00038 ± 0.00031	0.00600 ± 0.00165


Figure 6. Analysis of the shock front position and its blurring width:
 (a) displacement of the shock front; (b) width of the shock front

The mean shock front displacement remains below half of a spatial grid cell across all time quartiles, slightly increasing from 2.0×10^{-4} in quartile Q1 ($t_{\text{mean}} = 0.143$) to 3.9×10^{-4} in quartile Q4 ($t_{\text{mean}} = 0.793$). The shock front smearing width ω_{pred} remains stable at approximately $(2.5 - 3.0) \times 10^{-3}$ across quartiles Q1–Q3, but increases to 6.0×10^{-3} in quartile Q4.

An inference timing benchmark comparing the proposed FNO surrogate with a high-resolution

numerical solver (Godunov TVD scheme) across grid resolutions $N_x \in [512, 4096]$ is summarized in Table 4.

While the numerical TVD solver runtime increases from 5,964.33 ms at $N_x = 512$ to 44,332.01 ms at $N_x = 4096$, the inference time of the FNO model remains nearly constant at 3.07–3.57 ms across all tested resolutions, achieving a speedup of $1,671 \times$ at $N_x = 512$ and up to $14,436 \times$ at $N_x = 4096$.

Table 4. Computational inference time comparison (in milliseconds, mean $\pm$ std) to compute the saturation profile at final time $t = 1.0$ across different spatial grid resolutions

Resolution (Nx)	Godunov TVD Scheme (ms)	FNO Model (ms)	FNO Speedup vs. Godunov
512	5964.33 $\pm$ 56.97	3.57 $\pm$ 0.48	1671 $\times$
1024	12531.00 $\pm$ 38.15	3.15 $\pm$ 0.09	3981 $\times$
2048	22844.66 $\pm$ 95.69	3.26 $\pm$ 0.55	7010 $\times$
4096	44332.01 $\pm$ 90.49	3.07 $\pm$ 0.44	14436 $\times$

4. Discussion

4.1. Error Localization and Physical Compliance Analysis

Across all considered time instants, the predicted saturation profiles qualitatively reproduce the analytical solution structure: the rarefaction wave is approximated with high accuracy, and the shock front position aligns with the theoretical location within $\Delta x_{\text{front}} < 4 \times 10^{-4}$ (less than half a spatial grid cell). As demonstrated in Figure 3, the relative L_2 error decreases monotonically from 0.0289 at $t = 0.10$ to 0.0124 at $t = 0.60$. This improvement occurs because the early-time saturation profile exhibits steeper gradients in the rarefaction zone, presenting a greater approximation challenge for the low-frequency spectral kernel. As the front propagates deeper into the spatial domain, the solution gradient relaxes, and the coefficient of determination R^2 remains consistently high within 0.9991--0.9995.

The mean absolute error (MAE) remains in a narrow band (0.0016--0.0019), indicating uniform accuracy outside the discontinuity. In contrast, the maximum pointwise error L_∞ reaches 0.20--0.33 near the front. As localized in Figure 4, peak errors ($\max |\Delta S_w|$) strictly coincide with the theoretical shock front position $x_{\text{shock}}(t) = v_{\text{shock}} \cdot t$. Outside this localized frontal zone, pointwise errors remain below 0.015, confirming that the rarefaction region and undisturbed domain are predicted with high fidelity. The spatio-temporal error map (Figure 5) further confirms that error accumulation is strictly confined along the moving shock front trajectory throughout the entire evolution $t \in [0.01, 1.0]$.

Physical admissibility analysis revealed that 20.8% of test nodes yielded minor negative saturations ($S_w < 0$), while 0.25% exceeded unity. These violations were small in magnitude, averaging 0.002 and 0.001, respectively (bounded within $S_w \in [-0.015, 1.012]$). Spatial tracking confirms that negative values occur exclusively ahead of the front ($x > x_{\text{shock}}$) in the unflooded

zone ($S_w = 0$), where truncated Fourier modes induce minor Gibbs undershoots. Mass conservation integration via the trapezoidal rule demonstrated a mean relative balance error under 0.033%, proving that local Gibbs excursions cancel out globally and preserve phase conservation.

The quartile-wise analysis explains the temporal evolution of prediction errors. In quartiles $Q1 - Q3$ ($t < 0.75$), the front smearing width remains virtually constant, proving that spectral smearing is systematic rather than cumulative over time.

The anomalous increase in both front displacement and smearing width observed in quartile $Q4$ ($t \in 0.75, 1.00$) is physically attributed to the shock front approaching the right boundary ($x = 1.0$) and the subsequent breakthrough event at $t_{\text{bt}} \approx 0.828$. After breakthrough, boundary interactions and the exit of the shock front from the spatial domain introduce transient spectral artifacts, driving the heavy-tailed error distribution observed in post-breakthrough slices.

4.2. Computational Efficiency and Real-Time Application Potential

To evaluate the practical efficiency of the proposed FNO surrogate, an inference timing benchmark was conducted (as summarized in Section 3, Table 4). All GPU executions were evaluated on an NVIDIA GeForce RTX 3060 Laptop GPU (6 GB), while CPU baselines were run on an 12th Gen Intel(R) Core(TM) i7-12650H CPU (2.30 GHz).

The difference in execution scaling highlights the fundamental computational advantage of neural operator surrogates over traditional shock-capturing schemes. Classical finite-volume methods (such as the Godunov TVD scheme) suffer from severe $\mathcal{O}(N_x^2)$ computational complexity due to the strict Courant–Friedrichs–Lewy (CFL) stability condition, which enforces increasingly smaller time steps as the spatial grid is refined.

In contrast, although the theoretical complexity of 1D FNO is $\mathcal{O}(N_x \log N_x)$ due to Fast Fourier Transforms, the measured inference time stays virtually flat ($\approx 3.1 - 3.5$ ms) on GPU hardware up to $N_x = 4096$. At these grid sizes, CUDA kernel dispatch overhead dominates the execution time rather than grid resolution.

By executing an instantaneous forward pass $t \rightarrow S_w(\cdot, t)$ without step-by-step temporal integration or nonlinear flux iterations, the FNO surrogate achieves a $3981 \times$ speedup at the primary grid resolution ($N_x = 1024$) and $14436 \times$ zero-shot speedup at $N_x = 4096$. Since generating

the training dataset requires offline numerical integration, this computational advantage is amortized in multi-query scenarios, such as real-time reservoir optimization, history matching, and uncertainty quantification.

4.3. Hyperparameter Sensitivity Analysis

To evaluate model sensitivity to network architecture, an ablation study was conducted across three key structural parameters: the number of Fourier modes $k_{\max}$, feature dimension d , and spectral layers L . The results are summarized in Tables 5–7.

Table 5. Ablation study for $k_{\max}$

Configuration	$k_{\max} = 8$	$k_{\max} = 16$	$k_{\max} = 32$ (baseline)
Rel L_2 (mean $\pm$ std)	0.0257 ± 0.0273	0.0230 ± 0.0254	0.0191 ± 0.0253
R^2 (mean $\pm$ std)	0.9956 ± 0.0153	0.9961 ± 0.0148	0.9970 ± 0.0127
MAE (mean $\pm$ std)	0.0029 ± 0.0017	0.0022 ± 0.0012	0.0023 ± 0.0016
L_∞ (mean $\pm$ std)	0.2067 ± 0.1134	0.1967 ± 0.1153	0.1585 ± 0.1185
Parameters	132,929	198,465	329,537

Table 6. Ablation study for the feature dimension d

Configuration	$d = 32$	$d = 64$ (baseline)	$d = 128$
Rel L_2 (mean $\pm$ std)	0.0258 ± 0.0284	0.0191 ± 0.0253	0.0161 ± 0.0246
R^2 (mean $\pm$ std)	0.9956 ± 0.0153	0.9970 ± 0.0127	0.9974 ± 0.0114
MAE (mean $\pm$ std)	0.0033 ± 0.0018	0.0023 ± 0.0016	0.0015 ± 0.0015
L_∞ (mean $\pm$ std)	0.2064 ± 0.1109	0.1585 ± 0.1185	0.1355 ± 0.1187
Parameters	82,849	329,537	1,314,433

Table 7. Ablation study for network depth L

Configuration	$L = 3$	$L = 4$ (baseline)	$L = 5$
Rel L_2 (mean $\pm$ std)	0.0211 ± 0.0261	0.0191 ± 0.0253	0.0187 ± 0.0258
R^2 (mean $\pm$ std)	0.9972 ± 0.0100	0.9970 ± 0.0127	0.9966 ± 0.0156
MAE (mean $\pm$ std)	0.0018 ± 0.0010	0.0023 ± 0.0016	0.0020 ± 0.0015
L_∞ (mean $\pm$ std)	0.1849 ± 0.1157	0.1585 ± 0.1185	0.1538 ± 0.1202
Parameters	251,489	329,537	407,585

Increasing $k_{\max}$ from 8 to 32 monotonically improved approximation quality, reducing Rel. L_2 from 0.0257 to 0.0191. Feature dimension d proved to be the most sensitive hyperparameter: reducing d to 32 degraded Rel. L_2 by 35%

(0.0191 $\rightarrow$ 0.0258). Varying depth L between 3 and 5 yielded minor performance changes, confirming that the baseline configuration ($k_{\max} = 32, d = 64, L = 4$) balances accuracy and parameter efficiency.

4.4. Evaluation of Discretization Invariance and Zero-Shot Grid Transfer

To verify discretization invariance, the model trained on $N_x = 1024$ was evaluated zero-shot across grids $N_x \in \{512, 2048, 4096\}$ at time instants $t \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$. The results are summarized in Table 8 and Figure 7.

Integral accuracy metrics (Rel. L_2 , R^2 , MAE) show strong stability across all grid resolutions, confirming zero-shot discretization invariance without retraining. The moderate increase in local peak error ($L_\infty \approx 0.41$ on $N_x = 4096$) is attributed to finer grid resolution measuring localized high-frequency Gibbs oscillations at the discontinuous front more sharply.

Table 8. Discretization-invariance performance across spatial resolutions

N_x	t	Rel L_2	R^2	MAE	L_∞
512	0.10	0.0138	0.9997	0.0012	0.0810
512	0.20	0.0102	0.9998	0.0014	0.0788
512	0.30	0.0279	0.9987	0.0018	0.3118
512	0.40	0.0170	0.9994	0.0018	0.2140
512	0.50	0.0046	0.9999	0.0014	0.0400
512	0.60	0.0139	0.9993	0.0020	0.2122
2048	0.10	0.0312	0.9988	0.0015	0.3213
2048	0.20	0.0215	0.9993	0.0016	0.3012
2048	0.30	0.0184	0.9994	0.0015	0.3404
2048	0.40	0.0152	0.9995	0.0017	0.2891
2048	0.50	0.0144	0.9994	0.0017	0.3190
2048	0.60	0.0141	0.9992	0.0020	0.3331
4096	0.10	0.0345	0.9986	0.0016	0.3887
4096	0.20	0.0243	0.9992	0.0018	0.3781
4096	0.30	0.0204	0.9993	0.0017	0.4145
4096	0.40	0.0165	0.9994	0.0019	0.3778
4096	0.50	0.0157	0.9993	0.0020	0.4072
4096	0.60	0.0152	0.9991	0.0021	0.4076

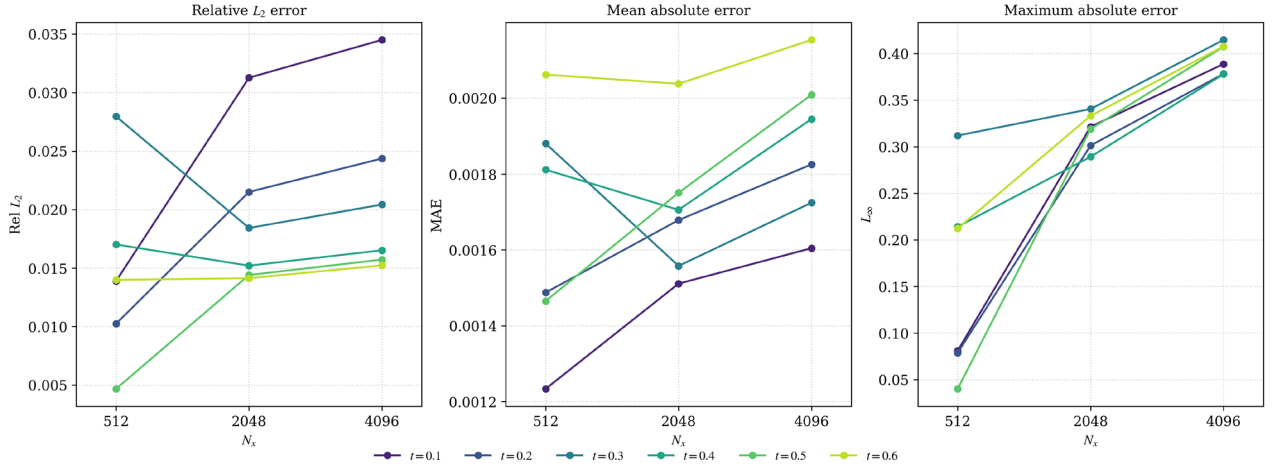


Figure 7. Discretization-invariance experiment plots

4.5. Mitigation Strategies for Gibbs Oscillations

Several physics-informed regularization losses were evaluated during training to suppress Gibbs oscillations, including Total Variation loss (L_{TV}),

Second-Order Smoothness loss (L_{smooth}), Physical Bound Penalty (L_{clip}), and Binarization Penalty (L_{bin}).

Empirical testing revealed that penalizing spatial derivatives or Fourier modes during

optimization produced negligible error reductions over the baseline MSE loss. In contrast, post-processing physical value clipping ($S_w \in [0,1]$) served as an optimal, zero-cost mitigation technique: it completely eliminated non-physical bound violations ($20.8\% \rightarrow 0.0\%$), removed maximum undershoot ($-0.0150 \rightarrow 0.0000$), and reduced the average relative L_2 error ($0.0191 \rightarrow 0.0187$).

5. Conclusions

In this work, the applicability of the Fourier Neural Operator (FNO) architecture for approximating a surrogate of the analytical solution to a fixed Buckley–Leverett problem was investigated. The considered problem is a nonlinear hyperbolic conservation law describing two-phase flow in porous media. The training dataset consisted of 1000 temporal snapshots of a single analytical solution to a Riemann problem with fixed parameters, generated using the method of characteristics together with the Rankine–Hugoniot condition and the Oleinik entropy condition for the case of equal phase viscosities ($M = 1$).

The trained model demonstrated high accuracy in approximating temporal snapshots of this fixed Riemann solution in the smooth regions of the saturation profile. The mean relative error was $\text{Rel. } L_2 = 0.0191 \pm 0.0253$, the coefficient of determination reached $R^2 = 0.9970$, and the mean absolute error was ($MAE = 0.0023$). The model successfully reproduced both the rarefaction region and the shock-front position. The displacement of the predicted shock-front location remained below 4×10^{-4} in all temporal quartiles, corresponding to less than one spatial grid spacing.

The primary limitation of the model is the presence of Gibbs-type oscillations in the vicinity of the shock front, manifested as localized prediction artifacts with a maximum error of $L_\infty = 0.1585 \pm 0.1185$. However, these errors remain strongly localized near the discontinuity, while outside the shock region the absolute error does not exceed 0.030. Analysis of the physical admissibility of the predictions showed that all violations of the saturation constraint $S_w \in [0,1]$ occur at locations where the true solution satisfies $S_w = 0$ and arise from Gibbs-type oscillations rather than from a systematic model bias.

A fundamental limitation of the present study is that the model approximates temporal snapshots of a single fixed solution corresponding to a Riemann problem with a specific initial condition and $M = 1$. Generalization to arbitrary initial conditions would require either the inclusion of problem parameters in the input space or retraining of the model, which lies beyond the scope of this work.

The discretization-invariance experiment showed that when the trained model was applied to grid resolutions different from the training resolution ($N_x = 1024$), including both coarser ($N_x = 512$) and finer ($N_x = 2048, 4096$) spatial grids, the model maintained high accuracy and consistency across time slices, with $\text{Rel } L_2$ error remaining low ($0.0046 - 0.0345$) and fully aligned with the training resolution baseline ($\text{Rel } L_2 = 0.0191$). While local maximum pointwise error L_∞ exhibits a minor increase on finer meshes ($L_\infty \approx 0.38 - 0.41$) due to the sharper spatial resolution of local Gibbs oscillations, these results confirm the zero-shot discretization-invariance of the Fourier Neural Operator across multi-resolution grids without retraining.

Acknowledgements

The authors would like to thank Al-Farabi Kazakh National University for providing computational resources and support for this research.

Funding

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant number AP23489431.

Conflicts of Interest

The authors declare no conflict of interest

Author Contributions

Conceptualization, D. Stolyarov, D. Suleimanov and Y. Tolegenuly; *Methodology*, D. Stolyarov and D. Suleimanov; *Formal Analysis*, D. Stolyarov, D. Suleimanov and Y. Tolegenuly; *Investigation*, D. Suleimanov; *Software*, D. Suleimanov; *Data Curation*, D. Suleimanov; *Visualization*, D. Suleimanov; *Writing – Original Draft Preparation*, D. Suleimanov; *Writing – Review & Editing*, D. Stolyarov, D. Suleimanov and Y. Tolegenuly;

References

1. S. E. Buckley and M. C. Leverett, 'Mechanism of Fluid Displacement in Sands', *Trans. AIME*, vol. 146, no. 01, pp. 107–116, Dec. 1942, doi: 10.2118/942107-G.
2. K.-J. Bathe, 'Finite Element Method', in *Wiley Encyclopedia of Computer Science and Engineering*, John Wiley & Sons, Ltd, 2008, pp. 1–12. doi: 10.1002/9780470050118.ecse159.
3. R. Eymard, T. Gallouët, and R. Herbin, 'Finite volume methods', in *Handbook of Numerical Analysis*, vol. 7, in Solution of Equation in $\mathbb{R}$ (Part 3), Techniques of Scientific Computing (Part 3), vol. 7, Elsevier, 2000, pp. 713–1018. doi: 10.1016/S1570-8659(00)07005-8.
4. R. J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*. in Cambridge Texts in Applied Mathematics. Cambridge: Cambridge University Press, 2002. doi: 10.1017/CBO9780511791253.
5. E. F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. doi: 10.1007/b79761.
6. A. Harten, 'High resolution schemes for hyperbolic conservation laws', *J. Comput. Phys.*, vol. 49, no. 3, pp. 357–393, Mar. 1983, doi: 10.1016/0021-9991(83)90136-5.
7. M. Raissi, P. Perdikaris, and G. E. Karniadakis, 'Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations', *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019, doi: 10.1016/j.jcp.2018.10.045.
8. L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, 'Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators', *Nat. Mach. Intell.*, vol. 3, no. 3, pp. 218–229, Mar. 2021, doi: 10.1038/s42256-021-00302-5.
9. Z. Li *et al.*, 'Fourier Neural Operator for Parametric Partial Differential Equations', Oct. 18, 2020, *arXiv: arXiv:2010.08895*. doi: 10.48550/arXiv.2010.08895.
10. V. Duruisseaux, J. Kossaiifi, and A. Anandkumar, 'Fourier Neural Operators Explained: A Practical Perspective', Jan. 22, 2026, *arXiv: arXiv:2512.01421*. doi: 10.48550/arXiv.2512.01421.
11. Z. Li, D. Z. Huang, B. Liu, and A. Anandkumar, 'Fourier Neural Operator with Learned Deformations for PDEs on General Geometries', May 02, 2024. doi: 10.5555/3648699.3649087.
12. Y. Yang, A. F. Gao, K. Azizzadenesheli, R. W. Clayton, and Z. E. Ross, 'Rapid Seismic Waveform Modeling and Inversion With Neural Operators', *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, 2023, doi: 10.1109/TGRS.2023.3264210.
13. G. S. O'Brien, C. J. Bean, H. Meiland, and P. Witte, 'Imaging and seismic modelling inside volcanoes using machine learning', *Sci. Rep.*, vol. 13, no. 1, p. 630, Jan. 2023, doi: 10.1038/s41598-023-27738-6.
14. T. Kim and M. Kang, 'Approximating Numerical Fluxes Using Fourier Neural Operators for Hyperbolic Conservation Laws', *Commun. Comput. Phys.*, vol. 37, no. 2, pp. 420–456, Sep. 2025, doi: 10.4208/cicp.OA-2024-0123.
15. W. Diab, O. Chaabi, S. Alkobaisi, A. Awotunde, and M. Al Kobaisi, 'Learning generic solutions for multiphase transport in porous media via the flux functions operator', *Adv. Water Resour.*, vol. 183, p. 104609, Jan. 2024, doi: 10.1016/j.advwatres.2023.104609.
16. H. J. Welge, 'A Simplified Method for Computing Oil Recovery by Gas or Water Drive', *J. Pet. Technol.*, vol. 4, no. 04, pp. 91–98, Apr. 1952, doi: 10.2118/124-G.
17. A. Heibig, 'Entropy estimates for conservation laws', *Appl. Math. Lett.*, vol. 6, no. 5, pp. 93–97, Sep. 1993, doi: 10.1016/0893-9659(93)90109-Z.
18. O. A. Oleinik, 'Discontinuous solutions of non-linear differential equations', in *American Mathematical Society Translations: Series 2*, vol. 26, Providence, Rhode Island: American Mathematical Society, 1963, pp. 95–172. doi: 10.1090/trans2/026/05.
19. D. Gottlieb and C.-W. Shu, 'On the Gibbs Phenomenon and Its Resolution', *SIAM Rev.*, vol. 39, no. 4, pp. 644–668, Jan. 1997, doi: 10.1137/S0036144596301390.
20. D. P. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization', 2014, *arXiv*. doi: 10.48550/ARXIV.1412.6980.
21. I. Loshchilov and F. Hutter, 'SGDR: Stochastic Gradient Descent with Warm Restarts', May 03, 2017, *arXiv: arXiv:1608.03983*. doi: 10.48550/arXiv.1608.03983.

Information about the authors:

Damir Suleimanov is a Junior Researcher at the Faculty of Information Technology and Artificial Intelligence, Al-Farabi Kazakh National University (Almaty, Kazakhstan, damirsuleimanov1008w@gmail.com). His research interests include machine learning, parallel algorithms, neural networks, and their applications in mathematical modeling. ORCID iD: 0009-0003-0122-4807.

Danila Stolyarov is a Junior Researcher at the Faculty of Information Technology and Artificial Intelligence, Al-Farabi Kazakh National University (Almaty, Kazakhstan, danilastolyarov97@gmail.com). His research interests include high-performance computing, parallel algorithms, and numerical simulation of physical processes. ORCID iD: 0009-0001-3137-9231.

Yermukassan Tolegenuly is a Junior Researcher at the Faculty of Information Technology and Artificial Intelligence, Al-Farabi Kazakh National University (Almaty, Kazakhstan, tolegenulyermukasan18@gmail.com). His research interests focus on computational fluid dynamics, data science, and mathematical modeling. ORCID iD: 0009-0003-7546-0861.

Submission received: 25 June, 2026

Revised: 24 August, 2026

Accepted: 24 August, 2026