

An encoder-decoder ConvLSTM surrogate model for simulating geological CO₂ sequestration with dynamic well controls



Zhao Feng ^{a,b,c}, Zeeshan Tariq ^c, Xianda Shen ^{a,b,*}, Bicheng Yan ^{c, **}, Xuhai Tang ^{d,e}, Fengshou Zhang ^{a,b}

^a Department of Geotechnical Engineering, College of Civil Engineering, Tongji University, Shanghai, 200092, China

^b Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education, Tongji University, Shanghai, 200092, China

^c Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955, Saudi Arabia

^d School of Civil Engineering, Wuhan University, Wuhan, 430072, China

^e Wuhan University Shenzhen Research Institute, Shenzhen, 518108, China

ARTICLE INFO

Keywords:

Geological carbon sequestration
Deep learning
Surrogate model
Encoder-decoder
Convolutional LSTM

ABSTRACT

In Geological Carbon Sequestration (GCS), effectively managing the project requires predicting state variables such as pressure and saturation. However, numerical simulation of multiphase flow in subsurface porous media involves solving large linear algebra systems, resulting in a substantial computational burden. This can make it impractical for real-time history matching or optimization. Meanwhile, deep learning-based surrogate models are emerging as fast and accurate approximators. This study delves into the application of the Encoder-Decoder Convolutional Long Short-Term Memory (ED-ConvLSTM) neural network for predicting the complex evolution of state variables under dynamic CO₂ injection schemes. Inception blocks enhanced with light-weighted attention modules are introduced in the encoder to extract high-dimensional input features. ConvLSTM is employed to propagate spatial temporal information in the low-dimensional latent space. Further, progressive upsampling blocks are used to reconstruct the latent features for the desired output. Instead of taking discrete time steps as an input feature, the proposed network captures the dynamic dependencies with the inherent ConvLSTM cell. The network has access to data at only portion of the initial time steps during training stage, while it is used to predict the state variables at unseen time steps during testing stage. Results show that the network can produce excellent predictions for both pressure and saturation, even at unseen future time steps. The remarkable generalizability to different geological permeability fields is also evaluated. ED-ConvLSTM outperforms the standard U-Net by far, especially when predicting beyond the training time period. These numerical experiments demonstrate the advantages of ED-ConvLSTM in terms of prediction accuracy, extrapolability and generalizability. This study highlights the importance of incorporating recurrent connections into the deep neural networks for simulating time-dependent multiphase flow problems. The proposed methodology has great potential in GCS surrogate modeling and offers a possible approach for real-time optimization of CO₂ injection.

1. Introduction

The rapid increase of anthropogenic carbon emissions has posed unprecedented challenges to global climate change. Over the past few decades, Geological Carbon Sequestration (GCS) has emerged as a promising strategy for mitigating the accumulation concentration of atmospheric carbon dioxide (CO₂) (Ajayi et al., 2019; Aminu et al., 2017). In the process of GCS, CO₂ is injected into geological formations such as deep saline aquifers (Bachu and Adams, 2003; Celia et al., 2015),

depleted hydrocarbon reservoirs (Liu et al., 2022; Lyu et al., 2021; Zhou et al., 2019), unmineable coal seams (Busch and Gensterblum, 2011) or basalts (Matter et al., 2016; Raza et al., 2022; Snæbjörnsdóttir et al., 2020). The characterization of CO₂ plume migration and injection induced over-pressure in target reservoirs is critical for site selection, operation management, and risk assessment (Birkholzer et al., 2015; Nordbotten and Celia, 2006; Silva et al., 2023).

Under the temperature and pressure conditions of in-situ reservoirs (above 7.38 MPa, 31.8 °C), injected CO₂ is in its supercritical state,

* Corresponding author. Department of Geotechnical Engineering, College of Civil Engineering, Tongji University, Shanghai, 200092, China.

** Corresponding author. Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955, Saudi Arabia.

E-mail addresses: xshen@tongji.edu.cn (X. Shen), bicheng.yan@kaust.edu.sa (B. Yan).

which forms a complex multiphase flow system with subsurface fluids. Recent years there have seen considerable progress in numerical models for simulating subsurface flow problems (Koch et al., 2021; Mora et al., 2023; Rinaldi et al., 2022). The mass and energy conservation equations for multiphase, multicomponent flows are spatially and temporally discretized based on traditional numerical schemes. The numerical models stick to the rigorous governing Partial Differential Equations (PDEs), thus representing the high-fidelity physics processes. However, numerical modeling of CO₂ subsurface flows is usually computationally expensive and time-consuming, due to the geological heterogeneity, complex fluid thermodynamics and long-term operational lifespan. Particularly when dealing with inverse modeling and uncertainty quantification (Chen et al., 2020; Kitanidis, 2015; Oladyshkin et al., 2013), a large number of forward numerical simulations need to be conducted, which makes it unfeasible for real-time decision-making and operation optimization. For such cases it would be beneficial to have access to the desired outputs by a surrogate model, obviating the need for the resolution of intricate physical processes.

With the booming of accessible data and advances in GPU-accelerated paralleling processing, deep learning techniques, specifically Deep Neural Networks (DNNs), have gained widespread popularity in surrogate modeling for their universal approximation ability. Substantial research has been dedicated to developing state-of-the-art network architectures for accurate prediction of target variables. Among them, Convolutional Neural Networks (CNNs) have witnessed great success in surrogate modeling of subsurface flow problems due to the efficient filter operations and fast convergence rate. Encoder-decoder architectures like U-Net and its variations (Mo et al., 2019; Stepien et al., 2023; Tang et al., 2021; Tariq et al., 2023; Wen et al., 2021, 2022; Xu et al., 2023; Yan et al., 2022b; Zhong et al., 2019) have been widely explored for predicting CO₂ pressure, saturation or other state variables of interest. These studies treated the prediction of subsurface CO₂ flows as an image-to-image regression task, in which CNNs were employed to approximate the nonlinear relationship between the input feature maps (e.g., permeability field) and the output state variables. In addition, neural operator-based Fourier Neural Operator (FNO) networks (Jiang et al., 2023; Li et al., 2021; Wen et al., 2022, 2023; Yan et al., 2022a) exhibit potential in directly learning the mappings between two infinite dimensional spaces through Fourier transformation and inverse Fourier transformation. The primary advantage of FNO lies in its resolution-invariant property, which enables its generalization across different discretized input-output pairs. Nevertheless, in the aforementioned studies, discrete time steps were explicitly input as an image with constant values denoting the current moment. This approach treated the evolution of the state response as a static process, resulting in the model unable to learn the temporal dependencies of the state variables in the subsurface flow system. Tang et al. (2022, 2020) proposed a “one-to-many” network architecture called recurrent R-U-Net to make time-sequence predictions. The geological feature maps were fed into the network to make simultaneous predictions of the state variables at the desired time steps. The limitation of recurrent R-U-Net was that it could only generate a fixed number of outputs, and the ability to extrapolate to unseen time steps was not proven by the authors. The extrapolation capability of surrogate models is a primary concern considering the long-term operational lifespan of GCS. Therefore, the development of a general surrogate model to approximate the complex spatial-temporal evolution of state variables over an extended duration becomes essential for effective GCS operation.

In this paper, an encoder-decoder convolutional-recurrent neural network (ED-ConvLSTM) is proposed to address the aforementioned challenges. Convolutional Long Short-Term Memory (ConvLSTM) is employed to capture the spatial-temporal dynamics, while the encoder and decoder serve as the information extraction and reconstruction module, respectively. The contributions of this study are summarized as follows: (1) Powerful encoder and decoder blocks enhanced with light-weighted attention modules are developed. ConvLSTM is applied to

efficiently propagate the low-dimensional spatial features through time, enabling the network to capture the dynamic patterns of the state variables. (2) An innovative loss function with gradient-based spatial regularization and temporal penalty terms is implemented to improve the prediction accuracy. (3) The capability of the proposed model to extrapolate to the future time steps that are unseen during training is proven. ED-ConvLSTM is capable of predicting complex long-term evolution of state variables with access to only portion of training data. (4) The generalizability of the proposed model to different geological settings is also illustrated. ED-ConvLSTM can produce decent predictions on different random geological fields which are unseen during training.

This manuscript is structured as follows. Section 2 elaborates the methodology of this study, including the governing equations for multiphase flow, the numerical simulation method used to generate training data, the detailed description of the neural network architecture, and the training strategy. In Section 3, the results of the proposed ED-ConvLSTM for extrapolating to unseen time steps and different geological settings are presented using U-Net as a baseline model. In Section 4, the model performance is evaluated, followed by conclusions of this study in Section 5.

2. Methodology

This section outlines the essential procedures for creating a surrogate model aimed at predicting the state variables in GCS. First, the governing equations of CO₂ flow in subsurface reservoirs are introduced, along with the details of numerical simulation and high-fidelity dataset preparation. Then, the proposed ED-ConvLSTM is presented, including the illustration of the overall architecture and the constituent blocks. After that, the training strategy of the neural network is detailed.

2.1. Governing equations

During the process of GCS in saline aquifers, the fluid phases mainly consist of a water-rich aqueous phase and a CO₂-rich gaseous phase. The aqueous phase is mainly water with some dissolved CO₂ and the gaseous CO₂ phase also contains minor vaporized water. Salt dissolution and precipitation may also be presented but are not considered in this study. The mass conservation equation for each component in porous media is given, as follows,

$$\frac{\partial}{\partial t}M^\kappa = -\nabla \cdot \mathbf{F}^\kappa + Q^\kappa \quad (1)$$

where κ denotes component CO₂ or water, M^κ is the mass accumulation term, \mathbf{F}^κ is the mass flux, and Q^κ is the sink or source term. For each component, the mass accumulation is computed as:

$$M^\kappa = \varphi \sum_\beta S_\beta \rho_\beta X_\beta^\kappa \quad (2)$$

where φ is porosity, S_β is the saturation of phase β (e.g., β = aqueous, gas), ρ_β is the density of phase β , and X_β^κ is the mass fraction of component κ in phase β .

The mass flux for each component κ can be written as the sum of advective fluxes (molecular diffusion is not considered for its minor impact in this scenario) over phases β ,

$$\mathbf{F}^\kappa = \sum_\beta X_\beta^\kappa \mathbf{F}_\beta \quad (3)$$

where individual phase flux is given by multiphase Darcy's law:

$$\mathbf{F}_\beta = -k \frac{k_{r\beta} \rho_\beta}{\mu_\beta} (\nabla p_\beta - \rho_\beta \mathbf{g}) \quad (4)$$

Here, k is the absolute permeability, $k_{r\beta}$ is the relative permeability to

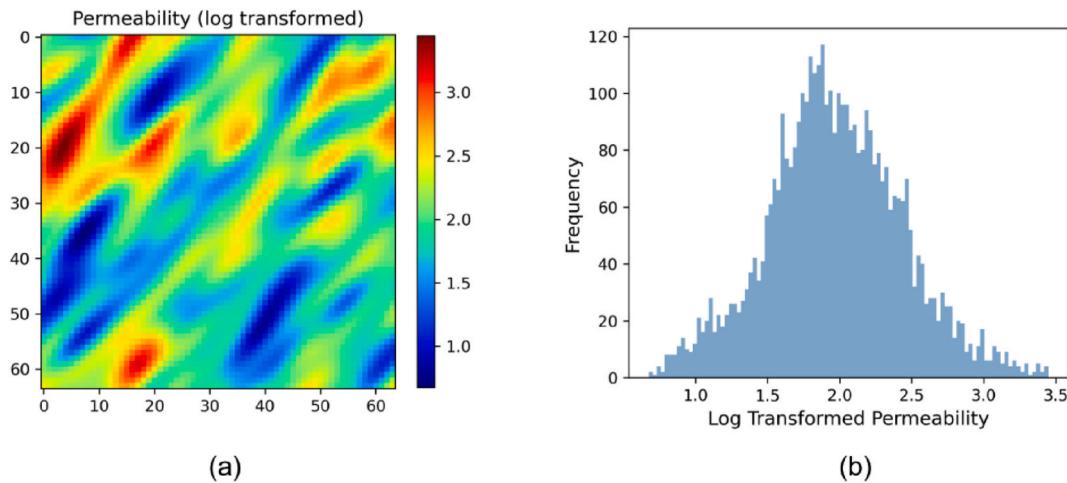


Fig. 1. The heterogeneous anisotropic permeability field (a) and its histogram (b).

phase β , g is the vector of gravitational acceleration, μ_β is the dynamic viscosity of phase β , and p_β is the fluid pressure in phase β , which is the sum of the pressure p of a reference phase (i.e., the gas phase) and the capillary pressure $p_{c\beta}$ (≤ 0),

$$p_\beta = p + p_{c\beta} \quad (5)$$

Besides, the saturation of each phase adds up to 1 according to the definition of saturation,

$$\sum_\beta S_\beta = 1 \quad (6)$$

And the sum of mass fraction of components in each phase is constrained by:

$$\sum_\kappa X_\beta^\kappa = 1 \quad (7)$$

2.2. Numerical simulation and dataset generation

The fully implicit finite volume simulator, TOUGH3 (Jung et al., 2017), is employed in this study to simulate the flow of CO₂ in the geologic porous media. TOUGH3 utilizes an integral finite difference method for spatial discretization, and a first-order fully implicit scheme for temporal differencing. It solves the discretized form of Eq. (1) using the Newton-Raphson iterative methods. Primary variables including pressure and saturation can be obtained for each grid block at each time step.

This study proposes a pseudo two-dimensional model, which is a squared sandstone reservoir with constant pressure boundaries to mimic an infinite large hydraulic system. The model has a length and width of 3 km and is discretized by a uniform grid of 64 × 64 blocks in the x-y plane. The initial pore pressure for this isothermal scenario is set at 24 MPa. An injection well is positioned at the center of the squared model and is operated under dynamic CO₂ injection. The injection rates exhibit monthly fluctuations, and this study accounts for a five-year operational period. Corey's relative permeability model and van Genuchten capillary pressure model are adopted in this study, with their respective parameter values sourced from Wen et al. (2021). The simulated state variables are the phase averaged pressure and CO₂ saturation. For simplicity, the two variables are respectively referred to as pressure and saturation in the rest of this manuscript.

The deep learning model is trained to predict the state variables given the geological setting and operational parameters. The used dataset is generated by the various combination of random geologic fields and injection rates. The perforation layer is set to be a heterogeneous anisotropic reservoir. The permeability is assumed to follow the

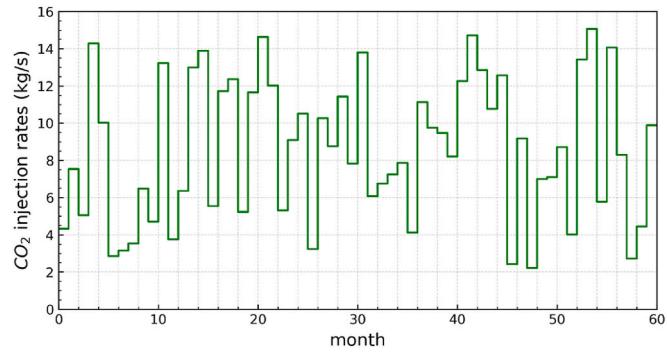


Fig. 2. The dynamic injection rates of CO₂.

log-normal distribution with a random anisotropy angle. The correlation lengths are fixed to be 421.875m and 140.625m along the two anisotropy directions respectively. One realization of such random permeability field and its histogram are shown in Fig. 1. The porosity is correlated to the permeability k by an empirical relationship (Zhong et al., 2019)

$$\varphi = 0.05 \log_{10}(k) + 0.15 \quad (8)$$

A total number of 5000 realizations of random fields are generated by the open-source python library GSTools (Müller et al., 2022). Each geologic realization is implemented with a distinct injection scheme. For one injection scheme, the time-dependent injection rate for each month is sampled via the Latin Hypercube Sampling method (Fig. 2). The surrogate model is trained using 4500 realizations, and subsequently validated and tested respectively with 250 realizations.

2.3. Deep learning-based surrogate model

Deep learning-based surrogate models are designed to map the high-dimensional input parameters to the desired output space by leveraging the universal approximation capacity of DNNs. In this study, a novel neural network is proposed to predict the spatial temporal evolution of the state variables during the injection period of GCS. In the following subsections, a detailed description of each component of the proposed neural network is presented.

2.3.1. ConvLSTM block

Convolutional Long Short-Term Memory (ConvLSTM) is a hybrid module that is capable of characterizing both spatial and temporal dependencies simultaneously (Shi et al., 2015). It combines the strengths

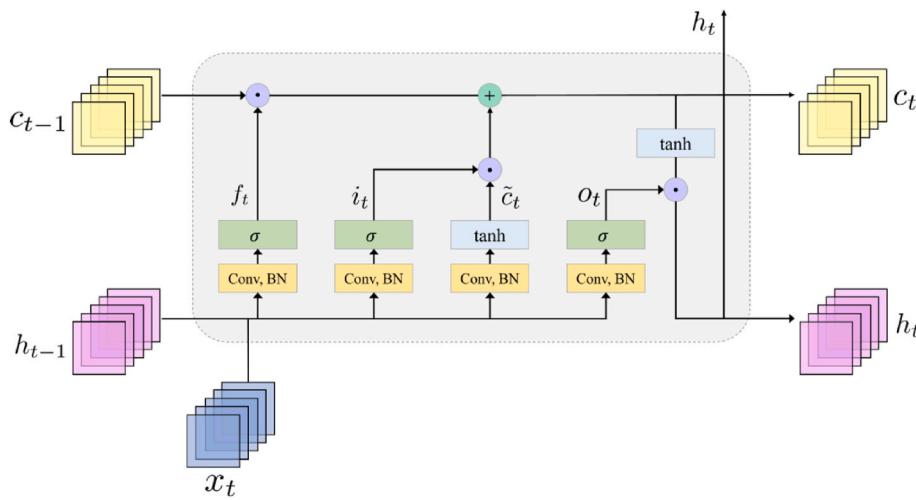


Fig. 3. Diagram of ConvLSTM cell.

of CNNs in processing spatial data and the memory retention capability of LSTM networks for sequence modeling. Essentially, ConvLSTM extends the classic LSTM network by incorporating convolutional operations into the memory cell, which enables the network to propagate spatial information through time. Moreover, the intricate gate operations in ConvLSTM cell mitigate the issue of gradient vanishing, enhancing its ability to effectively model long-term dependencies.

As shown in Fig. 3, the key mathematical equations of a ConvLSTM cell are expressed, as follows:

$$i_t = \sigma(W_{xi} * x_t + W_{hi} * h_{t-1} + b_i) \quad (9a)$$

$$f_t = \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + b_f) \quad (9b)$$

$$o_t = \sigma(W_{xo} * x_t + W_{ho} * h_{t-1} + b_o) \quad (9c)$$

$$\tilde{c}_t = \text{Tanh}(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c) \quad (9d)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (9e)$$

$$h_t = o_t \odot \text{Tanh}(c_t) \quad (9f)$$

where i_t, f_t, o_t are the input gate, forget gate and output gate, respectively, x_t is the input, h_t is the hidden state, c_t is the cell memory, \tilde{c}_t is the candidate cell memory. W and b are trainable weights and biases of the convolutional filters, respectively. The operation $*$ denotes the convolutional operation and \odot represents the element-wise multiplication

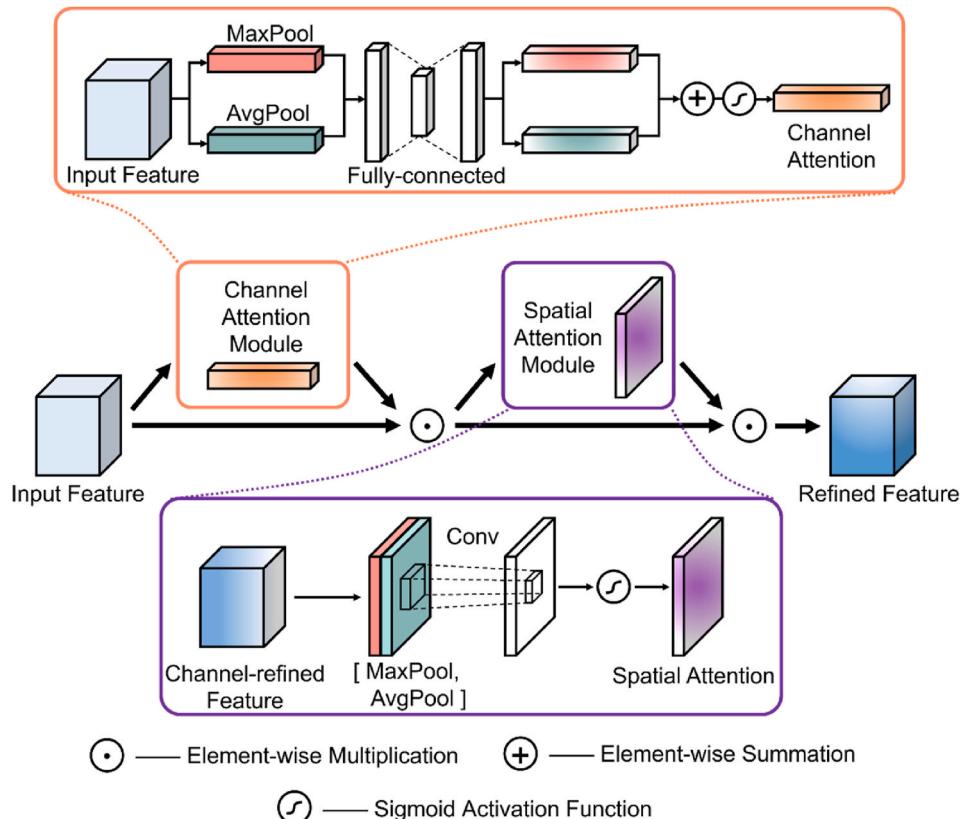


Fig. 4. Diagram of convolutional block attention module (CBAM) (after Woo et al. (2018)).

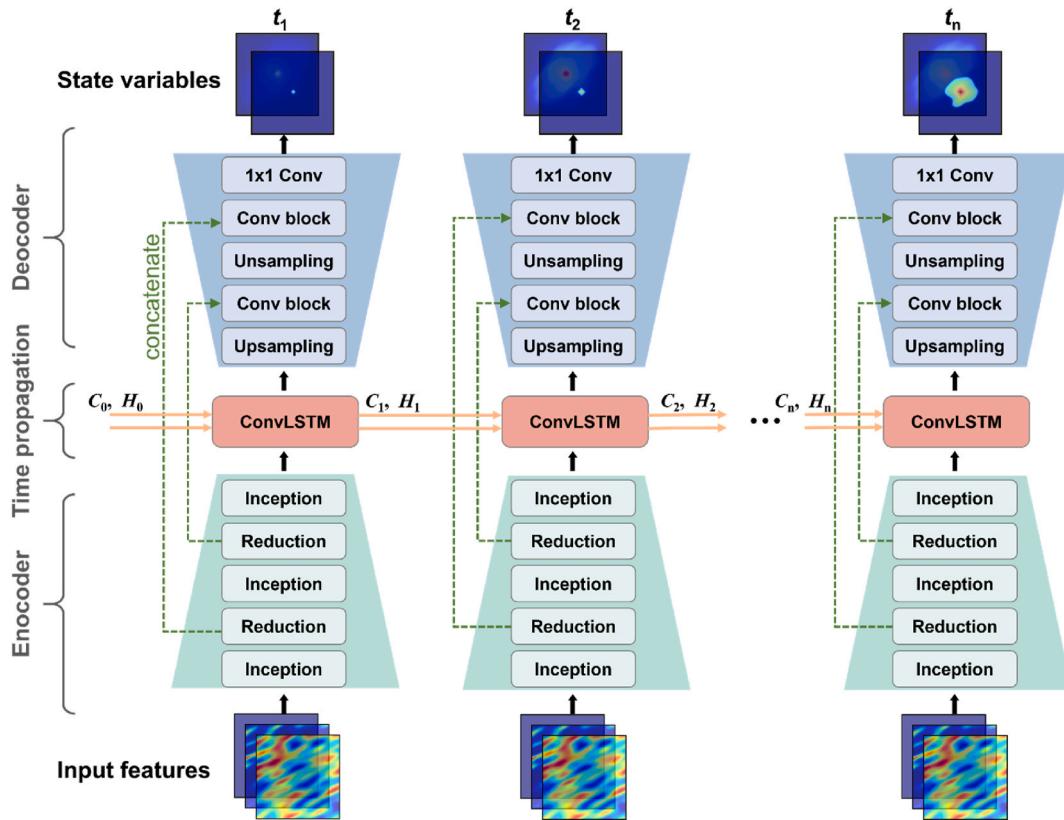


Fig. 5. The architecture of the proposed ED-ConvLSTM. It consists of an encoder, a decoder and one layer of ConvLSTM.

(i.e., Hadamard product). Activation functions like Sigmoid ($\sigma(\cdot)$) and Hyperbolic Tangent ($\text{Tanh}(\cdot)$) are used for non-linear mapping. Batch normalization (Ioffe and Szegedy, 2015) is also added after each convolutional operation for stability.

Given the input x_t , and the hidden state h_{t-1} and the cell memory c_{t-1} from the previous time step, ConvLSTM cell utilizes the gate operations to effectively propagate the temporal information. Specifically, the values of i_t , f_t and o_t are in the range of 0–1 due to the Sigmoid non-linear mapping, and they serve as weights to control the flow of history information used to update the current cell memory and hidden state. \bar{c}_t serves as a candidate new information by hyperbolic tangent scaling. The cell memory is updated with the information selected by the input gate and forgetting the unnecessary information determined by the forget gate (Eq. (9e)), while the hidden state is updated through the element-wise multiplication of output gate and hyperbolic tangent scaled cell memory (Eq. (9f)). By introducing the innovative cell memory and elaborate gate operations, ConvLSTM can selectively retain and update information over time, allowing it to capture long-term dependencies in sequential data.

2.3.2. Convolutional block attention module

Attention mechanism has been widely used in modern deep learning architectures (Ba et al., 2015; Niu et al., 2021; Vaswani et al., 2017; Woo et al., 2018; Xu et al., 2015). Attention forces the neural network to focus on important features and enhances the representation ability without increasing network complexity. This study explores improving the performance of the convolutional layers through the incorporation of Convolutional Block Attention Module (CBAM), a lightweight and effective attention module for CNNs (Woo et al., 2018). CBAM adaptively emphasizes or suppresses the feature to make the network focus on important information. As depicted in Fig. 4, CBAM consists of two sub-modules, i.e., the channel attention module and the spatial attention module, which facilitate the network to learn “what” and “where” to

focus on in the channel and spatial axes, respectively.

Given an input feature $F \in \mathbb{R}^{C \times H \times W}$, in order to obtain the refined feature $F' \in \mathbb{R}^{C \times H \times W}$, CBAM sequentially generates a 1D channel attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$ and a 2D spatial attention map $M_s \in \mathbb{R}^{1 \times H \times W}$. These attention maps are broadcasted accordingly and then multiplied with the feature as:

$$F' = M_c(F) \odot F \quad (10a)$$

$$F' = M_s(F') \odot F' \quad (10b)$$

in which the channel attention map is computed as:

$$M_c(F) = \sigma(W_1(W_0(\text{MaxPool}(F)) + W_1(W_0(\text{AvgPool}(F)))) \quad (11)$$

and the spatial attention map is computed as:

$$M_s(F') = \sigma(W_s * [\text{MaxPool}(F); \text{AvgPool}(F)]) \quad (12)$$

where $\{W_0, W_1, W_s\}$ are the only trainable parameters of the attention module. The lightweight characteristic of this module enhances the performance of traditional convolutional layers without the risk of overfitting. Note that CBAM takes an intermediate feature as input and produces the adaptively refined feature in the same dimension. This capability allows for seamless integration with various convolutional layers.

2.3.3. Network architecture of ED-ConvLSTM

ED-ConvLSTM, a novel network architecture, is proposed as the surrogate model for simulating GCS. As illustrated in Fig. 5, the network is comprised of an encoder, a decoder and one-layer of ConvLSTM for information flow through time. The encoder compresses the information from the original input features progressively. A low-dimensional latent feature is obtained after the encoder and then sent into the ConvLSTM layer to capture the dynamic dependencies in the temporal axis. The

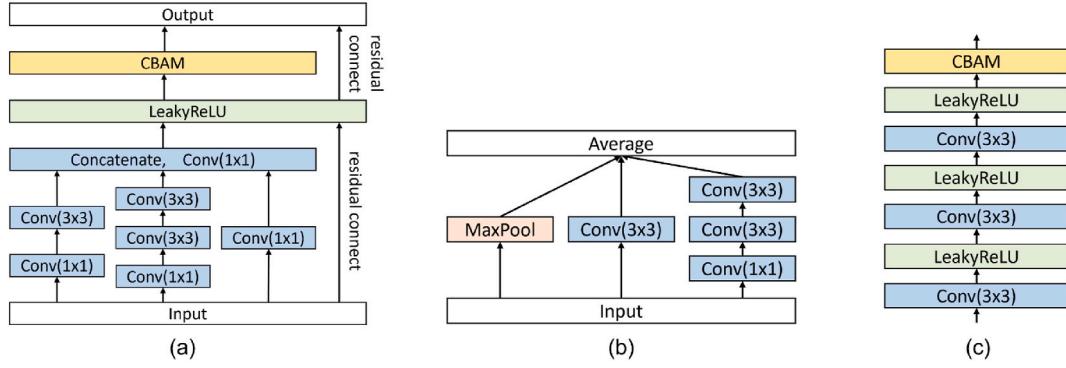


Fig. 6. Details of the (a) Inception block, (b) Reduction block and (c) convolutional block.

practice of learning the compressed features facilitates ConvLSTM to learn abstract pattern in the latent space, thus increasing the generalization ability of the network. Furthermore, the utilization of low-resolution compressed features alleviates the computational burden associated with convolutional operations. Zero initialization is employed to initialize the first hidden state and cell memory. ConvLSTM then learns and updates the hidden states and cell memories at the following time steps. The dynamic dependencies of this problem are automatically captured by ConvLSTM and stored in the hidden states and cell memories. The adoption of ConvLSTM is motivated by the recognition that GCS is a time-dependent physics process, where the preceding states exert influence on the states at the subsequent time steps. It should be highlighted that thanks to the captured dynamic patterns stored in ConvLSTM, the proposed network is capable of extrapolating accurately to unseen future time steps. The decoder maps the features from the latent space to the target solution space, i.e., the state variables of GCS. Two jump connections are applied to copy and concatenate the intermediate features from the encoder to the decoder to preserve crucial information. The input features include geological maps like permeability and porosity. The information of the well location and time-varying injection rates are merged into one feature map, in which the well position is represented as a non-zero number with its value denoting the current injection rate, and the reservoir is denoted as zero values, as Yan et al. (2022b) used. The output of the network is the state variables including pressure and saturation. The network is not designed as autoregressive to mitigate error accumulation.

A detailed description of the blocks in the encoder and the decoder is presented here. The graphical illustration of these blocks is shown in Fig. 6. The Inception and Reduction block proposed by Google (Szegedy et al., 2016) are employed as powerful feature extractors. Instead of going deeper, the Inception block widens its convolutional layer by parallel filters to learn the spatial features hierarchically. Instead of using ReLU as in the original paper (Szegedy et al., 2016), LeakyReLU is applied after the fusion of multi-level features for its better performance based on numerical experiments. The Inception block is further enhanced by CBAM to seamlessly suppress unnecessary information. Residual connections (He et al., 2016) are added for the efficient flow of information. The Reduction block reduces the dimension of the input feature by a scale of 2. There are 3 branches for dimension reduction. The output of each branch is fused by element-wise average, which is different from the concatenation operation in the original paper (Szegedy et al., 2016). The Reduction block produces a denser representation in a lower dimension, which leads to a decrease of the computational burden. For the convolution blocks in the decoder, 3 layers of convolutional layer and LeakyReLU are stacked upon each other. CBAM is also integrated into this block to enhance its performance. The Upsampling block is applied to upscale the low-resolution feature maps into the high-resolution space. A bilinear upsampling method is employed to upscale the input by a factor of 2. In essence, the Reduction block and the Upsampling block enable the ConvLSTM to focus on the compressed

representation of features in the latent space, which improves the computational efficiency and mitigates the memory burden for long sequence data.

2.4. Loss function and model training

In this study, the surrogate model takes an input feature with dimensions $n_t \times 3 \times 64 \times 64$, where n_t denotes the total number of time steps (i.e., one time step represents one month). All the input features are normalized into 0 and 1 by min-max scaling. The output of the model is the state variables (pressure and saturation). Two networks of the same architecture are trained to predict pressure and saturation, respectively. Therefore, the dimension of the output is $n_t \times 1 \times 64 \times 64$ for either pressure or saturation. Notice that the model is trained using data from a segment of the entire period. The first 24 months of data is used to train the surrogate model, while the model is required to predict all the 60 months of data to test the extrapolation ability to unseen future time steps.

The aim of this sequence-to-sequence regression task is to find the optimal model which minimizes the discrepancy between the sequence of the state maps generated from the high-fidelity simulations and that predicted by the surrogate model. During the training process, the parameters of the optimal model (e.g., weights and biases of each layer) are determined by minimizing the defined loss function. A customized loss function usually helps enforce the network to be optimized to the desired direction, as explored with physics-informed neural network (Raissi et al., 2019), theory-guided neural network (Wang et al., 2021) and thermodynamics-informed neural network (Zhang et al., 2024). Inspired by the Sobolev training technique proposed by Czarnecki et al. (2017), this study minimizes not only the difference between the target values and the predictions, but also the difference in their derivatives. Furthermore, ConvLSTM usually struggles at initial time steps because of the “cold start” issue (Zhu et al., 2016). An extra temporal penalty term is added to mitigate the relatively large errors at initial time steps. Therefore, the loss function is defined as follows:

$$L(\theta) = \frac{1}{n_s} \frac{1}{n_t} \sum_{i=1}^{n_s} \sum_{t=1}^{n_t} \|y_i^t - \hat{y}_i^t\|_p + \lambda_1 \frac{1}{n_s} \frac{1}{n_t} \sum_{i=1}^{n_s} \sum_{t=1}^{n_t} \|y_i^t - \hat{y}_i^t\|_p \\ \times \sum_{j=1}^k \|D^j y_i^t - D^j \hat{y}_i^t\|_p + \lambda_2 \frac{1}{n_s} \frac{1}{n_t} \sum_{i=1}^{n_s} \sum_{t=1}^{n_t} \|y_i^t - \hat{y}_i^t\|_p \quad (13)$$

where n_s is the number of training samples, n_t is the number of training time steps, k is the highest order of derivatives, n_{ti} is the number of penalized initial time steps, y_i^t is the ground truth state variable, \hat{y}_i^t is the prediction made by the surrogate model, D^j denotes the j -th order of derivatives with respect to spatial coordinates, $\|\cdot\|_p$ represents the p -order norm, and λ_1 and λ_2 are weights for the spatial and temporal penalty terms, respectively. For both pressure and saturation, the initial ten time steps and the first spatial derivatives are penalized. λ_1 equals

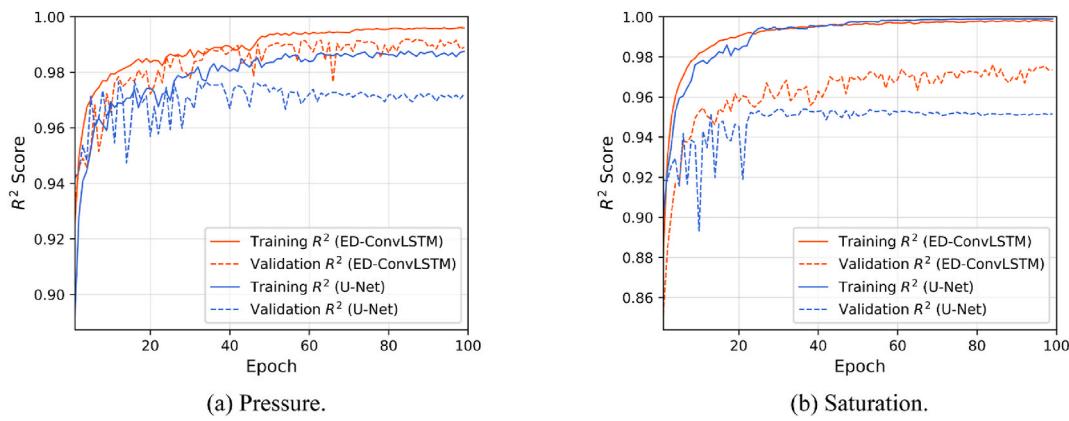


Fig. 7. Evolution of R^2 scores of ED-ConvLSTM and U-Net during the training and validation stage for (a) pressure and (b) saturation.

0.8 and λ_2 equals 10.

In order to calculate the derivatives efficiently, non-trainable 3×3 convolutional filters are employed, using the central difference discrete numerical approximations based on the finite difference method (Wang et al., 2021). Replicate padding is used considering the constant value boundary condition. This study uses the L^1 norm to predict the pressure maps, while uses the L^2 norm to predict the saturation maps based on trial and error. The goal of training is to find θ^* , and θ^* is obtained by minimizing $L(\theta)$:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} L(\theta) \quad (14)$$

The model parameter θ is automatically updated by the back-propagation through time (Werbos, 1990). Hyper-parameters are tuned to find their optimal values. In this study, the learning rate is initialized at 1e-3 and decreases automatically according to the performance of the validation set. Weight decay is set to be 1e-3 to alleviate overfitting. The number of channels of the convolutional filters is fine-tuned as well. The details of the optimal architecture are shown in Appendix A. The Adam optimizer is employed to minimize the loss function. The mini-batch size is chosen to be 32. The proposed neural network is implemented with the open-source deep learning framework PyTorch (Paszke et al., 2019).

3. Results

In this section, the model performance evaluating metrics are first presented, followed by the illustration that ED-ConvLSTM outperforms U-Net within the training time period, and is also capable of extrapolating to unseen future time steps. The generalizability of ED-ConvLSTM to different geological settings (e.g., different permeability fields) is then analyzed.

3.1. Model performance evaluating metrics

The performance of the surrogate model is quantitatively evaluated based on the mathematical metrics including normalized root mean square error (NRMSE), normalized absolute percentage error (NAPE), coefficient of determination (R^2) and structural similarity index measure (SSIM). The definition of each metric is given below:

$$NRMSE = \sqrt{\frac{\frac{1}{n_e} \sum_{i=1}^{n_e} (y_i - \hat{y}_i)^2}{y_{max} - y_{min}}} \quad (15)$$

$$NAPE = \frac{|y_i - \hat{y}_i|}{y_{max} - y_{min}} * 100 \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_e} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_e} (y_i - \bar{y})^2} \quad (17)$$

where n_e is the number of testing samples to be evaluated, y_i and \hat{y}_i are the ground truth and the network prediction, respectively, and y_{max} and y_{min} are the maximum and minimum of the true state variable values, respectively. The residual errors are normalized to enable the comparison between different state variables. R^2 is a normalized metric. A value of R^2 closer to 1 represents a higher prediction accuracy.

SSIM is used for measuring the similarity between two images. The maximum value of SSIM is 1, indicating identical images, while the minimum value is 0, indicating different images. SSIM is expressed as:

$$SSIM = \frac{1}{M} \sum_{m=1}^M \frac{(2\mu_{u,m}\mu_{v,m} + K_1^2)(2\sigma_{uv,m} + K_2^2)}{(\mu_{u,m}^2 + \mu_{v,m}^2 + K_1^2)(\sigma_{u,m}^2 + \sigma_{v,m}^2 + K_2^2)} \quad (18)$$

where u and v are the respective true and predicted state variables (images), M is the number of local windows used to evaluate the images, μ and σ denote the mean and standard deviation, respectively, and K_1 and K_2 are small constants to avoid zeros values in the denominator.

To further demonstrate the advantages of the proposed ED-ConvLSTM, the U-Net model (Ronneberger et al., 2015) is implemented as a baseline model. Note that the discrete time steps are taken as additional input features as U-Net does not have recurrent connections to take into account time dependencies. The well control parameters at specific time steps along with geological information are used in U-Net, as in Section 2.3.3. The training dataset is same for both networks. More details about the architecture of U-Net are provided in Appendix B.

3.2. Model performance within the training time period

ED-ConvLSTM and U-Net are trained with 100 epochs on a GPU (NVIDIA GeForce RTX 4090). Pressure and saturation are trained, separately. Fig. 7 shows the evolution of R^2 scores during the training and validation stages. For either ED-ConvLSTM or U-Net, the R^2 score gradually increases and converges to a stable level as epoch goes on, indicating the learning process of the two networks. The R^2 score of ED-ConvLSTM on the validation dataset takes over that of U-Net after certain epochs. The prediction of saturation is prone to the issue of overfitting due to the more localized pattern of CO₂ plume.

The model performance is evaluated on the test dataset with 250 samples. These samples have distinct random fields and injection schemes, which have not seen by the model during the training stage. The results within the training time period (i.e., from 1 to 24 months) are

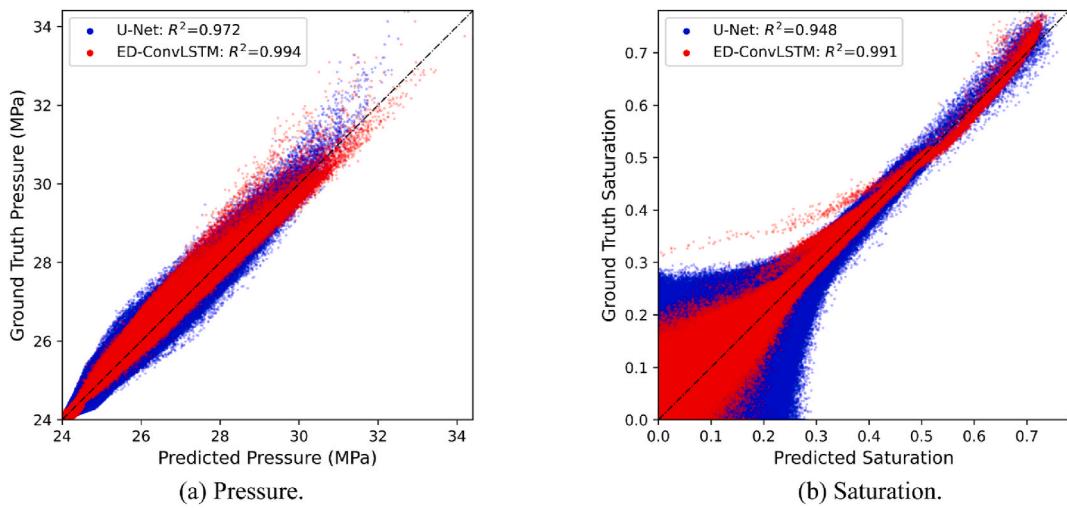


Fig. 8. Scatter plots of (a) pressure and (b) saturation comparing predicted and true values in the test dataset.

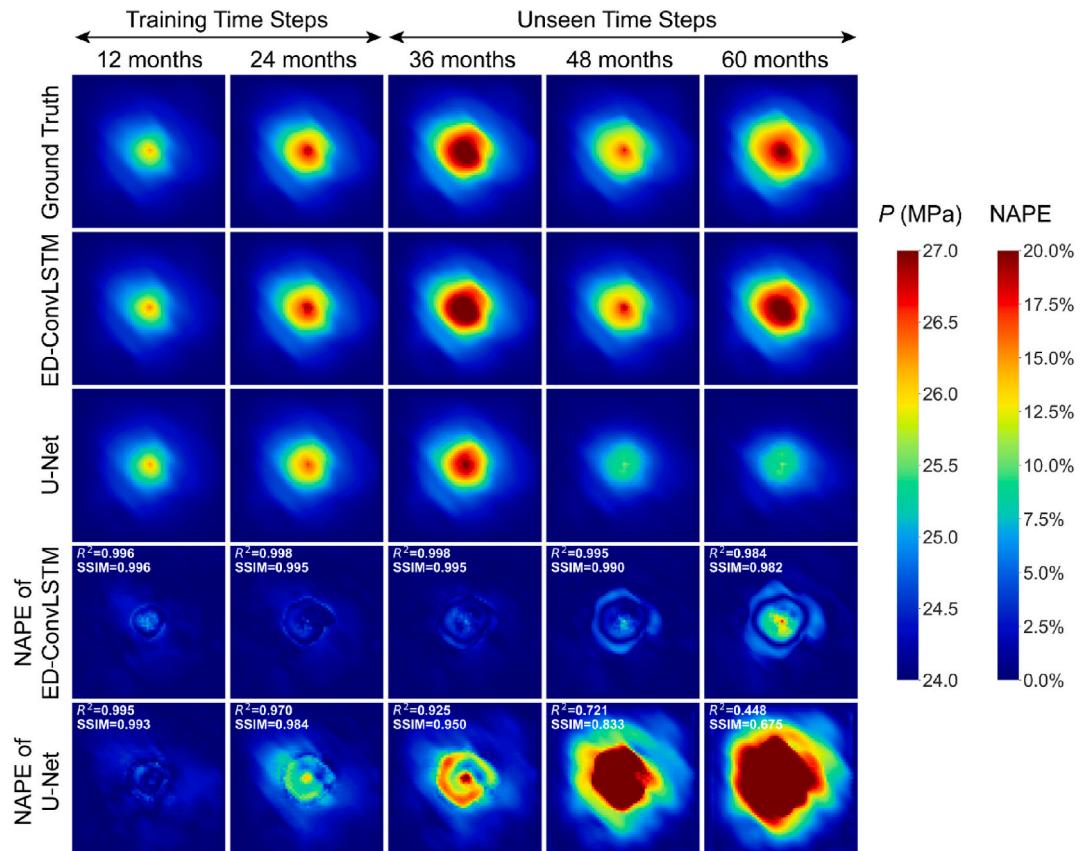


Fig. 9. Comparison of pressure between the ground truth, the predictions by ED-ConvLSTM and the predictions by U-Net. The first row is the pressure distribution obtained from high-fidelity numerical simulation. The second and third rows show the pressure predictions by ED-ConvLSTM and U-Net, respectively. The fourth and fifth rows are the normalized absolute percentage error (NAPE) of the two models. The first two columns represent the results within the training time period while the last three columns are the results at unseen time steps.

first presented. Fig. 8 shows the scatter plots comparing the predicted state variables and the ground truth values for each grid in the test dataset. The proposed ED-ConvLSTM achieves R^2 scores of 0.994 and 0.991 for pressure and saturation, respectively, while U-Net predicts the two state variables with R^2 of 0.972 and 0.948. It is observed that the predictions generated by both models are more accurate for lower pressure values, while the result of saturation shows the opposite trend. Overall, the proposed ED-ConvLSTM outperforms the U-Net model for

both pressure and saturation predictions within the training time period.

3.3. Extrapolation to unseen future time steps

In this subsection, the aim is to test if the model is able to predict the long-term evolution of the state variables beyond the training time period (i.e., 25–60 months). During training stage, the model is fed with data from 1 to 24 months. Note that the injection rates beyond the

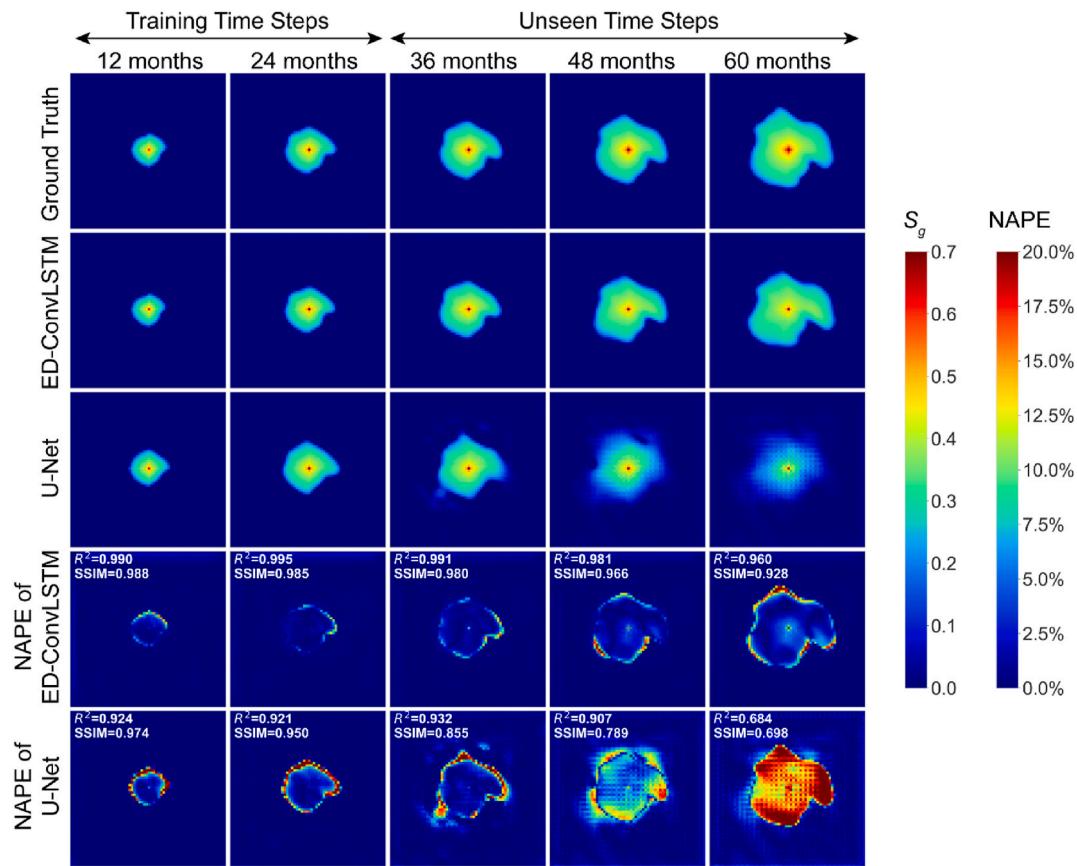


Fig. 10. Comparison of CO_2 saturation between the ground truth, the predictions by ED-ConvLSTM and the predictions by U-Net. The first row includes the saturation fields obtained from high-fidelity numerical simulation. The second and third rows respectively show the predictions generated by ED-ConvLSTM and U-Net. The fourth and fifth rows are the residual errors (represented by NAPE) of the two models. The first two columns show the results within the training time period while the last three columns are the results beyond the training time period.

training time period vary from month to month as well. The results of the baseline U-Net model are also presented for comparison.

Fig. 9 shows the pressure predictions produced by ED-ConvLSTM and U-Net and the comparison between the predictions and the ground truth for a randomly selected sample from the test dataset. The first row of

Fig. 9 is the ground truth obtained from high-fidelity numerical simulation. The second and third row are the predicted pressure by ED-ConvLSTM and U-Net, respectively. The last two rows show the normalized absolute percentage error (NAPE) of the two models. The evaluation metrics including R^2 and SSIM are calculated as well. Five snapshots ($t = 12, 24$ months within the training period, $t = 36, 48, 60$ months beyond the training period) are collected, depicting the evolving pressure distribution with time. Under the circumstance of time-varying injection rates, the distribution of pressure shows a complex dynamic pattern because of the interaction between pore pressure dissipation and injection induced over-pressurization. Results show that the pressure field predicted by ED-ConvLSTM exhibits a better match with the ground truth compared with that predicted by U-Net. The expansion and shrinkage of the over-pressurized zone are well captured by ED-ConvLSTM, except some minor overestimations around the injection well at the 60th month. R^2 scores and SSIM of ED-ConvLSTM are in the range from 0.982 to 0.996 at the five snapshots, while the performance of U-Net degrades dramatically at unseen time steps, with its R^2 and SSIM decreasing to 0.448 and 0.675, respectively. The region of mismatch between U-Net and the ground truth expands very fast, and the maximum NAPE value exceeds 20%. It is shown that ED-ConvLSTM performs better than U-Net in terms of pressure approximation, especially extrapolating to unseen time steps.

The results of saturation are presented in Fig. 10. The CO_2 plume expands as injection proceeds. The boundary delineating CO_2 saturation exhibits a complex dynamic pattern due to the highly heterogeneous geological fields and the time-varying injection rates. The predictions made by ED-ConvLSTM show a good agreement with the ground truth at the five snapshots. The evolution of CO_2 saturation is well captured by

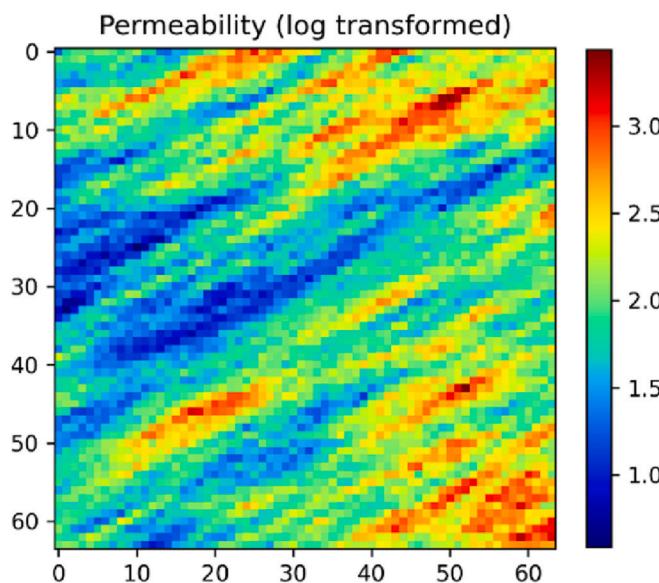


Fig. 11. Exponential anisotropy random permeability field.

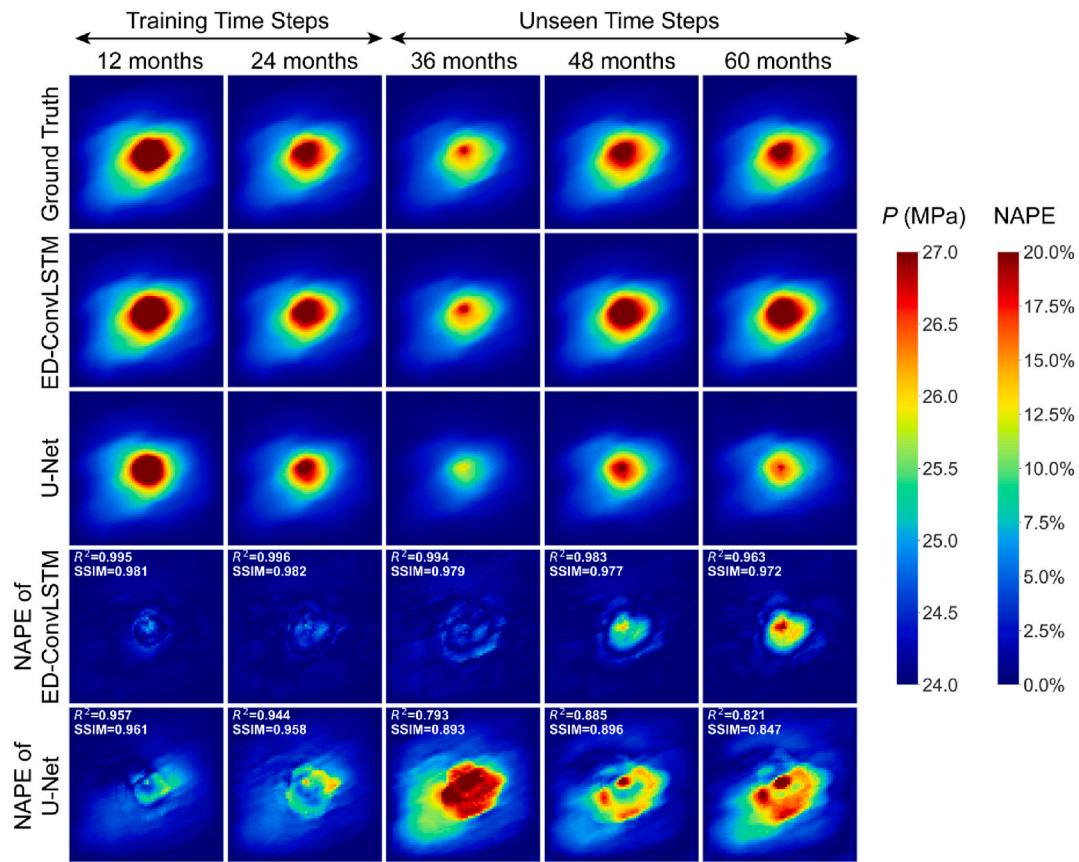


Fig. 12. Comparison of pressure between the ground truth, the predictions by ED-ConvLSTM and the predictions by U-Net on the exponential random field. From top to bottom, the rows show the ground truth from high-fidelity simulation, ED-ConvLSTM predictions, U-Net predictions, NAPE of ED-ConvLSTM, and NAPE of U-Net. The first two columns show the results within the training time period while the last three columns are the results at unseen time steps.

ED-ConvLSTM, with minor disparities at the boundaries of CO₂ plume and the well location. R^2 scores and SSIM of ED-ConvLSTM stay in the range from 0.928 to 0.995. However, the CO₂ plume predicted by U-Net become blurry and noisy when predicting beyond the training time period. R^2 scores and SSIM of U-Net respectively decrease to 0.694 and 0.698. The residual errors exist everywhere throughout the CO₂ plume at the 60th month, with maximum NAPE values surpassing 20%. These results demonstrate that ED-ConvLSTM is capable of extrapolating well beyond the training time period for both pressure and saturation prediction.

3.4. Generalization to different geological settings

In this subsection, different random geologic fields are considered to test the generalizability of ED-ConvLSTM on different geological settings that are not included in the training dataset. To be specific, different from the Gaussian random field adopted in the training dataset, exponential random permeability fields are generated with different orientations of anisotropy. The porosity is correlated to the permeability using Eq. (8). 250 realizations are generated to provide another blind testing dataset. One realization of such field is depicted in Fig. 11. In contrast to the Gaussian random field, the exponential random field exhibits abrupt spatial variations.

Fig. 12 shows the results of predicted pressure with the above exponential random field. Five snapshots, including the three unseen time steps, are presented. The over-pressurized zone induced by injection can be predicted by ED-ConvLSTM with excellent accuracy at all the time steps. R^2 scores and SSIM vary between 0.963 and 0.996 for ED-ConvLSTM, which are higher than those of U-Net at all five snapshots. A small region of mismatch around the well location only appears after

48 months. There are obvious discrepancies between U-Net and the ground truth throughout the simulated time, with the largest mismatch region appearing at the 36th month. The results of CO₂ saturation are shown in Fig. 13. The profile of saturation is more complex than the cases with Gaussian distribution, where zigzags are observed on the boundaries. The saturation prediction of ED-ConvLSTM exhibits a decent accuracy (i.e., R^2 scores is above 0.930 and SSIM is above 0.896). The extension of CO₂ plume is well characterized by ED-ConvLSTM despite minor difference at the boundaries and the well location, even in the cases where the large uncertainty of geological fields is considered. In contrast, the predictions by U-Net are less accurate, especially at unseen time steps. R^2 scores and SSIM respectively decrease to 0.598 and 0.624 at the 60th month. The shape of CO₂ plume cannot be captured after 24 months due to the poor extrapolating ability of U-Net. Therefore, it is validated that ED-ConvLSTM is able to either generalize to the predictions with exponential random fields or extrapolate to unseen future time steps. This indicates the reliability of ED-ConvLSTM for predicting complex long-term evolving configurations in the carbon sequestration scenario.

4. Discussion

In this section, ablative experiments are firstly presented, followed by the comparison of time consumption of different models. Then, the temporal performance is evaluated.

4.1. Ablative experiments

To justify the effectiveness of the proposed deep learning model, ablative experiments are implemented. The benefits of the use of the

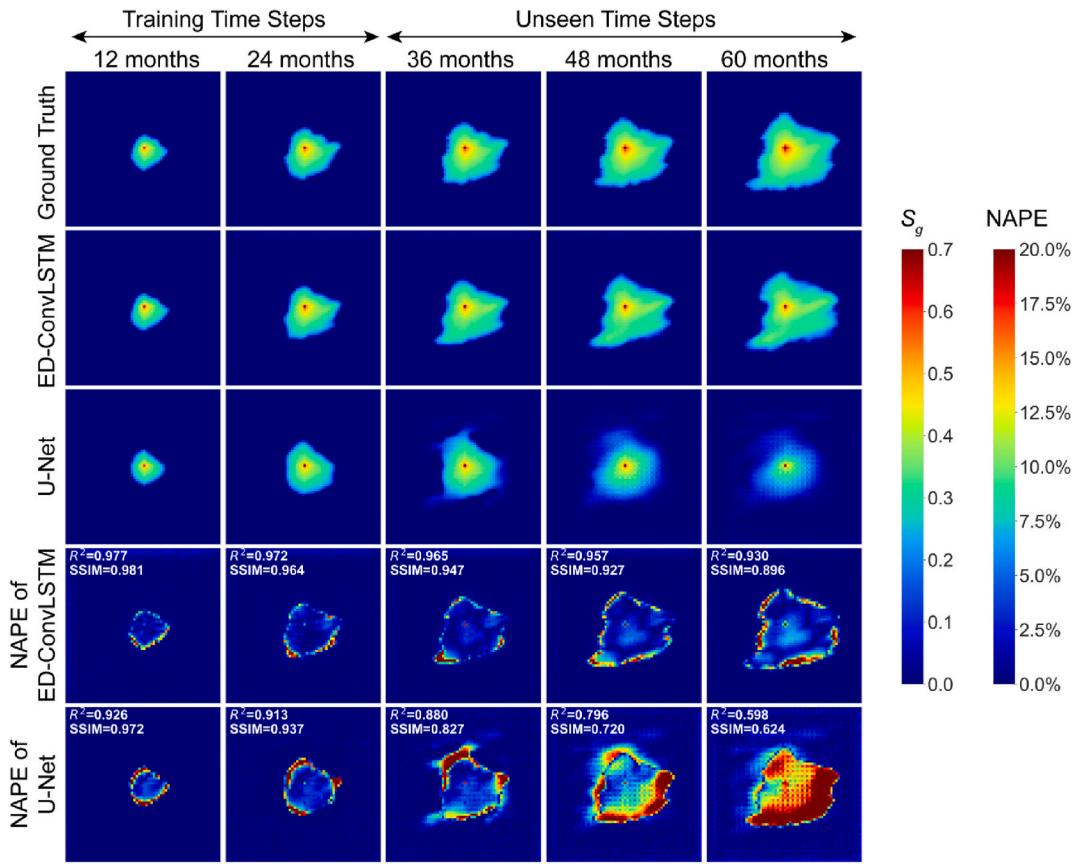


Fig. 13. Comparison of saturation between the ground truth, the predictions by ED-ConvLSTM and the predictions by U-Net on the exponential random field. From top to bottom, the rows show the ground truth from high-fidelity simulation, ED-ConvLSTM predictions, U-Net predictions, NAPE of ED-ConvLSTM, and NAPE of U-Net. The first two columns show the results within the training time period while the last three columns are the results beyond the training period.

Table 1
Results of ablative experiments.

	R^2 for pressure	R^2 for saturation
ED-ConvLSTM	0.9890	0.9642
ED-ConvLSTM without CBAM module	0.9764	0.9298
ED-ConvLSTM without penalty loss terms	0.9884	0.9135

Table 2
Time consumption of ED-ConvLSTM, U-Net and numerical simulation.

	Training time (minute)	Inference time (second/case)
ED-ConvLSTM	246	0.7
U-Net	60	1.4
Numerical simulation	–	91.8

CBAM module and the customized loss function (Eq. (13)) are investigated. The experimental settings include full ED-ConvLSTM, ED-ConvLSTM without CBAM module, ED-ConvLSTM trained without penalty terms (i.e., vanilla L^1 / L^2 loss). The comparison results are presented in Table 1. The structure of ED-ConvLSTM without CBAM module achieves the lowest R^2 score for both pressure and saturation. In addition, ED-ConvLSTM trained without penalty terms also exhibits inferior performance compared to the one trained with the proposed loss function. Therefore, the ablative experiments validate the significance and effectiveness of the CBAM module and the customized loss function.

4.2. Computational time comparison

Time consumptions of ED-ConvLSTM, U-Net and numerical simulation are compared at the training and inference stages. The results are presented in Table 2. At the training stage, ED-ConvLSTM takes longer time compared with U-Net. This is caused by the recurrent architecture of ED-ConvLSTM, disabling the parallel computing ability of GPU. Nevertheless, once trained, ED-ConvLSTM takes two folds less time than U-Net at the inference stage. It is also noteworthy that ED-ConvLSTM saves up to 99% inference time compared to traditional numerical simulations. These comparisons show the outstanding efficiency of ED-ConvLSTM.

4.3. Temporal performance evaluation

Fig. 14 (a) to (d) show the evolution of NRMSE and SSIM for pressure and saturation prediction. Curves marked with circles denotes predictions by ED-ConvLSTM, and those marked with triangles represents predictions by U-Net. Solid and dashed curves represent results from Gaussian and exponential random fields, respectively. The shaded regions around curves are the 95% confidence interval. The training time period and the unseen time period are depicted as yellow and cyan colored areas. In general, the performance of ED-ConvLSTM is better than U-Net. The errors of U-Net monotonically increase as time proceeds and even explode at time steps that are not seen during training stage. In contrast, ED-ConvLSTM does not suffer from significant error accumulation, leading to the stable and remarkable performance in both the training and extrapolation phases. The accuracy of ED-ConvLSTM is much higher than that of U-Net at the unseen time steps, which further demonstrates the excellent extrapolation ability of ED-ConvLSTM. It can

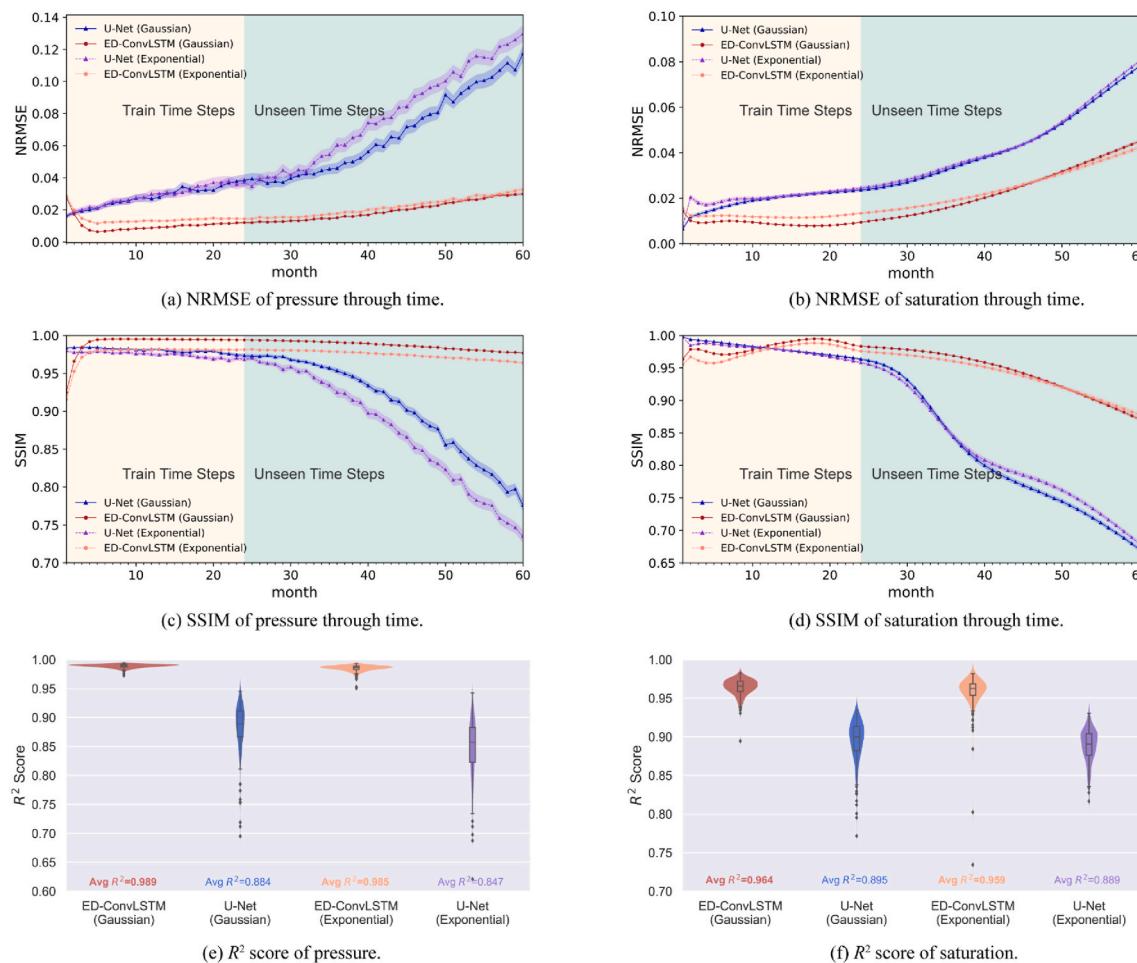


Fig. 14. Comparison of the performance between ED-ConvLSTM and U-Net. (a)–(d) Temporal variation of NRMSE for pressure, NRMSE for saturation, SSIM for pressure, and SSIM for saturation of ED-ConvLSTM and U-Net on Gaussian and exponential random fields. The shaded regions surrounding the curves denote the 95% confidence interval. The curves in yellow and cyan areas represent model performance within and beyond the training time period, respectively. (e)–(f) Violin plots of R^2 score for pressure and saturation prediction. The violin geometry denotes the probability distribution and the dots denotes the outliers.. (To understand the color references in this figure legend, please refer to the web version of this article.)

be observed that R^2 score of ED-ConvLSTM is always higher than U-Net for either Gaussian or exponential random fields (Fig. 14 (e) and (f)). More importantly, the predictions by ED-ConvLSTM exhibit a relatively centralized distribution, whereas those by U-Net tend to be more dispersive. This further shows the stable and robust performance of ED-ConvLSTM.

It can be observed that NRMSE of ED-ConvLSTM starts from an initial high value, and then decrease at the subsequent several time steps, especially for pressure. Similar trend can also be observed by SSIM. This phenomenon can be attributed to the following three factors. First, zero initialization of the first hidden state and cell memory is adopted in this study, leading to the issue named “cold start” (Zhu et al., 2016). There is no historical information contained in the first hidden state or cell memory, which makes the network struggle to produce decent predictions at the initial time steps. Nevertheless, the network learns to capture the temporal dependencies after several time steps thanks to the retained hidden states and cell memories by ConvLSTM, thus capable of producing more accurate predictions at the subsequent time steps. Second, the network is optimized by Backpropagation Through Time (BPTT), where the parameters are updated by summing the gradients at all the time steps. The gradients for the initial time steps might become vanishingly small due to BPTT, making it challenge for the network to learn temporal correlations for these time steps (Bengio et al., 1994; Pascanu et al., 2013). Third, numerical simulations usually employ small time step intervals at the initial period to ensure numerical

stability, whereas ED-ConvLSTM adopts uniform time step scheme in a data-driven manner. Although the data-driven model does not suffer from the numerical stability issue, it finds difficulty in capturing the evolution pattern marching from the homogeneous initial condition to the transient response during the initial period. Pretraining on dataset obtained from smaller time steps or refining the initial time steps might help mitigate this issue but is out of the focus of this study. However, from the perspective of practical application, the minor under-performance during the first several months can be acceptable considering the long-term operational lifespan (100–1000 years) of GCS projects.

5. Concluding remarks

This study presents a novel deep learning surrogate model ED-ConvLSTM to predict the complex spatial temporal evolution of pressure and saturation under dynamic injection rates. ED-ConvLSTM is able to map the high dimensional input features to the desired output state variables fast and accurately by obviating the need for resolving large algebra linear systems. It is demonstrated that ED-ConvLSTM is capable of predicting the state variables beyond the training period, while the standard U-Net fails in such extrapolation tasks. The generalizability of ED-ConvLSTM is also evaluated on different geological settings. It is shown that ED-ConvLSTM can robustly and accurately characterize the evolution of pressure and saturation under complex geological

conditions.

This study highlights the importance of incorporating recurrent connections for the prediction of state variables at unseen future time steps. Due to the elaborate gate operations of the ConvLSTM, there is limited error accumulation with time and the dynamic patterns of the state variables are “remembered” by the cell memories. These enable ED-ConvLSTM to achieve good accuracy within the training time period and even extrapolate to unseen time steps. The proposed ED-ConvLSTM is compared with U-Net, a one-shot architecture without recurrent connections. Discrete time steps are taken as additional input features in U-Net so as to consider time dependencies. In fact, this practice regards the state variables at each time step as time-independent samples and ignores the influence of history on the evolution of state variables. These conclusions have key implications in the effective operation of GCS, since accurately capturing pressure and saturation of CO₂ in the long-term run is critical. By leveraging the recurrent connections, ED-ConvLSTM has the potential to accelerate the prediction of state variables and be applied in real-time CO₂ injection optimization tasks, even with a limited amount of initial training data.

Future work in this area should address the following topics. Realistic 3-dimension model needs to be developed. This can be seamlessly integrated into the proposed framework by introducing the 3-dimension convolution operations. Moreover, the application in history matching and real-time CO₂ injection optimization can be explored by employing ED-ConvLSTM as a surrogate model. In addition, the inclusion of physics constraints into the loss function can be considered as a viable strategy to enhance the interpretability of the model.

CRediT authorship contribution statement

Zhao Feng: Writing – original draft, Visualization, Validation,

Appendix A. Detailed architecture of ED-ConvLSTM

Table A.1
ED-ConvLSTM model summary. n_t denotes the channel for time step.

Name	Type	Output shape
Input		$n_t, 64, 64, 3$
Encoder 1	Inception Residual Block	$n_t, 64, 64, 16$
Encoder 2	Reduction Block	$n_t, 32, 32, 16$
Encoder 3	Inception Residual Block	$n_t, 32, 32, 64$
Encoder 4	Reduction Block	$n_t, 16, 16, 64$
Encoder 5	Inception Residual Block	$n_t, 16, 16, 128$
ConvLSTM	Convolutional LSTM Layer	$n_t, 16, 16, 128$
Decoder 1	Upsampling Block	$n_t, 32, 32, 128$
Decoder 2	Convolutional Block	$n_t, 32, 32, 64$
Decoder 3	Upsampling Block	$n_t, 64, 64, 64$
Decoder 4	Convolutional Block	$n_t, 64, 64, 32$
Decoder 5	1 × 1 Convolutional Layer	$n_t, 64, 64, 1$
Output		$n_t, 64, 64, 1$

Appendix B. Detailed architecture of U-Net

U-Net is a typical encoder-decoder network. First proposed by Ronneberger et al. (2015) for medical image segmentation, U-Net has now been widely used in image-to-image regression tasks. This study adopts U-Net as a baseline model. As shown in Fig. B1, U-Net is comprised of contracting layers for downsampling (left side) and expansive layers for upsampling (right side). The model architecture is detailed in Table B1.

Software, Methodology, Investigation, Conceptualization. **Zeeshan Tariq:** Writing – review & editing, Validation, Methodology, Data curation. **Xianda Shen:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Bicheng Yan:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Xuhai Tang:** Software, Resources. **Fengshou Zhang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by the National Key Research and Development Project, China (No. 2023YFE0110900) and National Natural Science Foundation of China (No. 42320104003, 42077247). Z.F., Z.T., B.Y. acknowledge the sponsorship from King Abdullah University of Science and Technology (KAUST), Saudi Arabia (BAS/1/1423-01-01).

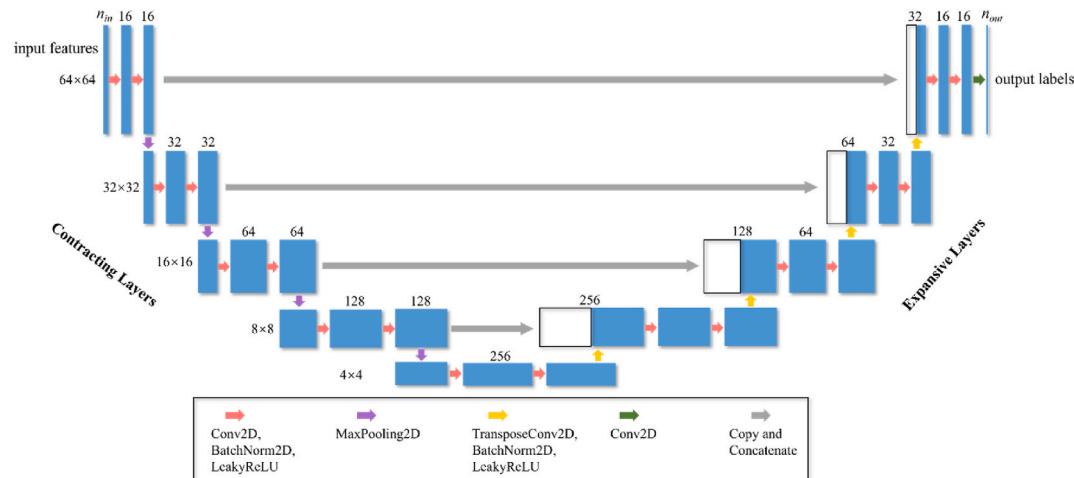


Fig. B.1. Architecture of U-Net in this study.

Table B.1
U-Net model summary

Name	Type	Output shape
Input		64, 64, 4
Encoder 1	Conv2d/Conv2d/MaxPool2d	32, 32, 16
Encoder 2	Conv2d/Conv2d/MaxPool2d	16, 16, 32
Encoder 3	Conv2d/Conv2d/MaxPool2d	8, 8, 64
Encoder 4	Conv2d/Conv2d/MaxPool2d	4, 4, 128
Conv	Conv2d/Conv2d	4, 4, 256
Decoder 1	TransposeConv2d/Conv2d/Conv2d	8, 8, 128
Decoder 2	TransposeConv2d/Conv2d/Conv2d	16, 16, 64
Decoder 3	TransposeConv2d/Conv2d/Conv2d	32, 32, 32
Decoder 4	TransposeConv2d/Conv2d/Conv2d	64, 64, 16
Conv	1 × 1 Conv2d	64, 64, 1
Output		64, 64, 1

References

- Ajaiyi, T., Gomes, J.S., Bera, A., 2019. A review of CO₂ storage in geological formations emphasizing modeling, monitoring and capacity estimation approaches. Pet. Sci. 16, 1028–1063. <https://doi.org/10.1007/s12182-019-0340-8>.
- Aminu, M.D., Nabavi, S.A., Rochelle, C.A., Manovic, V., 2017. A review of developments in carbon dioxide storage. Appl. Energy 208, 1389–1419. <https://doi.org/10.1016/j.apenergy.2017.09.015>.
- Ba, J., Mnih, V., Kavukcuoglu, K., 2015. Multiple Object Recognition with Visual Attention. <https://doi.org/10.48550/arXiv.1412.7755>.
- Bachu, S., Adams, J.J., 2003. Sequestration of CO₂ in geological media in response to climate change: capacity of deep saline aquifers to sequester CO₂ in solution. Energy Convers. Manag. 44, 3151–3175. [https://doi.org/10.1016/S0196-8904\(03\)00101-8](https://doi.org/10.1016/S0196-8904(03)00101-8).
- Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. 5, 157–166. <https://doi.org/10.1109/72.279181>.
- Birkholzer, J.T., Oldenburg, C.M., Zhou, Q., 2015. CO₂ migration and pressure evolution in deep saline aquifers. Int. J. Greenh. Gas Control, Special Issue commemorating the 10th year anniversary of the publication of the Intergovernmental Panel on Climate Change Special Report on CO₂ Capture and Storage 40, 203–220. <https://doi.org/10.1016/j.ijggc.2015.03.022>.
- Busch, A., Gensterblum, Y., 2011. CBM and CO₂-ECBM related sorption processes in coal: a review. Int. J. Coal Geol. 87, 49–71. <https://doi.org/10.1016/j.coal.2011.04.011>.
- Celia, M.A., Bachu, S., Nordbotten, J.M., Bandilla, K.W., 2015. Status of CO₂ storage in deep saline aquifers with emphasis on modeling approaches and practical simulations. Water Resour. Res. 51, 6846–6892. <https://doi.org/10.1002/2015WR017609>.
- Chen, B., Harp, D.R., Lu, Z., Pawar, R.J., 2020. Reducing uncertainty in geologic CO₂ sequestration risk assessment by assimilating monitoring data. Int. J. Greenh. Gas Control 94, 102926. <https://doi.org/10.1016/j.ijggc.2019.102926>.
- Czarnecki, W.M., Osindero, S., Jaderberg, M., Świrszcz, G., Pascanu, R., 2017. Sobolev training for neural networks. Adv. Neural Inf. Process. Syst. 30.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (pp. 448–456). pmlr.
- Jiang, Z., Zhu, M., Li, D., Li, Q., Yuan, Y.O., Lu, L., 2023. Fourier-mionet: fourier-enhanced multiple-input neural operators for multiphase modeling of geological carbon sequestration (preprint). SSRN. <https://doi.org/10.2139/ssrn.4408632>.
- Jung, Y., Pau, G.S.H., Finsterle, S., Polleyea, R.M., 2017. TOUGH3: a new efficient version of the TOUGH suite of multiphase flow and transport simulators. Comput. Geosci. 108, 2–7. <https://doi.org/10.1016/j.cageo.2016.09.009>.
- Kitanidis, P.K., 2015. Persistent questions of heterogeneity, uncertainty, and scale in subsurface flow and transport. Water Resour. Res. 51, 5888–5904. <https://doi.org/10.1002/2015WR017639>.
- Koch, T., Gläser, D., Weishaupt, K., Ackermann, S., Beck, M., Becker, B., Burbulla, S., Class, H., Coltrman, E., Emmert, S., Fettner, T., Grüninger, C., Heck, K., Hommel, J., Kurz, T., Lipp, M., Mohammadi, F., Scherrer, S., Schneider, M., Seitz, G., Stadler, L., Utz, M., Weinhardt, F., Flemisch, B., 2021. DuMuX 3 – an open-source simulator for solving flow and transport problems in porous media with a focus on model coupling. Comput. Math. Appl., Development and Application of Open-source Software for Problems with Numerical PDEs 81, 423–443. <https://doi.org/10.1016/j.camwa.2020.02.012>.
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A., 2021. Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv: 2010, 08895.
- Liu, S., Yuan, L., Zhao, C., Zhang, Y., Song, Y., 2022. A review of research on the dispersion process and CO₂ enhanced natural gas recovery in depleted gas reservoir. J. Pet. Sci. Eng. 208, 109682. <https://doi.org/10.1016/j.petrol.2021.109682>.
- Lyu, Q., Tan, J., Li, L., Ju, Y., Busch, A., Wood, D.A., Ranjith, P.G., Middleton, R., Shu, B., Hu, C., Wang, Z., Hu, R., 2021. The role of supercritical carbon dioxide for recovery of shale gas and sequestration in gas shale reservoirs. Energy Environ. Sci. 14, 4203–4227. <https://doi.org/10.1039/DOEE03648J>.
- Matter, J.M., Stute, M., Snæbjörnsdóttir, S.Ó., Oelkers, E.H., Gislason, S.R., Aradottir, E. S., Sigfusson, B., Gunnarsson, I., Sigurdardottir, H., Gunnlaugsson, E., Axelsson, G.,

- Alfredsson, H.A., Wolff-Boenisch, D., Mesfin, K., Taya, D.F. de la R., Hall, J., Dideriksen, K., Broecker, W.S., 2016. Rapid carbon mineralization for permanent disposal of anthropogenic carbon dioxide emissions. *Science* 352, 1312–1314. <https://doi.org/10.1126/science.aad8132>.
- Mo, S., Zhu, Y., Zabaras, N., Shi, X., Wu, J., 2019. Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. *Water Resour. Res.* 55, 703–728. <https://doi.org/10.1029/2018WR023528>.
- Mora, P., Morra, G., Yuen, D.A., Patil, S., Juanes, R., 2023. Convection-diffusion with the colour gradient lattice Boltzmann method for three-component, two-phase flow. *Transp. Porous Media* 147, 259–280. <https://doi.org/10.1007/s11242-023-01906-8>.
- Müller, S., Schüler, L., Zech, A., Heße, F., 2022. GSTools v1.3: a toolbox for geostatistical modelling in Python. *Geosci. Model Dev. (GMD)* 15, 3161–3182. <https://doi.org/10.5194/gmd-15-3161-2022>.
- Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. *Neurocomputing* 452, 48–62. <https://doi.org/10.1016/j.neucom.2021.03.091>.
- Nordbotten, J.M., Celia, M.A., 2006. Similarity solutions for fluid injection into confined aquifers. *J. Fluid Mech.* 561, 307. <https://doi.org/10.1017/S0022112006000802>.
- Oladyshkin, S., Class, H., Nowak, W., 2013. Bayesian updating via bootstrap filtering combined with data-driven polynomial chaos expansions: methodology and application to history matching for carbon dioxide storage in geological formations. *Comput. Geosci.* 17, 671–687. <https://doi.org/10.1007/s10596-013-9350-6>.
- Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks. In: International Conference on Machine Learning. Pmlr, pp. 1310–1318.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.
- Raiossi, 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. *J. Comput. Phys.* 378, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>.
- Raza, A., Glatz, G., Gholami, R., Mahmoud, M., Alafnan, S., 2022. Carbon mineralization and geological storage of CO₂ in basalt: mechanisms and technical challenges. *Earth Sci. Rev.* 229, 104036. <https://doi.org/10.1016/j.earscirev.2022.104036>.
- Rinaldi, A.P., Rutqvist, J., Luu, K., Blanco-Martín, L., Hu, M., Sentí, M.L., Eberle, L., Kaestli, P., 2022. TOUGH3-FLAC3D: a modeling approach for parallel computing of fluid flow and geomechanics. *Comput. Geosci.* 26, 1563–1580. <https://doi.org/10.1007/s10596-022-01176-0>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference. Springer International Publishing, Munich, Germany, pp. 234–241. October 5–9, 2015, Proceedings, Part III 18.
- Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* 28.
- Silva, J.A., Saló-Salgado, L., Patterson, J., Dasari, G.R., Juanes, R., 2023. Assessing the viability of CO₂ storage in offshore formations of the Gulf of Mexico at a scale relevant for climate-change mitigation. *Int. J. Greenh. Gas Control* 126, 103884. <https://doi.org/10.1016/j.ijggec.2023.103884>.
- Snaebjörnsdóttir, S.Ó., Sigfusson, B., Marieni, C., Goldberg, D., Gislason, S.R., Oelkers, E. H., 2020. Carbon dioxide storage through mineral carbonation. *Nat. Rev. Earth Environ.* 1, 90–102. <https://doi.org/10.1038/s43017-019-0011-8>.
- Stepien, M., Ferreira, C.A.S., Hosseinzadehsadati, S., Kadeethum, T., Nick, H.M., 2023. Continuous conditional generative adversarial networks for data-driven modelling of geologic CO₂ storage and plume evolution. *Gas Sci. Eng.* 115, 204982. <https://doi.org/10.1016/j.jgsce.2023.204982>.
- Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).
- Tang, H., Fu, P., Sherman, C.S., Zhang, J., Ju, X., Hamon, F., Azzolina, N.A., Burton-Kelly, M., Morris, J.P., 2021. A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage. *Int. J. Greenh. Gas Control* 112, 103488. <https://doi.org/10.1016/j.ijggec.2021.103488>.
- Tang, M., Ju, X., Durlofsky, L.J., 2022. Deep-learning-based coupled flow-geomechanics surrogate model for CO₂ sequestration. *Int. J. Greenh. Gas Control* 118, 103692. <https://doi.org/10.1016/j.ijggec.2022.103692>.
- Tang, M., Liu, Y., Durlofsky, L.J., 2020. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. *J. Comput. Phys.* 413, 109456. <https://doi.org/10.1016/j.jcp.2020.109456>.
- Tariq, Z., Yildirim, E.U., Gudala, M., Yan, B., Sun, S., Hoteit, H., 2023. Spatial-temporal prediction of minerals dissolution and precipitation using deep learning techniques: an implication to Geological Carbon Sequestration. *Fuel* 341, 127677. <https://doi.org/10.1016/j.fuel.2023.127677>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.
- Wang, K., Chen, Y., Mehana, M., Lubbers, N., Bennett, K.C., Kang, Q., Viswanathan, H.S., Germann, T.C., 2021. A physics-informed and hierarchically regularized data-driven model for predicting fluid flow through porous media. *J. Comput. Phys.* 443, 110526. <https://doi.org/10.1016/j.jcp.2021.110526>.
- Wang, N., Chang, H., Zhang, D., 2021. Efficient uncertainty quantification for dynamic subsurface flow with surrogates by Theory-guided Neural Network. *Comput. Methods Appl. Mech. Eng.* 373, 113492. <https://doi.org/10.1016/j.cma.2020.113492>.
- Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A., Benson, S.M., 2022. U-FNO—an enhanced Fourier neural operator-based deep-learning model for multiphase flow. *Adv. Water Resour.* 163, 104180. <https://doi.org/10.1016/j.advwatres.2022.104180>.
- Wen, G., Li, Z., Long, Q., Azizzadenesheli, K., Anandkumar, A., Benson, S.M., 2023. Real-time high-resolution CO₂ geological storage prediction using nested Fourier neural operators. *Energy Environ. Sci.* 16, 1732–1741. <https://doi.org/10.1039/D2EE04204E>.
- Wen, G., Tang, M., Benson, S.M., 2021. Towards a predictor for CO₂ plume migration using deep neural networks. *Int. J. Greenh. Gas Control* 105, 103223. <https://doi.org/10.1016/j.ijggec.2020.103223>.
- Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. *Proc. IEEE* 78, 1550–1560. <https://doi.org/10.1109/5.58337>.
- Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision – ECCV 2018, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
- Xu, J., Fu, Q., Li, H., 2023. A novel deep learning-based automatic search workflow for CO₂ sequestration surrogate flow models. *Fuel* 354, 129353. <https://doi.org/10.1016/j.fuel.2023.129353>.
- Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y., 2015. Show, attend and tell: neural image caption generation with visual attention. In: Proceedings of the 32nd International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 2048–2057.
- Yan, B., Chen, B., Robert Harp, D., Jia, W., Pawar, R.J., 2022a. A robust deep learning workflow to predict multiphase flow behavior during geological CO₂ sequestration injection and Post-Injection periods. *J. Hydrol.* 607, 127542. <https://doi.org/10.1016/j.jhydrol.2022.127542>.
- Yan, B., Harp, D.R., Chen, B., Pawar, R., 2022b. A physics-constrained deep learning model for simulating multiphase flow in 3D heterogeneous porous media. *Fuel* 313, 122693. <https://doi.org/10.1016/j.fuel.2021.122693>.
- Zhang, T., Zhang, Y., Katterbauer, K., Al Shehri, A., Sun, S., Hoteit, I., 2024. Deep learning-assisted phase equilibrium analysis for producing natural hydrogen. *Int. J. Hydrog. Energy* 50, 473–486. <https://doi.org/10.1016/j.ijhydene.2023.09.097>.
- Zhong, Z., Sun, A.Y., Jeong, H., 2019. Predicting CO₂ plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. *Water Resour. Res.* 55, 5830–5851. <https://doi.org/10.1029/2018WR024592>.
- Zhou, J., Hu, N., Xian, X., Zhou, L., Tang, J., Kang, Y., Wang, H., 2019. Supercritical CO₂ fracking for enhanced shale gas recovery and CO₂ sequestration: results, status and future challenges. *Adv. Geo-Energy Res.* 3, 207–224. <https://doi.org/10.26804/agcr.2019.02.10>.
- Zhu, Q., Zhang, W., Zhou, L., Liu, T., 2016. Learning to start for sequence to sequence architecture. arXiv preprint arXiv:1608.05554. doi: 10.48550/arXiv.1608.05554.