

¹ Stochastic pix2vid: A new spatiotemporal deep learning
² method for image-to-video synthesis in geologic CO₂
³ storage prediction

⁴ Misael M. Morales^{1*}, Carlos Torres-Verdín^{1,2}, and Michael J. Pyrcz^{1,2}

⁵ 1. Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin

⁶ 2. Jackson School of Geosciences, The University of Texas at Austin

⁷ *Corresponding author; email: misaelmorales@utexas.edu

⁸ **Abstract**

⁹ Numerical simulation of multiphase flow in porous media is an important step in understanding the dynamic
¹⁰ behavior of geologic CO₂ storage (GCS). Scaling up GCS requires fast and accurate high-resolution modeling
¹¹ of the storage reservoir pressure and saturation plume migration; however, such modeling is challenging due
¹² to the high computational costs of traditional physics-based simulations. Deep learning models trained with
¹³ numerical simulation data can provide a fast and reliable alternative to expensive physics-based numerical
¹⁴ simulations. We propose a Stochastic pix2vid neural network architecture for solving multiphase fluid flow
¹⁵ problems with superior speed, accuracy, and efficiency. The Stochastic pix2vid model is designed based on
¹⁶ the principles of computer vision and video synthesis and is able to generate dynamic spatiotemporal predic-
¹⁷ tions of fluid flow from static reservoir models, closely mimicking the performance of traditional numerical
¹⁸ simulation. We apply the Stochastic pix2vid model to a highly-complex CO₂-water multiphase problem with
¹⁹ a wide range of reservoir models in terms of porosity and permeability heterogeneity, facies distribution, and
²⁰ injection configurations. The Stochastic pix2vid method is first-of-its-kind in static-to-dynamic prediction
²¹ of reservoir behavior, where a single static input is mapped to its dynamic response. The Stochastic pix2vid
²² method provides superior performance in highly heterogeneous geologic formations and complex estimation
²³ such as CO₂ saturation and pressure buildup plume determination. The trained model can serve as a general-
²⁴ purpose, static-to-dynamic (image-to-video) alternative to traditional numerical reservoir simulation of 2D
²⁵ CO₂ injection problems with up to 6,500× speedup compared to traditional numerical simulation.

²⁶ **Keywords:** Image-to-video synthesis, Spatiotemporal prediction, Convolutional neural network, Recur-
²⁷ rent neural network, Proxy model

28 1 Introduction

29 Geologic CO₂ sequestration (GCS) has emerged as a potential technology solution to reduce anthropogenic
30 greenhouse gas emissions to the atmosphere [1–3], and has become increasingly popular worldwide due to
31 the need to meet international climate protection agreements [4–6]. Modeling injected CO₂ movement in
32 the subsurface over and beyond the life of the project is a critical component to support optimum GCS
33 project decision making for safe and secure CO₂ sequestration. A schematic of typical GCS operations is
34 shown in Figure 1, including storage in depleted oil and gas reservoir and deep saline formations, and CO₂
35 enhanced oil and coalbed methane recovery [7–9]. However, there are several technical challenges associated
36 with the subsurface modeling to support GCS operations. To accurately forecast and monitor subsurface
37 multiphase flow, physics-based high-fidelity numerical simulations are required. These numerical simulations
38 are computationally intensive and time-consuming since they require iterative solutions of nonlinear systems
39 of equations applied over large volumes of the subsurface at sufficient resolution to represent heterogeneity
40 [10–13]. Also, due to the large degree of uncertainty in subsurface data, and the spatial distribution of
41 the properties of heterogeneous porous media between the sparsely sampled data, GCS operations require
42 a robust probabilistic-based uncertainty assessment for improved engineering decision-making [14–16]. In
43 order to capture the fine-scale multiphase flow behavior given an uncertain spatial distribution of subsurface
44 properties, a large number of numerical simulations are required, leading to very high computational costs
45 and delayed feedback unable to support timely decision making [17, 18].

46 To overcome this, machine learning techniques have emerged as candidate proxy models due to their
47 ability to perform dimensionality reduction for efficient problem parameterization and model complicated
48 systems to calculate fast predictions of subsurface flow and transport behavior for real-time feedback on
49 the impact of geological and engineering controls on CO₂ behavior in the subsurface over time [19–21].
50 Dimensionality reduction techniques are supervised or unsupervised machine learning methods that compress
51 (or encode) the data, X , into a lower-dimensional latent feature representation, z , and decompress (or decode)
52 the latent representation either: (1) back to the original data space, \hat{X} (unsupervised, AutoEncoder), or (2)
53 to a new response feature space, y (supervised, Encoder-Decoder) [22–24], as shown in Figure 2. The recent
54 advancements in deep learning algorithms and in computing architecture and power, enable GPU-enabled
55 neural network models that have accelerated the fields of forward and inverse modeling [25, 26]. Classical
56 statistical modeling methods are often hindered by the size of the models and their conditioning to big data,
57 i.e., that is data with volume, velocity, variety, value, and veracity [27, 28], and fail to generalize beyond
58 fit-for-purpose frameworks [29, 30]. By analyzing big data sets, machine learning techniques can uncover
59 complex patterns and relationships in lower-dimensional, latent feature representations that may not be

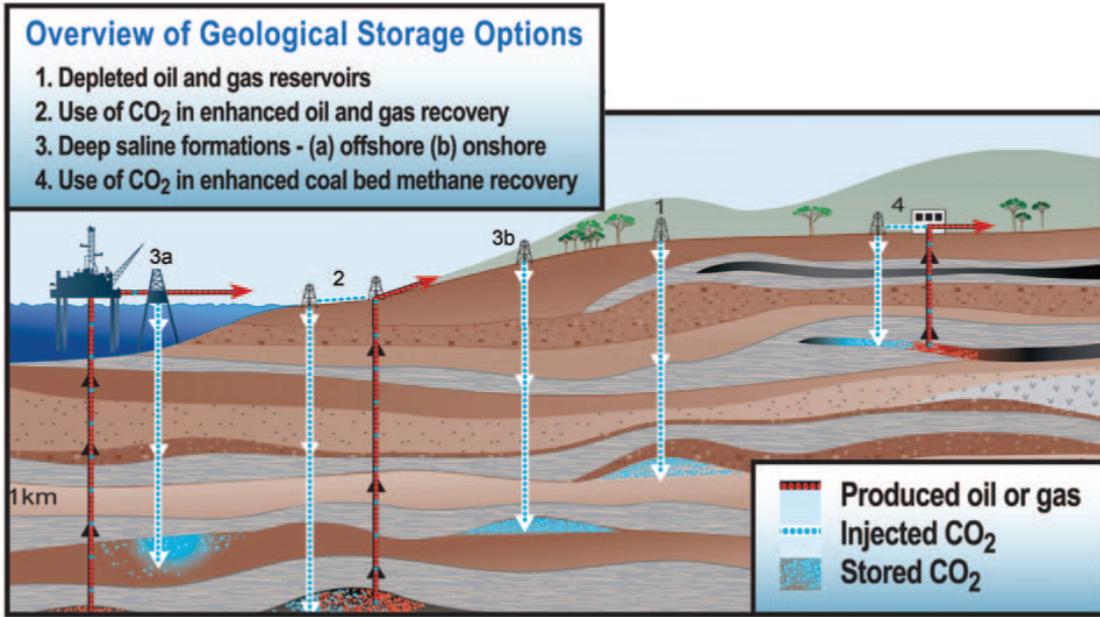


Figure 1: Types of geologic CO₂ storage operations and the geologic formations that can be used for sequestration. *Modified from the Carbon Dioxide Cooperative Research Center (CO2CRC), <http://www.co2crc.com.au/about/co2crc>*

discernible through traditional statistical and geostatistical methods [31–33]. When combined with a latent space modeling framework, machine learning approaches efficiently and accurately exploit hidden patterns and features in the data, remove redundancies or noise, and decrease the mathematical and computational complexity of the problem significantly [34, 35].

Supervised machine learning approaches applied to the subsurface are divided into two main categories, namely purely data-driven models or physics-informed models. Data-driven proxy models are neural network architectures trained with labeled data that produce a mapping from input predictor feature to output response features [36, 37]. On the other hand, the training process to match training data for PINNs is

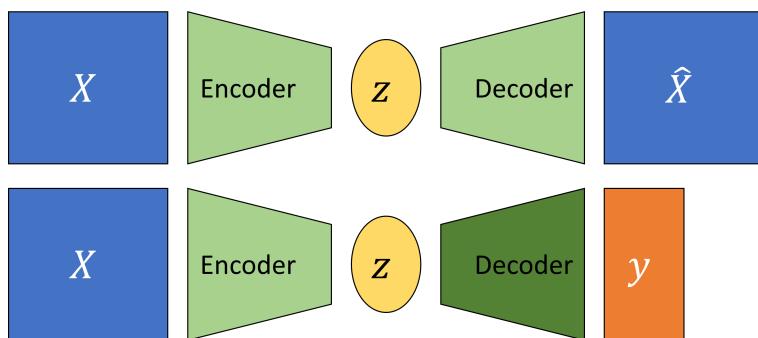


Figure 2: Dimensionality reduction model structures. Unsupervised AutoEncoder structure (top), and supervised Encoder-Decoder structure (bottom).

regularized with the minimization of the (physical) loss from the residual of the governing partial differential equations (PDEs) along with the losses associated with the initial and boundary conditions [38, 39]. However, other variants of PINNs such as physics-guided or physics-constrained neural networks where the PDE loss is not embedded in the training step, instead the models have specific architectures or parameters to mimic the physics in the system, have proven useful for subsurface energy resource engineering applications [40–42]. One disadvantage of machine learning techniques is that they require significant amounts of training data, but once trained these prediction models suffer from lack of generalization, i.e., inability to provide accurate predictions away from the training data beyond which they have been specifically trained [43, 44]. For both data-driven and physics-informed approaches, typically, spatial relationships are modeled through convolutional neural networks (CNNs) [45, 46] and the temporal relationships through recurrent neural networks (RNNs) [47, 48], but recent advancements in transformer-based architectures improve performance compared to the CNN and RNN methods for spatial and temporal latent feature representations [49–51].

A number of machine learning-based proxy models have been developed to estimate the reservoir behavior in subsurface energy resource applications. Most techniques rely on the concept of image translation, or pix2pix, where a target image(s) is predicted from an input image(s) [52–55], as shown in Figure 3. Maldonado-Cruz and Pyrcz [56] develop a convolutional U-Net model to predict pressure and saturation states given an uncertain geologic realization. This work is an example of image-to-image static forecasting, where the time state is given as an input, and the proxy model will predict a single response state of pressure and saturation at the given time. Wen et al. [57] develop a Fourier Neural Operator (FNO) architecture to predict image-to-image response states of pressure and saturation from an uncertain geologic realization and is further extended for multi-scale and nested domains [58]. These methods are based on a pix2pix, or image-to-image prediction, where a specific timestep is used as an input feature to predict the relationship between the geologic model and the reservoir response at that specific timestep. This implies that pix2pix or image-to-image methods are formulated as an even-determined or sometimes over-determined estimation problem, where the number of input features is equal to or greater than the number of output features. Moreover, numerous other proxy models have been developed for subsurface applications using more complex architectures such as generative adversarial networks (GANs) [59] and transformers [60, 61]. Despite showing consistent results and significant speedups compared to traditional numerical simulation, pix2pix models do not capture the spatiotemporal relationships and dynamic response of the subsurface system.

Moving beyond image-to-image predictions, Kim and Durlofsky [62] develop a convolutional-recurrent proxy for pix2time, or image-to-timeseries, forecasting and discuss its advantages for closed-loop reservoir management under geologic uncertainty. This method moves beyond the image-to-image forecasting and exploits a spatiotemporal latent space in an encoder-recurrent neural network architecture to obtain hy-

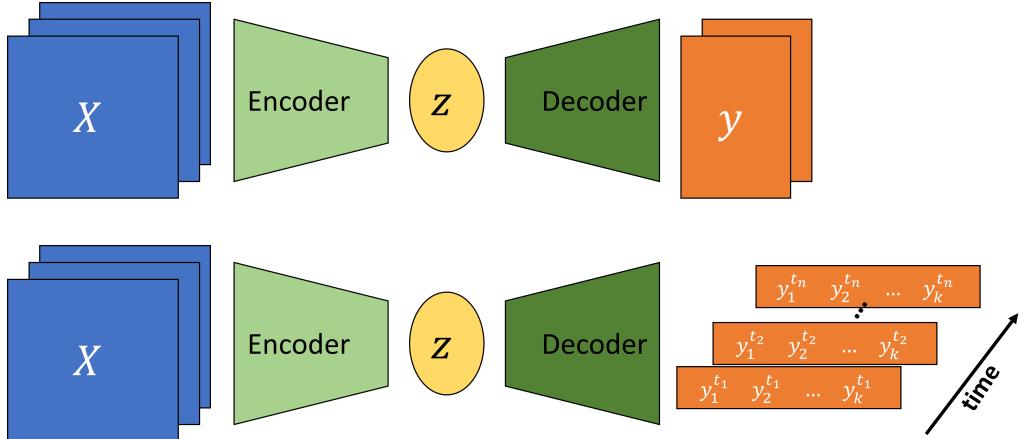


Figure 3: Image-to-image (pix2pix) (top) and image-to-timeseries (bottom) Encoder-Decoder structures.

drocarbon production forecasts. The image-to-series formulation can still be an even- or over-determined estimation problem, where we have equal or more inputs than outputs, as shown in Figure 3. Furthermore, Tang et al. [63, 64] and Jiang and Durlofsky [18] develop a recurrent residual U-net (R-U-net) proxy for the prediction of dynamic pressure- and saturation-over-time from uncertain geologic realizations using an encoder-recurrent-decoder architecture. These methods aim to obtain dynamic response states over time from a single static image. This type of proxy model is formulated to resolve the more complex underdetermined estimation problem (compared to even- or over-determined), where the number of input features is a fraction of the number of output features. However, the recurrent R-U-net proxy is limited by the fact that only the latent space receives spatiotemporal processing, while the model reconstruction is done via time-distributed deconvolutions, treating time as an additional “spatial” dimension, and not fully exploiting the spatiotemporal relations in the data and latent space as an image-to-video forecasting formulation.

The problem of image-to-video forecasting, also known as video synthesis, has been approached previously by researchers in the field of computer vision [65–69]. Iliadis et al. [70] are one of the first to develop a deep learning-based framework for video compressive sensing to reconstruct a video sequence from a single image using a deep fully-connected neural network, or artificial neural network (ANN). Despite excellent accuracy in the video predictions, this method is still limited by time-distributed fully-connected layers in the encoder and decoder portions of the network, thus not exploiting the spatiotemporal relationships in the data. Xu and Ren [71] develop a three-part encoder-recurrent-decoder network for video reconstruction from the estimated motion fields of the encoded frames. The implementation is similar to that of Jiang and Durlofsky [18] and Tang et al. [63, 64] in that it applies a recurrent update in the latent space but relies on time-distributed deconvolutions for the video frames reconstruction to exploit spatiotemporal relationships in the data. Dorkenwald et al. [72] develop a conditional invertible neural network (cINN) as a bijective mapping

123 between image and video domains using a dynamic latent representation. The cINN architecture allows for
124 video-to-image and image-to-video predictions, demonstrating possible the generation of video frames from
125 a static input image. Finally, Holynski et al. [73] implemented the idea of Eulerian motion fields to define
126 the moving portions of the image to accurately reconstruct a series of video frames from a static image using
127 a spatiotemporal latent space parameterization. These advancements in the field of computer vision and
128 video compressed sensing are the foundation for our image-to-video proxy model.

129 We propose a novel image-to-video spatiotemporal proxy model, Stochastic pix2vid, for the prediction of
130 dynamic reservoir behavior over time from a subsurface uncertainty model suite of static geologic realizations.
131 Our model exploits the spatial and temporal structures in latent space to dynamically reconstruct the time-
132 dependent pressure and multiphase saturation states from a static geologic realization. The model then
133 reconstructs the dynamic pressure and saturation distributions using a spatiotemporal decoder network
134 with convolutional long short-term memory (ConvLSTM) layers, which are concatenated with the residuals
135 of the spatial latent parameterizations from the encoder network. Thus, it is not an encoder-recurrent-
136 decoder architecture, but instead a fully spatiotemporal convolutional-recurrent image-to-video synthesis
137 model. Our stochastic pix2vid model has significant advantages compared to image-to-image and encoder-
138 recurrent-decoder models in terms of computational efficiency and prediction accuracy and can be used as
139 a replacement for physics-based numerical reservoir simulations, or high-fidelity simulations (HFS), in GCS
140 projects as an image-to-video mapping operator.

141 In the methodology section, we describe the governing equations of multiphase flow in GCS, and the
142 proposed spatiotemporal proxy model architecture. In the results and discussion sections, we describe the
143 geologic modeling and numerical reservoir simulation steps required to generate the training data, and
144 evaluate the training and performance of the proposed proxy model and compare its efficiency and accuracy
145 to high-fidelity numerical simulations using a 2D synthetic case for large-scale GCS operations.

146 2 Methodology

147 This section describes the governing equations, and the architecture and training strategy of the Stochastic
148 pix2vid model.

149 **2.1 Governing equations** For the CO₂-water multiphase flow problem, the general form of the mass
150 accumulation for component $\kappa = \text{CO}_2$ or water is given by [74]:

$$\frac{\partial M^k}{\partial t} = -\nabla \bullet F^\kappa + q^\kappa. \quad (1)$$

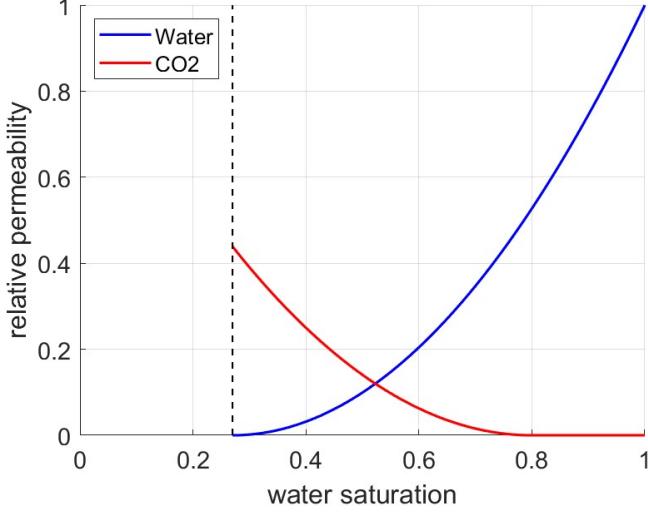


Figure 4: Relative permeability curves for the CO₂-water system. The residual saturations are 0.27 and 0.2 for water and CO₂, respectively.

151 For each component κ , the mass accumulation term M^κ is summed over all phases p ,

$$M^\kappa = \phi \sum_p S_p \rho_p X_p^\kappa \quad (2)$$

152 where ϕ is the porosity, S_p is the saturation of phase p , ρ_p is the density of phase p , and X_p^κ is the mass
153 fraction of component κ present in phase p . For each component κ , there is also the advective mass flux
154 $F^\kappa|_{adv}$ obtained by summing over all phases p ,

$$F^\kappa|_{adv} = \sum_p X_p^\kappa F_p \quad (3)$$

155 where each individual phase flux F_p is given by Darcy's equation:

$$F_p = \rho_p u_p = -k \frac{k_{r,p} \rho_p}{\mu_p} (\nabla P_p - \rho_p g) \quad (4)$$

156 where u_p is the Darcy velocity of phase p , k is the absolute permeability, $k_{r,p}$ is the relative permeability
157 of phase p , μ_p is the viscosity of phase p , and g is the gravitational acceleration constant. The relative
158 permeability curves for the CO₂-water system are shown in Figure 4. The fluid pressure of phase p ,

$$P_p = P + P_c \quad (5)$$

159 is given by the sum of the reference phase pressure P and the capillary pressure P_c . The numerical
160 simulation does not include molecular diffusion or hydrodynamic dispersion effects for practical purposes.

161 **2.2 Proxy Model Architecture**

162 Our proposed Stochastic pix2vid image-to-video data-driven method, is mapping operator from the static
163 realizations of geologic distributions of porosity, permeability and facies as well as the injector well(s) distri-
164 bution, to the dynamic responses of pressure and saturation distributions over time.

165 Let m represent a geologic model realization of porosity, permeability, facies, and injector well(s) distri-
166 butions, such that $m = \{\phi, k, \text{facies}, w\}$. The dynamic responses of pressure and saturation over time are
167 given by $d = f(m)$, such that $d = \{P(t), S(t)\}$ and f is the physics-based numerical reservoir simulation. Our
168 aim is to replace f with a more efficient data-driven proxy by training the Stochastic pix2vid model, which is
169 trained as a single model to predict both pressure and saturation distributions over time as a multi-channel
170 output from the multi-channel input features, m . For this purpose, we exploit the latent structures in space
171 and time of the static inputs and dynamic outputs through a spatiotemporal encoder-decoder architecture.

172 The encoder portion of the network is comprised of sequential convolutional layers to compress the spatial
173 features of the subsurface realizations into a latent parameterization z_m , given by $z_m = Enc(m)$. In their
174 compressed representation, these features represent the salient characteristics of the geologic distributions.
175 The decoder portion of the network is designed as a series of recursive residual convolutional-recurrent
176 layers, such that the latent space z_m is recursively decoded into the dynamic distributions of pressure and
177 saturation over time. The previous timestep latent representations, z_d^t , are used in the subsequent timesteps
178 of the decoder, such that the subsequent timesteps will predict the current and previous timestep(s) jointly
179 and iteratively, providing a reduction of systematic error in time as subsequent frames of the dynamic output
180 video are predicted. The full architecture is represented as:

$$\hat{d} = Dec^t([Enc(m); z_d^t]) \quad (6)$$

181 The encoder, $Enc(\cdot)$, compresses the geologic realizations, m , into a latent representation z_m through
182 the use of depthwise separable convolutions [75]. This type of convolution learns the parameters for each
183 channel in the image separately, avoiding mixing of variables or loss of resolution, as shown in Figure
184 5. This is especially important when dealing with discrete, non-smooth porosity and permeability spatial
185 distributions due to discrete facies and binary well(s) location distributions. Each separable convolution
186 layer is regularized with an l_1 -norm weight of 1×10^{-6} . Moreover, we use a Squeeze-and-Excite layer
187 to improve channel interdependence, and to avoid mixing and loss of resolution [76]. Each Squeeze-and-
188 Excite layer will provide the optimal network weights for each channel independent of the other channels by
189 passing the feature maps through a global pooling layer (squeeze) and a dense layer with nonlinear activation
190 (excite), to add content aware mechanism for re-weighting each channel adaptively, as shown in Figure 6.

191 Furthermore, by applying instance normalization, as opposed to the more common batch normalization, we
192 achieve channel-independent normalization of the convolved features [77]. Instance normalization is a special
193 case of group normalization, where the numbers of channels per group is exactly 1, such that each channels
194 gets its own normalization scheme, as shown in Figure 7. Parametric rectified linear units (PReLU) is used
195 as the activation function, where at each minibatch iteration, the network learns the optimal leaky slope for
196 activation in each layer, as shown in Figure 8. Finally, pooling and spatial dropout are applied to reduce in
197 half the input dimension of each layer and to provide a means of spatial regularization, respectively. Through
198 3 convolutional encoding layers with filter size 3×3 , we obtain the latent parameterizations z_m^1 , z_m^2 , and z_m^3 .
199 Table 1 summarizes the architecture and dimensions of each layer.

200 Step 1: **Depthwise separable encoding:** The first layer of Enc takes the geologic model realization, m ,
201 and computes the depthwise separable convolutional features channel-by-channel.

202 Step 2: **Squeeze-and-Excite encoding:** By taking the channel-wise global average of the feature space
203 from Step 1, a fully-connected predicts the appropriate weighting coefficients to best parameterize
204 the features.

205 Step 3: **Instance Normalization of the feature space:** Feature normalization is applied on a channel-
206 by-channels basis for each batch of the encoded feature space, avoiding mixing and blurring.

207 Step 4: **Activation, Pooling, and Spatial Dropout:** The PReLU nonlinear activation function is used,
208 and for each batch, an optimal leaky slope is learned. Pooling is used to reduce the feature space in
209 half, and Spatial Dropout of 5% is used to regularize the learning process and increase robustness
210 in prediction.

211 Step 5: **Final Encoding and Repeat:** From Steps 1-4, the geologic model realization m is encoded into
212 a latent representation z_m^k . We repeat Steps 1-4 three times to obtain three intermediate latent
213 representations, namely z_m^1 , z_m^2 , and z_m^3 .

214 The decoder, $Dec^t(\cdot)$, of the Stochastic pix2vid model extracts the spatiotemporal relationships from the
215 latent representations of m to reconstruct the dynamic pressure and saturation distributions over time, d .
216 To accurately reconstruct the spatiotemporal structure from the static latent space, z_m , we employ a series
217 of convolutional-recurrent layers, namely a convolutional long-short term memory layer (ConvLSTM). The
218 general form of a 2D ConvLSTM layer is shown in Figure 9. Through 3 convolutional-recurrent layers, we
219 obtain the dynamic prediction of \hat{d} as follows:

220 Step 6: **Spatiotemporal decoding of z_m^3 :** The first ConvLSTM layer takes the smallest latent represen-
221 tation, z_m^3 , and reconstructs the first decoded timestep z_d^3 .

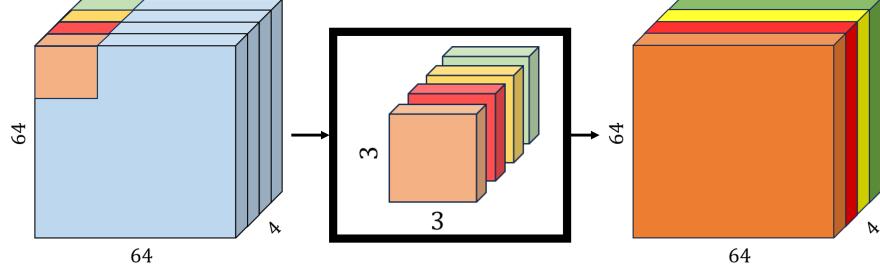


Figure 5: Schematic for a separable convolutional layer. Each channel is convolved with its own set of convolutional filters to obtain the best representation, as opposed to traditional convolutions where the same filter is applied to all channels in the data.

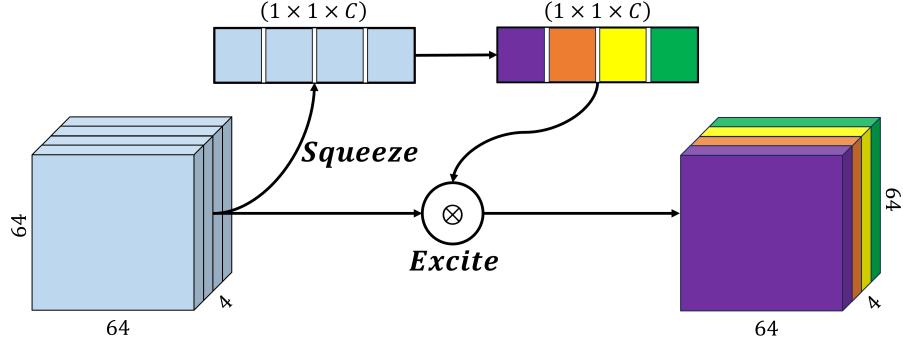


Figure 6: Schematic for a squeeze-and-excite layer. The "squeeze" layer takes the global average of the data for each channel, and the "excite" layer is a fully-connected layer with nonlinear activation to estimate the optimal weights for each channel in the data. The result is a weighted representation of the data based on their intrinsic global characteristics.

- 222 Step 7: **Residual concatenation of z_m^2 :** The first decoded timestep, z_d^3 , is concatenated with the inter-
223 mediate static encoding z_m^2 to retain multi-scale features and improve prediction performance and
224 resolution.
- 225 Step 8: **Intermediate spatiotemporal decoding:** The second ConvLSTM layer takes the residual con-
226 catenation of the intermediate latent representations, $[z_m^2, z_d^3]$, to predict the next spatiotemporal
227 representation z_d^2 .
- 228 Step 9: **Residual concatenation of z_m^1 :** The intermediate decoded timestep, z_d^2 , is concatenated with the
229 largest static encoding z_m^1 .
- 230 Step 10: **Final spatiotemporal decoding:** The third ConvLSTM layer takes the residual concatenation of
231 the larger latent representations, $[z_m^1, z_d^2]$, to predict the full-scale dynamic output, d .
- 232 To enhance the performance of the spatiotemporal decoding, each ConvLSTM layer is followed by a batch
233 normalization, activation, and a transpose convolutional layer, the latter for downscaling the latent features

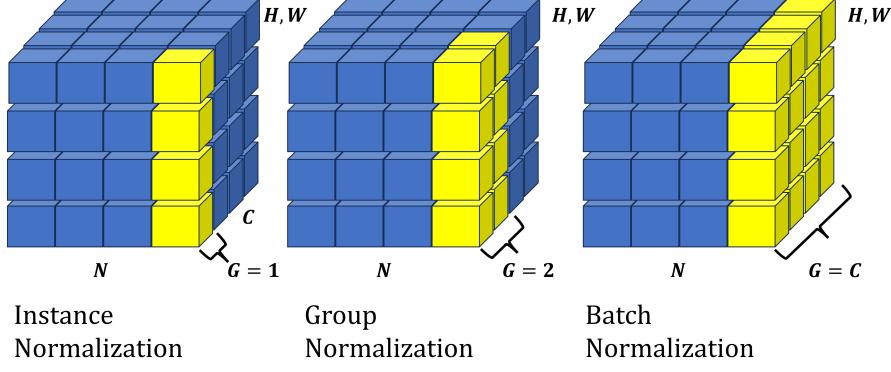


Figure 7: Schematic for instance normalization (left) compared to group normalization (center) and batch normalization (right). In an instance normalization layer, each channel will be normalized by themselves rather than normalizing the entire batch or a subset of channels (groups).

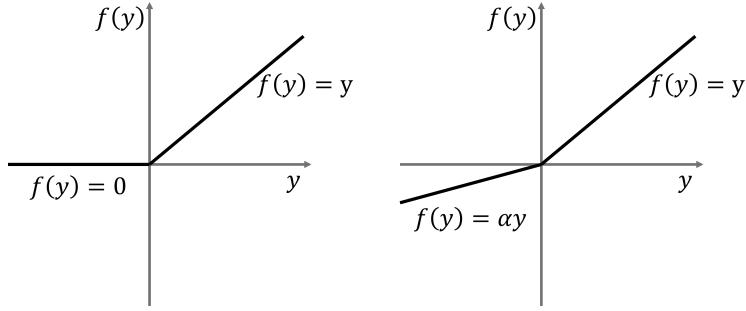


Figure 8: Schematic for the Parametric Rectified Linear Unit (PReLU) activation function (right) compared to the traditional ReLU activation function (left). The slope of the negative portion of the data, α , is learned for each batch.

234 to twice their dimension. Spatial dropout is applied, and the concatenated features are once more convolved
 235 and activated to obtain the layer prediction. Table 2 shows the architecture of the decoder network.

236 This process yields the first video frame prediction, d_1 , from the latent representation of the geologic
 237 realizations z_m . Each subsequent video frame prediction is obtained by another set of residual concatenation
 238 of the previous timestep dynamic decoded representation. The static latent representation z_m is concatenated
 239 at each timestep with the previous dynamic decoded representation for each layer such that we have $[z_m, z_{d_t}^i]$,
 240 where i is the decoding layer number and t is the timestep. By recursively implementing spatiotemporal
 241 decoding to the latent representation z_m , we obtain the prediction of the dynamic response d_t at times for
 242 each timestep $t = 1, \dots, n$.

243 The complete Stochastic pix2vid architecture is shown in Figure 10. Here we observe the spatial com-
 244 pression of the geologic models, m , through the encoding portion of the network, and the spatiotemporal
 245 decoding and residual multi-scale concatenations through the decoder portion of the network. The result-
 246 ing architecture provides proxy model from a subsurface static uncertainty model (images) to subsurface

Table 1: Encoder network architecture

Layer Number	Architecture	Shape in (h,w,c)	Shape out (h,w,c)
1	SeparableConv2D	$64 \times 64 \times 4 (m)$	
	Squeeze-and-Excite		
	Instance Norm		
	PReLU + Pooling		
	Spatial Dropout		$32 \times 32 \times 64 (z_m^1)$
2	SeparableConv2D	$32 \times 32 \times 64$	
	Squeeze-and-Excite		
	Instance Norm		
	PReLU + Pooling		
	Spatial Dropout		$16 \times 16 \times 128 (z_m^2)$
3	SeparableConv2D	$16 \times 16 \times 128$	
	Squeeze-and-Excite		
	Instance Norm		
	PReLU + Pooling		
	Spatial Dropout		$8 \times 8 \times 256 (z_m^3)$

247 dynamic response (videos).

248 2.3 Training Strategy

249 The inputs to the Stochastic pix2vid are the geologic realizations, comprised of the distributions of
 250 porosity, permeability, facies, and injection well(s) location, represented as a matrix m of dimensions $64 \times$
 251 64×4 . The outputs are the results from the numerical reservoir simulation, namely pressure and saturation
 252 distributions over time, represented as a matrix d of dimensions $64 \times 64 \times 60 \times 2$. This yields an ill-posed
 253 and under-determined estimation problem, which are difficult to resolve [78, 79]. To improve the training
 254 efficiency and performance, we subsample in time from 60 timesteps to 11. In other words, instead of monthly
 255 monitoring, we predict the dynamic outputs at the initial step and every 6 months afterward; therefore the
 256 output matrices, (d, \hat{d}) , have a final dimension of $64 \times 64 \times 11 \times 2$. This is done to make the problem
 257 more tractable and speed up the training and prediction process, while retaining majority of the temporal
 258 information.

259 We also perform min-max normalization so that the input and output features are in the range of $[0, 1]$,
 260 which greatly improves the performance of the nonlinear activation functions. Furthermore, we perform
 261 data augmentation by 90° image rotation, making the network agnostic to orientation and encourage ef-
 262 fectively learning the flow physics in the system rather than memorizing spatial distribution patterns. The
 263 total amount of training data is therefore 2,000 realizations (after augmentation), which is split into 1,500
 264 realizations for training and 500 realizations for testing. To improve model generalizability, at each epoch,
 265 each training set minibatch is further split into a training and validation subset using an 80/20 split. The
 266 validation set is only used to adjust the trainable model parameters for each batch at each epoch and is ran-

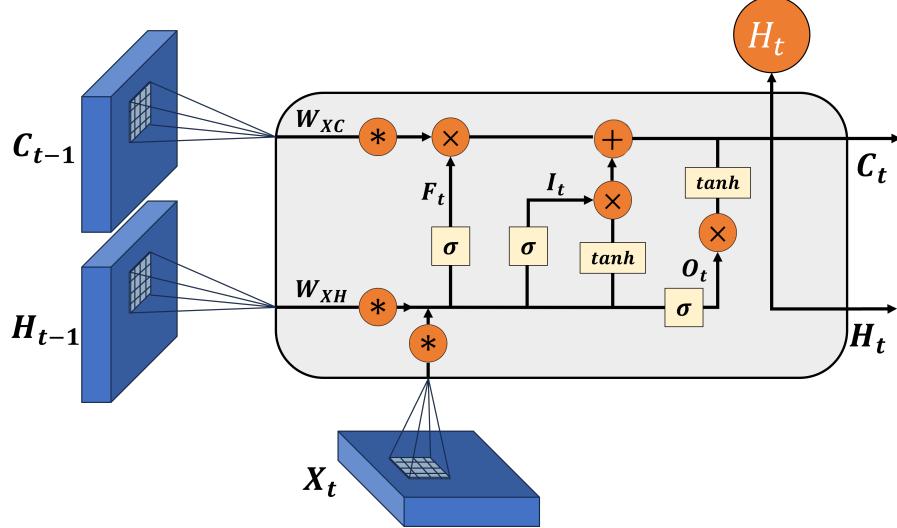


Figure 9: Schematic of a convolutional-LSTM (ConvLSTM) layer. The layer applies convolutional operations to the input data using a set of learnable filters to capture the spatial patterns. The recurrent part is a long short-term memory layer with memory and forget gates to capture the temporal patterns. LSTM units are applied to each spatial location separately allowing to capture both spatial and temporal dependencies in the data.

267 domly partitioned from the training batch at every epoch, while the testing data remains unseen to quantify
 268 the model performance after training.

269 A custom three-part loss function is used to accurately predict pixel-wise and perceptual information in
 270 the predictions. The mean squared error (MSE) is used to reconstruct the pixel-wise intensity values, while
 271 the mean absolute error (MAE) is used to optimize for the pressure and saturation plume edges. The third
 272 part is the structural similarity index metric (SSIM), which provides a perceptual image-to-image comparison
 273 of luminance, contrast, and structure [80]. For optimal training, the aim is to minimize the MSE and MAE
 274 while maximizing the SSIM for the true versus predicted outputs, d and \hat{d} , such that the total loss is given
 275 by $\mathcal{L} = \alpha(1 - SSIM) + (1 - \alpha)[\beta MSE + (1 - \beta) MAE]$, where α and β are weighting coefficients obtained
 276 empirically as 0.33 and 0.66, respectively.

277 The model is trained using the AdamW optimizer [81]. This variant of the well-known adaptive mo-
 278 mentum (Adam) optimizer [82] includes an added method to decay weights for the adaptive estimation of
 279 first-order and second-order moments. We implement a learning rate of 1×10^{-3} with a weight decay term
 280 of 1×10^{-5} .

Table 2: Decoder network architecture

Layer Number	Architecture	Shape in (t,h,w,c)	Shape out (t,h,w,c)
1	ConvLSTM2D	$1 \times 8 \times 8 \times 256$	
	BatchNorm + LeakyReLU		
	Conv2DTranspose		
	Spatial Dropout		
	Concatenate (z_m^3)		
2	Conv2D + Sigmoid		$t \times 16 \times 16 \times 128 (z_{d_t}^3)$
	ConvLSTM2D	$t \times 16 \times 16 \times 128$	
	BatchNorm + LeakyReLU		
	Conv2DTranspose		
	Spatial Dropout		
3	Concatenate (z_m^2)		
	Conv2D + Sigmoid		$t \times 32 \times 32 \times 64 (z_{d_t}^2)$
	ConvLSTM2D	$t \times 32 \times 32 \times 64$	
	BatchNorm + LeakyReLU		
	Conv2DTranspose		
	Spatial Dropout		
	Concatenate (z_m^1)		
	Conv2D + Sigmoid		$t \times 64 \times 64 \times 2 (z_{d_t}^1)$

281 3 Results

282 This section describes the geologic model generation, training performance and discusses the application of
 283 the Stochastic pix2vid proxy to rapidly forecast CO₂ plume migration for a large-scale GCS operation.

284 3.1 Reservoir Model and Simulation

285 We use SGEMS [83] to construct the subsurface uncertainty model, an ensemble of static feature realiations
 286 that is representative of various potential geologic scenarios for CO₂ storage. Using sequential Gaussian co-
 287 simulation [84], we generate a set of 1,000 random porosity (ϕ) and permeability (k) distributions with a
 288 wide range of values, as shown in Figure 11. Facies distributions are obtained from a library of deepwater
 289 fluvial training images [85, 86]. These encompass a wide range of possible geologic scenarios including
 290 marked point (lobe, ellipse, and bar), FluvSim (channel, channel-levee, and channel-levee-splay), surface
 291 based (compensational cycles of lobes), and bank retreat (channel complex). To generate consistent porosity
 292 and permeability distributions with the facies-based geologic scenarios, we condition the original porosity
 293 and permeability distributions to the facies distributions. The resulting fluvial distributions are shown in
 294 Figure 12.

295 The model has dimensions of 1km-1km-100m in the x-, y-, and z-directions, respectively. We use 64
 296 uniform grid cells in the x- and y-directions. The grid design is sufficiently refined to resolve the pressure
 297 and saturation plumes in highly heterogeneous reservoirs while remaining computationally tractable for the

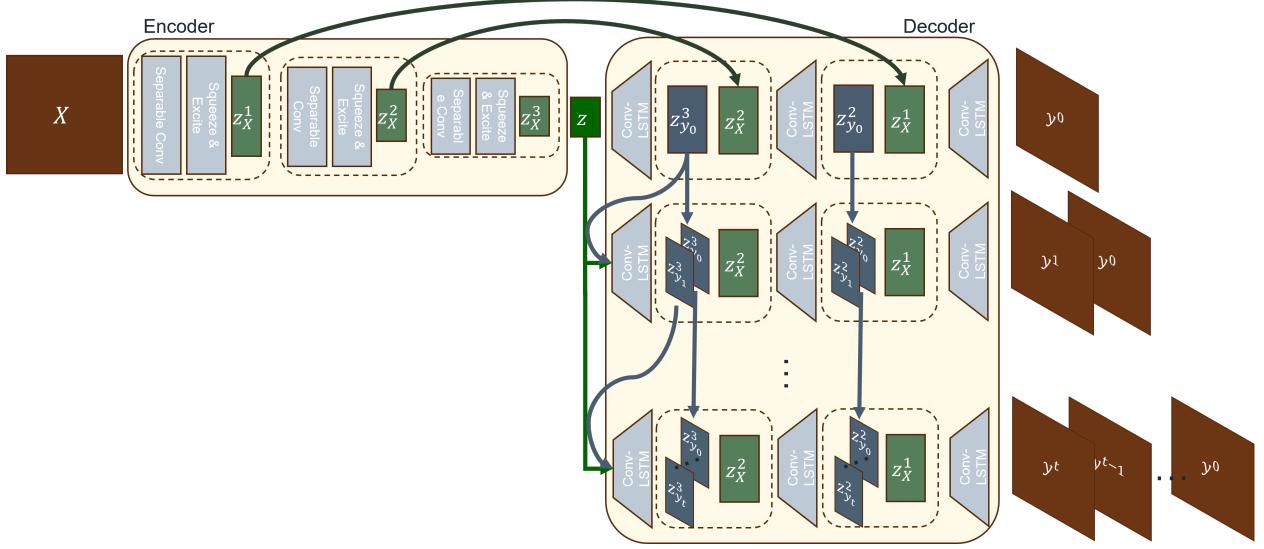


Figure 10: Architecture of our proposed Stochastic pix2vid method. The input data, $X \equiv m$, is encoded through a series of convolutional layers to capture the spatial dependencies in the geologic models. The latent representation, z_m , is recursively passed through a spatiotemporal decoder with convolutional-recurrent layers, and concatenated with the residuals of the encoder to reconstruct iteratively the frames of the output (video) data, $y \equiv d$.

purpose of training deep learning models. A random number of injection wells, $w \in [1, 3]$, are placed randomly along the reservoir for each of the 1,000 realizations, no closer than 250m from the boundaries, as shown in Figure 13. The injection well(s) are randomly placed and not conditioned to zones of preferential porosity, permeability, nor facies. Each injection well has a constant radius of 0.1m and a single and continuous perforation that injects pure supercritical CO₂ at a constant rate such that the total injection rate of the w well(s) is 0.5 megatons per year.

The conditional fluvial porosity and permeability distributions are used as input models for the numerical simulation of geologic CO₂ storage using MRST [87] to calculate the response models for training our proposed model. The reservoir is initialized as a fully water saturated zone (i.e., aquifer) with an initial pressure of 4,000 psi. The reservoir has constant isothermal conditions and constant pressure boundary conditions to represents a large-scale geologic CO₂ storage project with negligible dip, such as found in the Illinois Basin and parts of the North Sea and Gulf of Mexico.

The numerical simulation is run for 5 years, monitored monthly, for a total of 60 timesteps. At each grid cell and for each time step, we resolve the implicit pressure, explicit saturation (IMPES) formulation of Eq. (1) to obtain the corresponding dynamic pressure and saturation distributions over time (videos) from the static geologic realizations of porosity and permeability conditioned to the fluvial facies (images) with random well(s) configuration. The pressure and saturation responses corresponding to the geologic model realizations are shown in Figures 14 and 15, respectively.

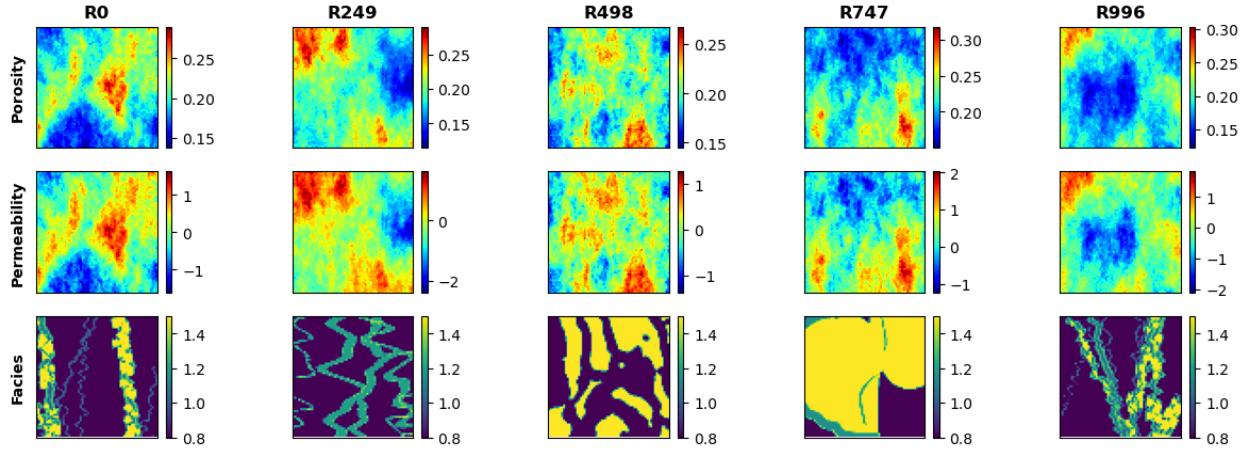


Figure 11: Spatial distribution of porosity (top), permeability (middle), and facies (bottom) for 5 random realizations.

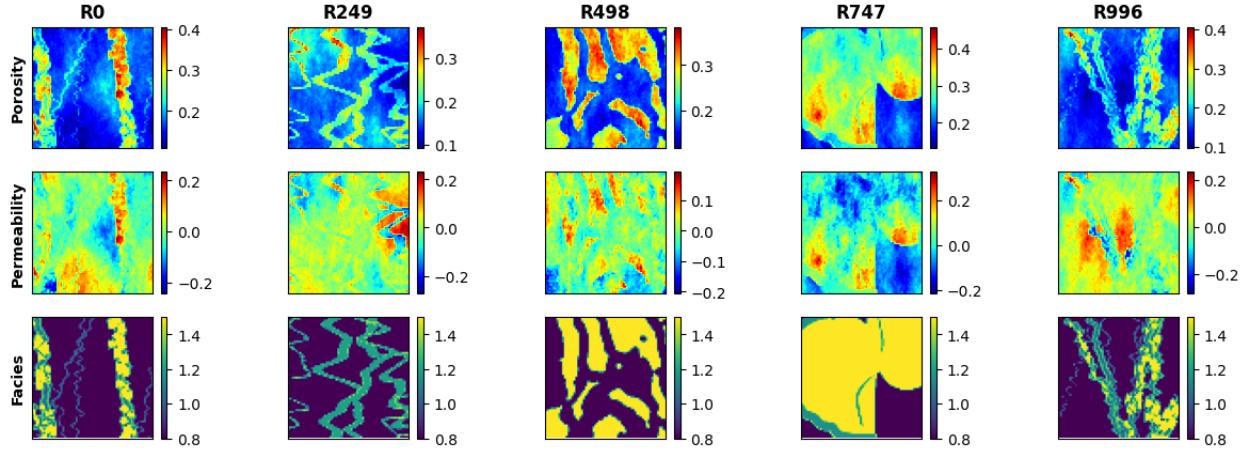


Figure 12: Spatial distribution conditioned to facies (top) for porosity (middle) and permeability (bottom) for 5 random realizations.

316 3.2 Training Performance

317 Using an NVIDIA Quadro M6000 GPU, we train for 100 epochs with batch size of 50. The model has
 318 in total 97,523,370 parameters, and the training time required is approximately 88 minutes for all 1,500
 319 training realizations. The training and validation performance per epoch is shown in Figure 16. We observe
 320 minimal overfit in the validation set, corresponding to good model generalizability and prediction accuracy
 321 within the training data. Using physics-based numerical simulation, each realization requires approximately
 322 30 seconds to obtain the dynamic pressure and saturation predictions from the static geologic models. Our
 323 Stochastic pix2vid model obtains the same results in approximately 4.59 milliseconds, corresponding to a
 324 6,500 \times speedup. The average MSE for the ensemble is 9.21×10^{-4} and 9.70×10^{-4} for training and testing,
 325 respectively. Similarly, the average SSIM for the ensemble is 98.97% and 97.91% for training and testing,

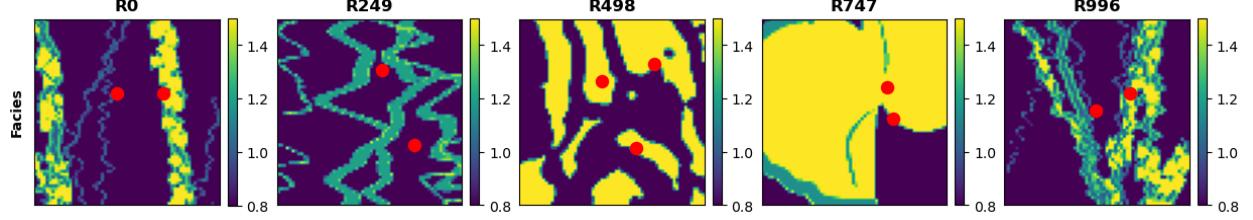


Figure 13: CO₂ injection well(s) location (red) overlaid over facies distributions for 5 random realizations.

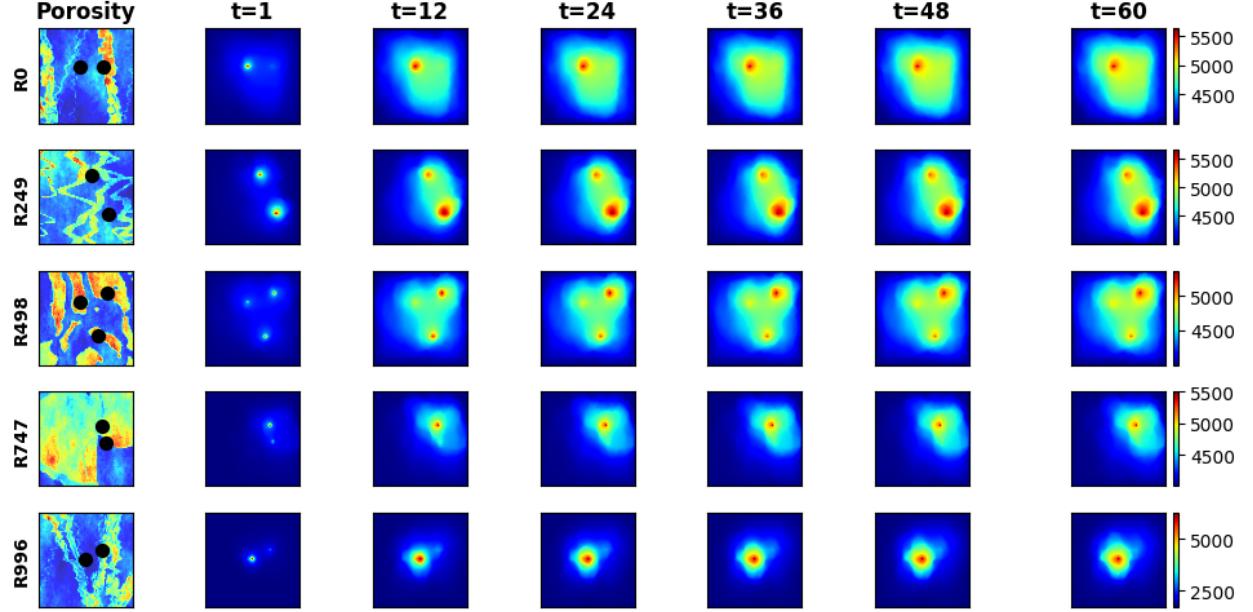


Figure 14: Pressure response distributions over time (in psia) obtained by HFS for the 5 random realizations from Fig. 12.

326 respectively.

3.3 Prediction Results

327 After training the Stochastic pix2vid model with 1,500 realizations of static geologic models, $m =$
 329 $\{\phi, k, \text{facies}, w\}$, to predict the dynamic reservoir response, $d = \{P(t), S(t)\}$, we can compare the per-
 330 formance of the predictions for the training and unseen testing data.

331 Figures 17 and 18 show the predicted dynamic pressure and saturation distributions, respectively, along
 332 with the absolute difference to HFS for 3 training realizations. We observe reasonable agreement between
 333 the true and predicted CO₂ pressure and saturation plumes over time, pixel-wise with an average MSE of
 334 3.25×10^{-4} and perceptually with SSIM of 98.59% for pressure predictions and MSE of 1.50×10^{-4} and
 335 SSIM of 97.31% for saturation predictions.

336 Similarly, Figures 19 and 20 show the pressure and saturation distributions predictions along with the
 337 absolute difference to HFS for 3 testing realizations. We observe a similar performance, with an average MSE

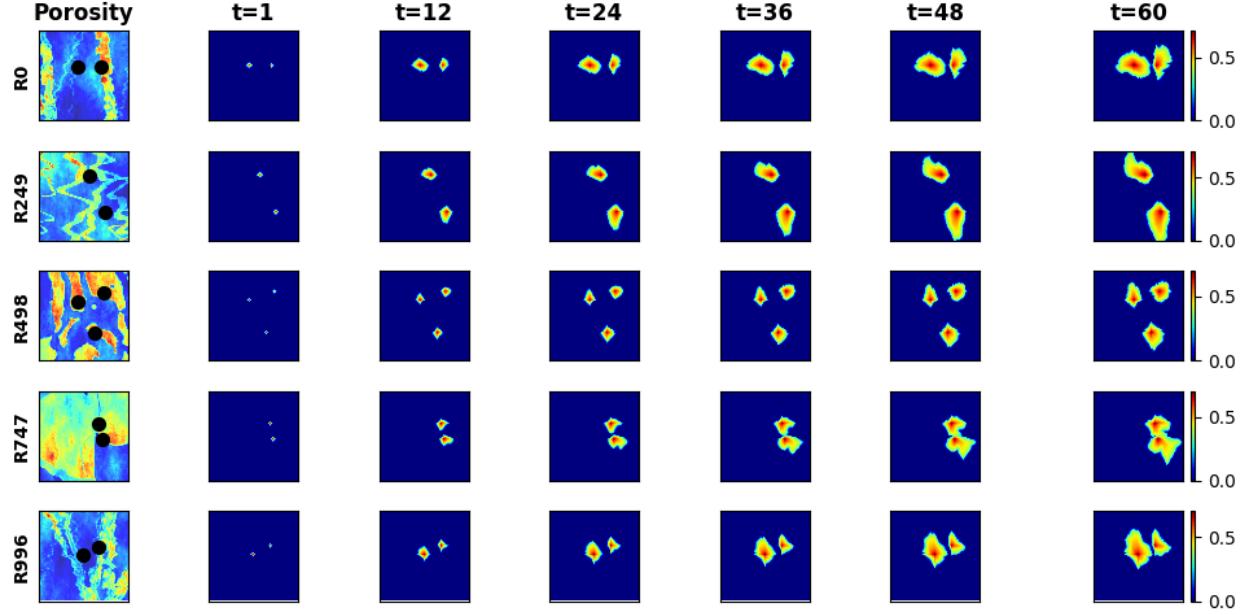


Figure 15: Saturation response distributions over time obtained by HFS for the 5 random realizations obtained from Fig. 12.

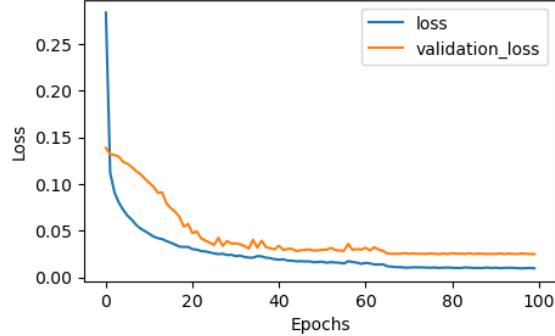


Figure 16: The total training and validation losses, \mathcal{L} , as a function of epoch number.

338 of 3.71×10^{-4} and SSIM of 97.55% for pressure predictions and MSE of 1.61×10^{-3} and SSIM of 96.19%
339 for saturation predictions. This indicates that the Stochastic pix2vid model is generalizable and achieves on
340 par performance with HFS at a fraction of the computational cost.

341 It is interesting to note that the Stochastic pix2vid model is trained on a triple-loss function with MSE,
342 MAE and SSIM. For both training and testing cases, we see that the average MSE for pressure is higher
343 than that of saturation, while the opposite is true for the average SSIM. This can be attributed to the fact
344 that there are more pixel-wise variations in pressure predictions, thus the loss focuses on matching those
345 individual pixel-wise values. On the other hand, for saturation predictions, the contrast, luminance, and
346 structure play a bigger role in the prediction than the pixel-wise intensity values. Therefore, it is important

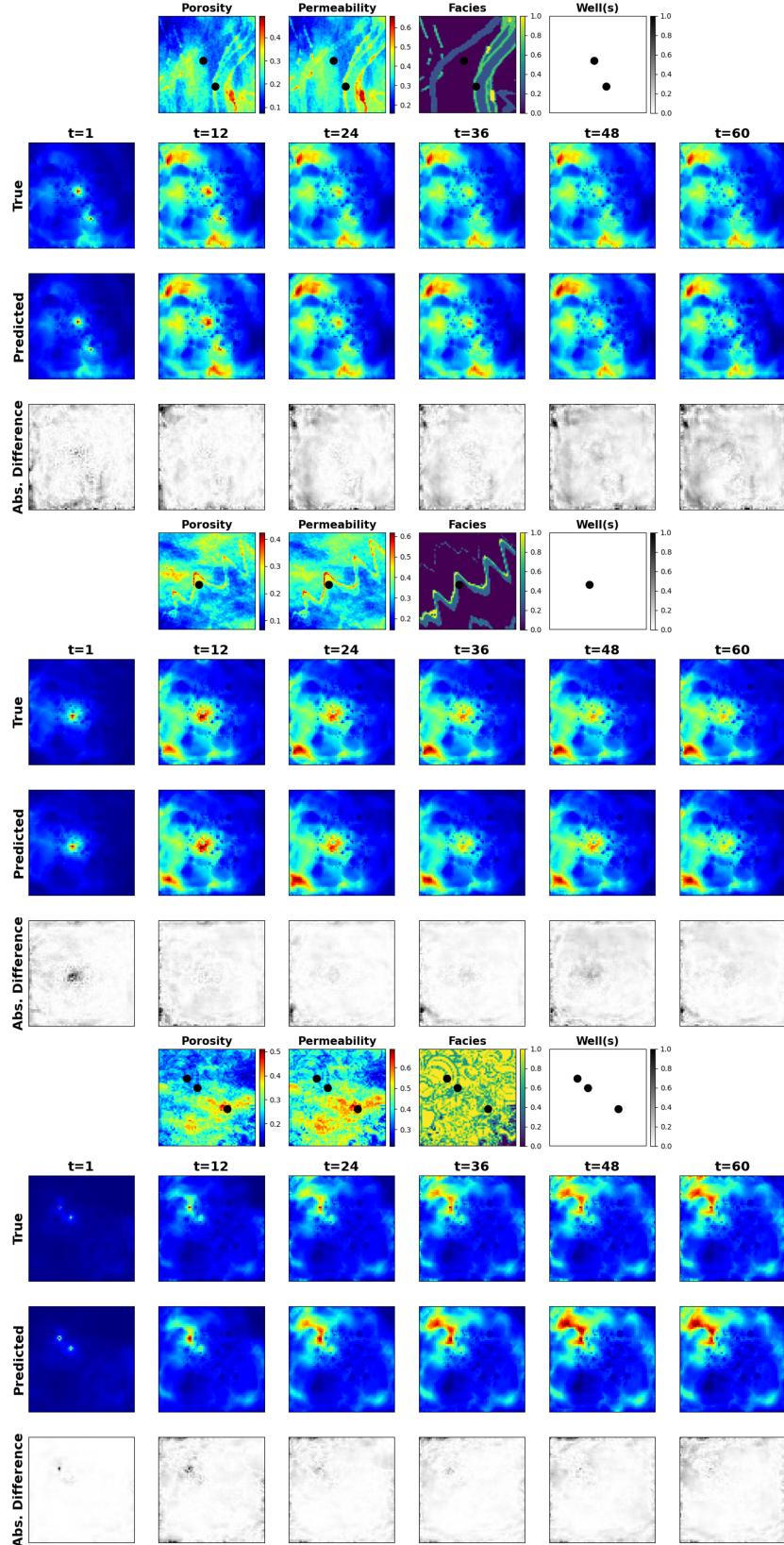


Figure 17: Normalized pressure distribution over time for 3 random training realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.

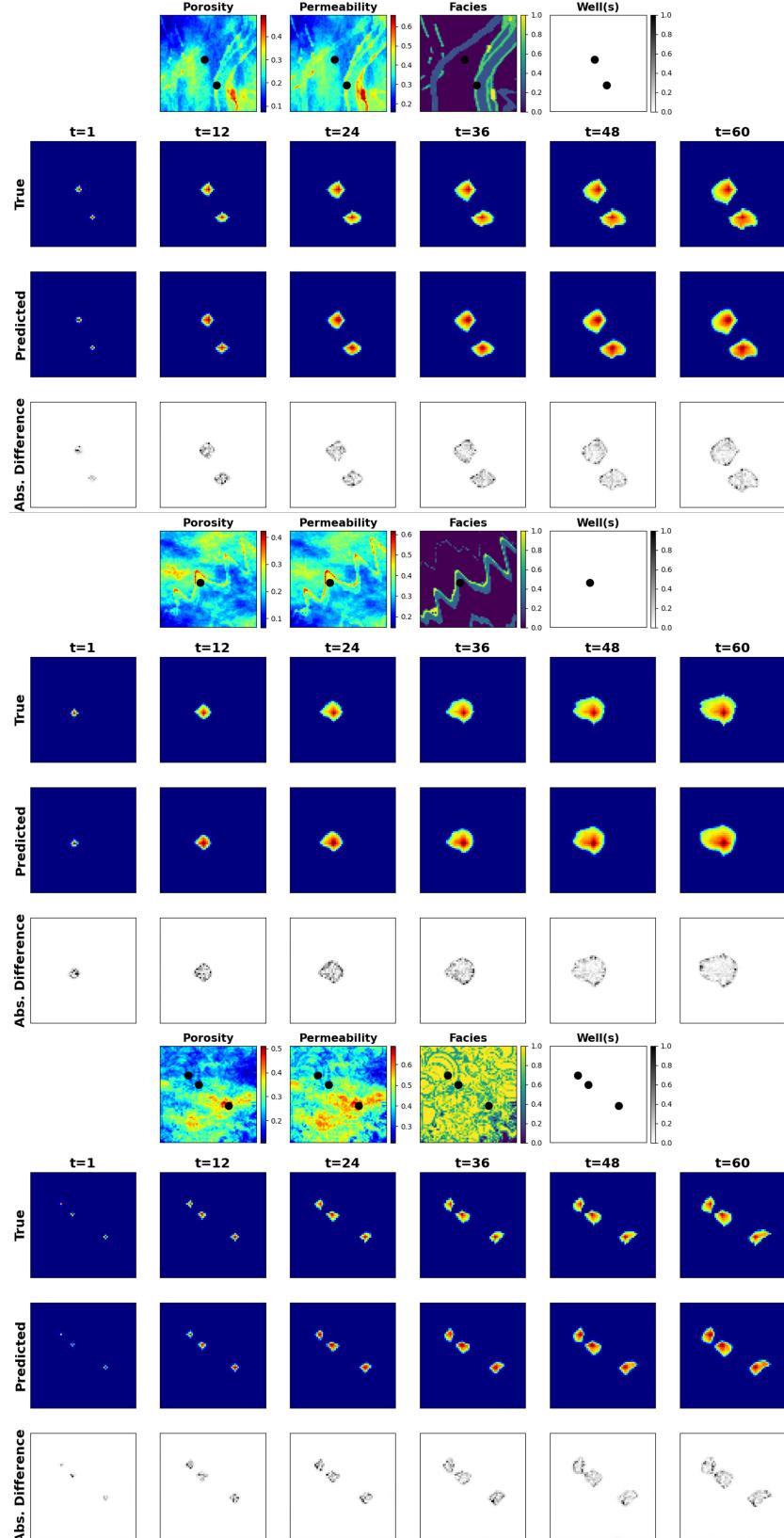


Figure 18: Saturation distribution over time for 3 random training realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.

347 to take into account both metrics for training and validating spatiotemporal subsurface prediction models.

348 From Section 2.2, the first step of the Stochastic pix2vid model is to take the static geologic realizations,
349 m , and compresses them into a latent space representation, z_m , using the spatial encoder structure. Figure
350 21 show a random selection of latent feature maps, along with their superposition on the porosity and facies
351 distribution. This can be interpreted as an analog to the attention head mechanisms recently developed
352 in transformer-based architectures [88]. We observe that the latent feature maps are essentially learning
353 the injection location(s) and direction of flow based on the geologic distributions. Thus, proving that the
354 Stochastic pix2vid model is learning multiphase flow physics and dynamic reservoir behavior appropriately.

355 These results imply that our Stochastic pix2vid is capable of learning the spatiotemporal relationship be-
356 tween the static geologic models and the dynamic reservoir response. Thus, our image-to-video architecture
357 can outperform current image-to-image and encoder-recurrent-decoder architectures to provide improved
358 reservoir behavior prediction closer to that of traditional numerical simulation. To quantify the uncertainty
359 in predictions, a comparison of true (d) versus predicted (\hat{d}) response for pressure and saturation distribu-
360 tions for the testing data is shown in Figure 22. The average R^2 over time is approximately 99% with narrow
361 95% prediction bands that recursively narrow over time. From Figure 22 we observe the advantage in imple-
362 menting recursive refining of predictions over time with recurrent residual connections in the spatiotemporal
363 decoder network, thus reducing the spatiotemporal uncertainty in the predictions.

364 CO₂ saturation and pressure buildup fronts are important quantities for geologic CO₂ storage projects
365 and are often used for regulatory oversight [89, 90], monitoring metrics or history matching purposes [91, 92].
366 The distance between the injection well(s) and the saturation fronts represents the maximum extent of the
367 CO₂ plume; however, these are often very difficult to capture accurately with data-driven proxy models.
368 Our Stochastic pix2vid method shows greater absolute error on and around the plume fronts compared to
369 within the plumes. However, the overall shape and intensity of the pressure and saturation distributions over
370 time is very well captured for all realizations despite being highly heterogeneous. Therefore, the Stochastic
371 pix2vid model can be used as a reliable replacement for expensive numerical reservoir simulations, especially
372 in cases where large number of runs are required to obtain dynamic estimates (e.g., well placement and
373 control optimization, history matching, uncertainty quantification).

374 3.4 Discussion

375 In our Stochastic pix2vid model, the encoder block is composed of separable convolutions, squeeze and
376 excite layers, and instance normalization. These three particular implementations allow for precise param-
377 eterization of the geologic realization into a latent representation, without mixing the effects of Gaussian-
378 distributed properties against binary or binomial-distributed properties. Using recursive residual ConvLSTM
379 layers, the decoder block iteratively predicts each dynamic state, or video frame, from the concatenation

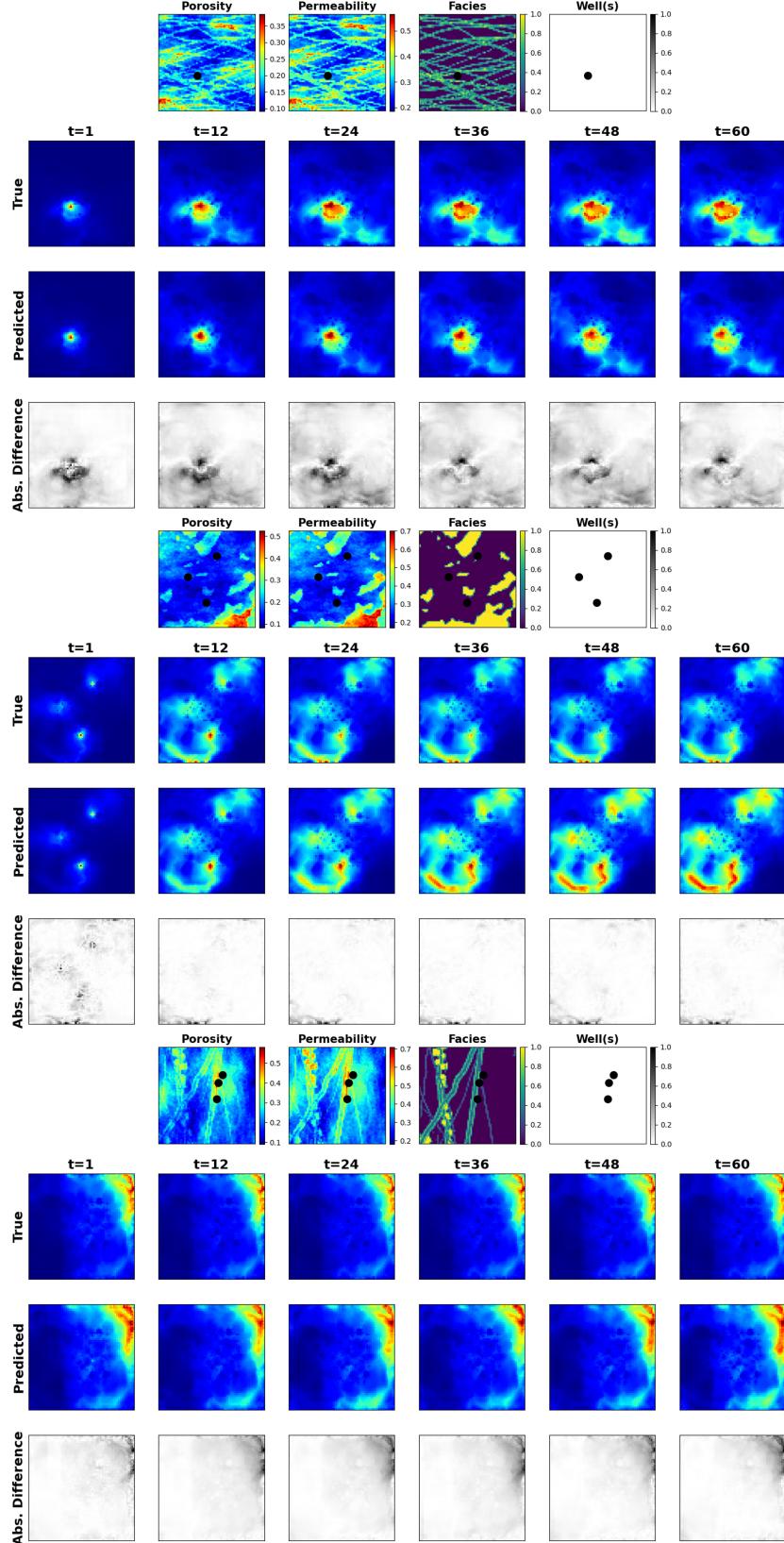


Figure 19: Normalized pressure distribution over time for 3 random testing realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.

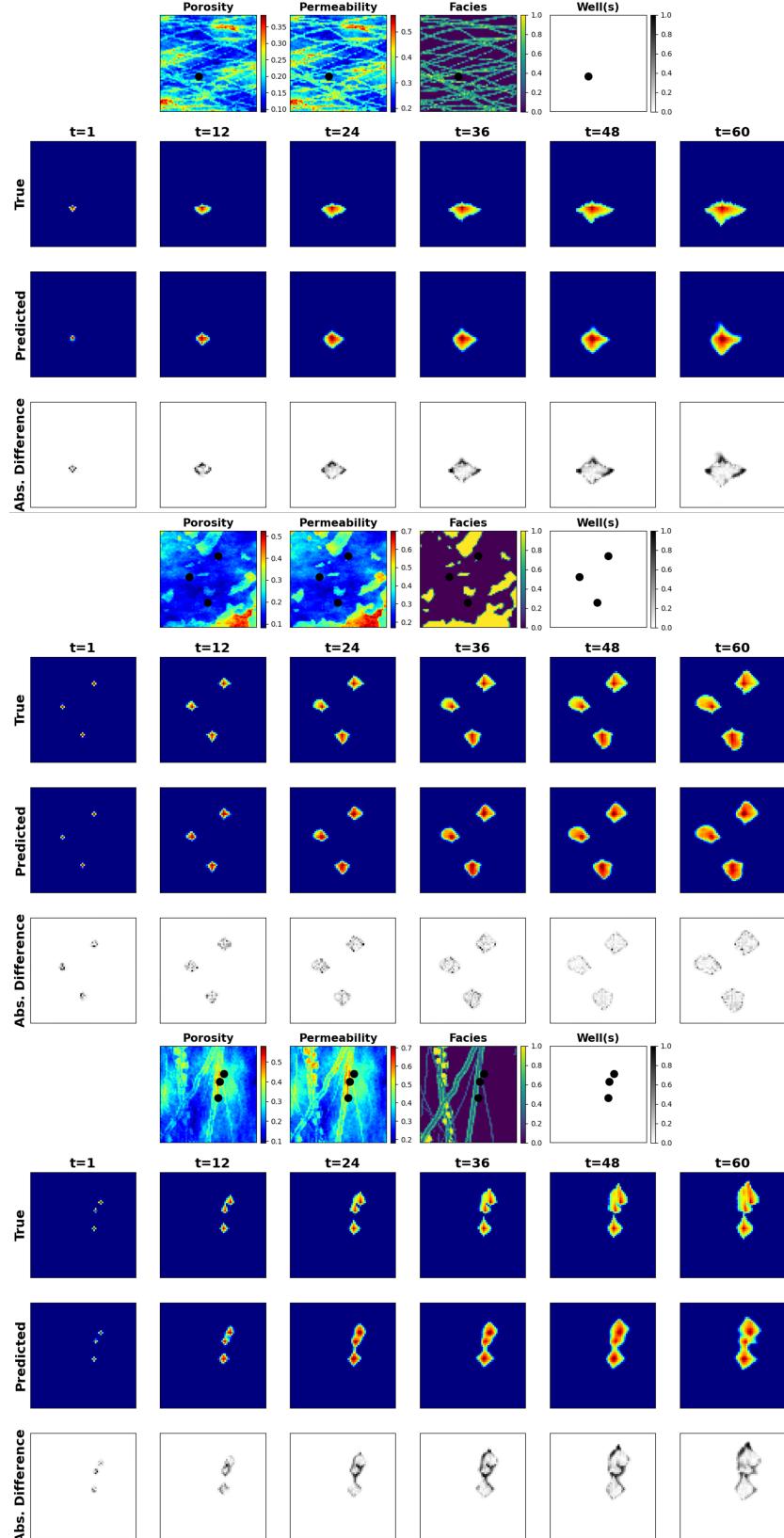


Figure 20: Saturation distribution over time for 3 random testing realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.

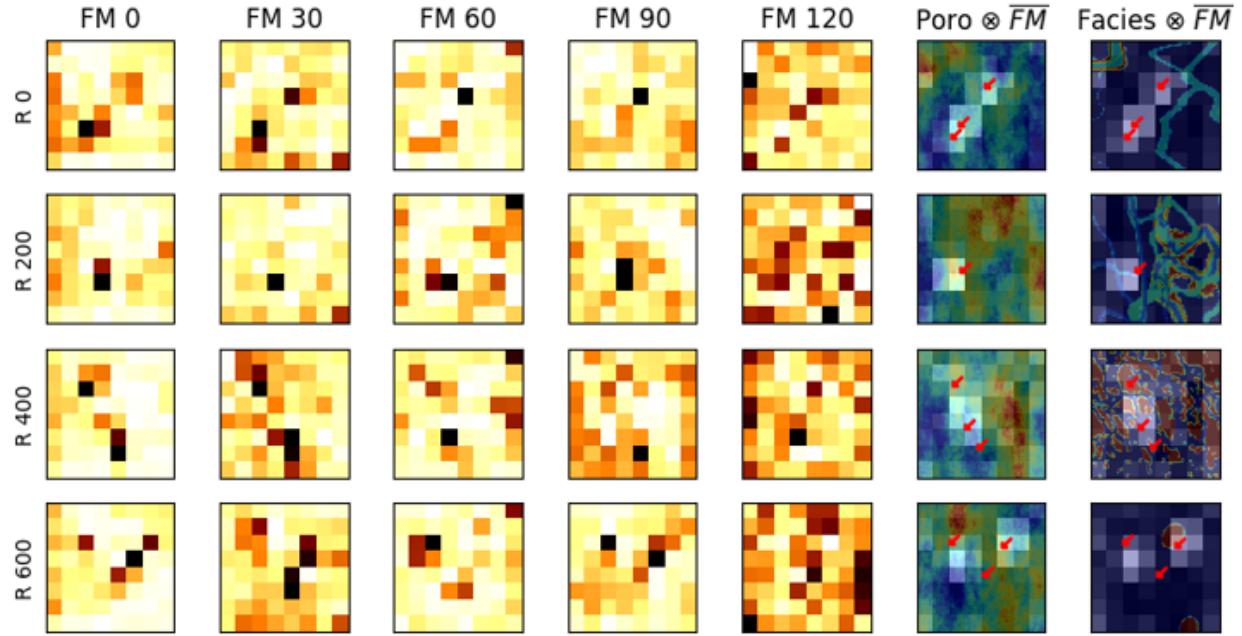


Figure 21: Five random feature maps (FM) of z_m^3 for 4 random realizations. Their average is superimposed on top of the porosity and facies distributions to show the attention mechanism of the encoder. Bright colors represent higher attention and dark colors represent lower attention.

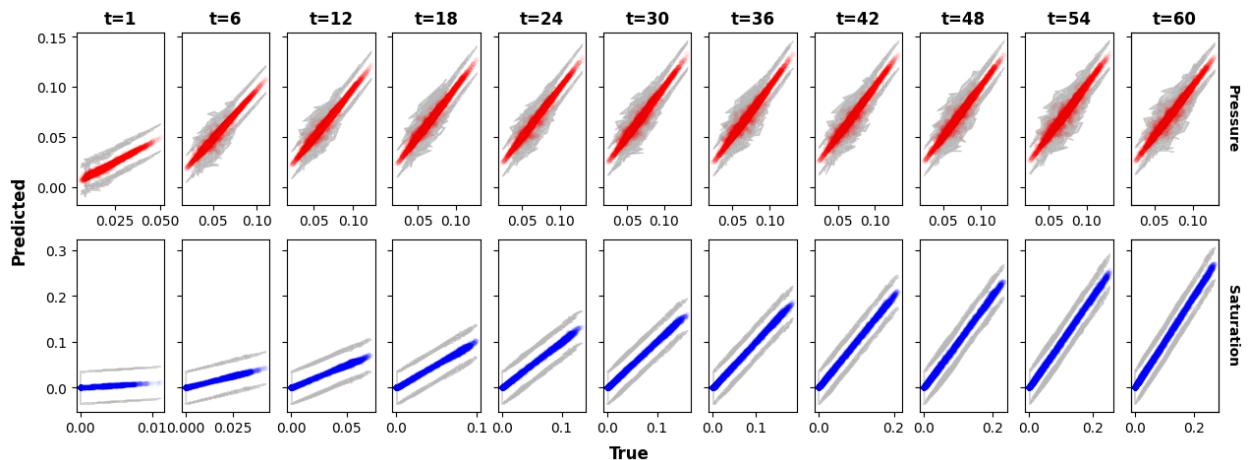


Figure 22: True versus predicted average normalized pressure (top) and saturation (bottom) over time for the testing data. The gray portion represents the 95% confidence bands, which narrow over time.

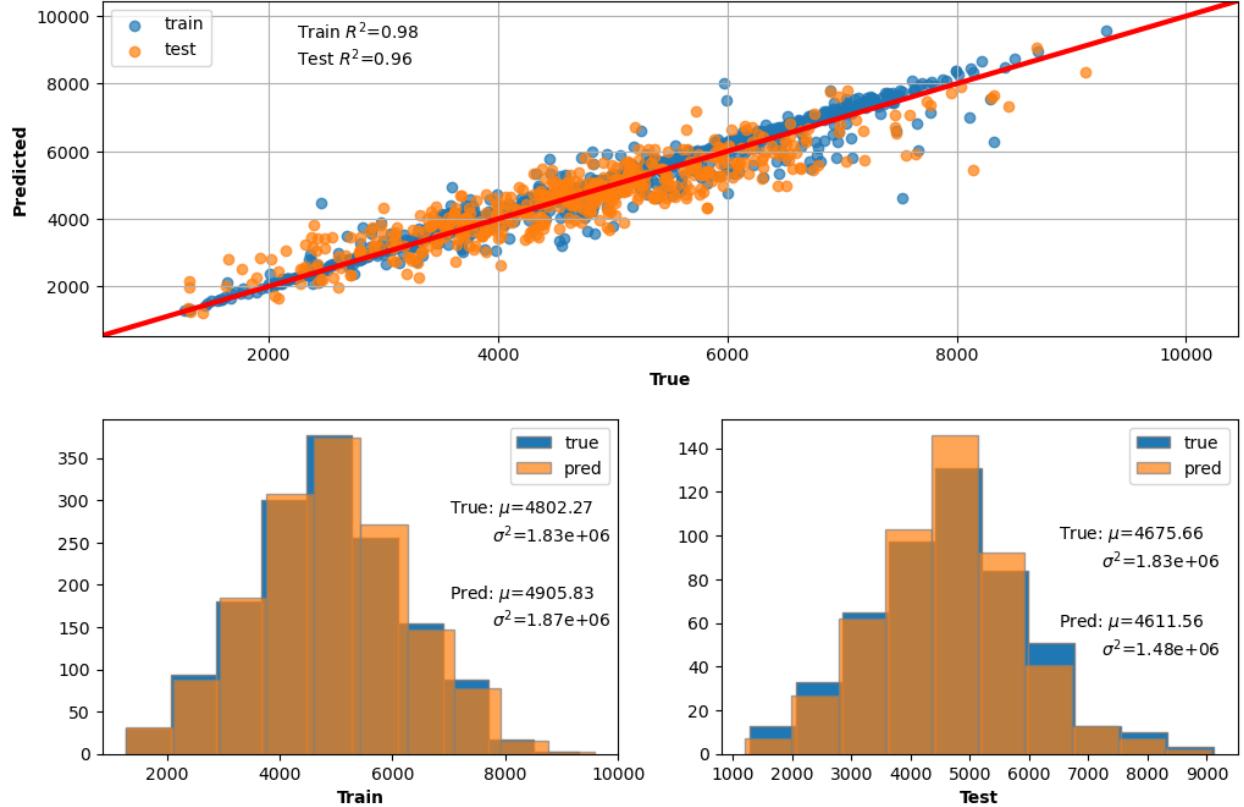


Figure 23: (Top) True vs. predicted cumulative CO₂ volume injected via pixel-wise saturation. (Bottom) True vs. predicted distributions of cumulative CO₂ saturation for training (left) and testing (right).

380 of the previous dynamic latent representation and the intermediate encoding parameterizations. Thus, our
 381 architecture makes the proxy model an image-to-video prediction formulation for dynamic reservoir states
 382 from a static geologic realization.

383 To further demonstrate the effectiveness of our Stochastic pix2vid model for geologic CO₂ storage op-
 384 erations, we plot the cumulative pixel-wise CO₂ saturation as a surrogate for the cumulative CO₂ volume
 385 injected. For all training and testing realizations, Figure 23 shows the sum of pixel-wise CO₂ saturation and
 386 the probability density function (PDF) of the true versus predicted saturations. We observe an R^2 of 98%
 387 for training and 96% for testing in the cumulative CO₂ saturation of true versus predicted results, and a
 388 conformable PDFs for both training and testing.

389 Our Stochastic pix2vid method has several limitations. In order to learn the spatiotemporal relationships
 390 between input images and output videos, the model requires substantial amounts of training data, which
 391 in turn require expensive physics-based numerical simulation runs. Moreover, the method would require
 392 retraining in order to apply to a different subsurface flow and transport problem, increasing the time required
 393 for generating the training data and the time required to retrain the model. One major limitation is the

394 inability to predict for timesteps beyond those present in the training data. The architecture of the Stochastic
395 pix2vid is designed to reconstruct only the 11 timesteps present in d , therefore it is capable of interpolation
396 for steps in between the training timesteps, but incapable to forecast beyond $t = 5$ years (60 months). Lastly,
397 the method is designed for images at the resolution of 64×64 pixels, and preprocessing is required to reshape
398 training data of other dimensions to this size.

399 4 Conclusions

400 We develop a deep learning-based spatiotemporal proxy model to provide efficient flow predictions for a
401 large-scale GCS operations to support optimum decision making. Our proposed method, Stochastic pix2vid,
402 introduces the use of a spatiotemporal convolutional-recurrent architecture for dynamic predictions of CO₂
403 pressure and saturation distributions over time from a static geologic realization representing the subsur-
404 face uncertainty model. The framework is developed as an image-to-video prediction, which is an under-
405 determined estimation problem. Specifically, the implementation expands upon the architectures of current
406 encoder-recurrent-decoder models and provides a fast and accurate proxy as a replacement for physics-based
407 numerical reservoir simulation.

408 The spatiotemporal proxy is applied to a synthetic 2D GCS project with multiple uncertain geologic
409 scenarios and random number and location of injection well(s). A total of 1,000 geologic models are obtained
410 from a variety of possible geologic scenarios including fluvial, turbidite, and deepwater lobe systems. The
411 spatial distribution of porosity, permeability and facies, and the spatial location of the injector well(s) are used
412 as the input data. The proxy model is used to predict the dynamic reservoir response over time, namely the
413 video frames, corresponding to the dynamic CO₂ pressure and saturation distributions, which are obtained
414 offline for training using HFS. The total training time is 88 minutes on a single NVIDIA Quadro M6000
415 GPU, and predictions are obtained with 98-99% accuracy within approximately 4.6 milliseconds, compared
416 to the approximate 30 seconds required for HFS, a $6,500\times$ speedup.

417 There are several opportunities for future work. First, an extension to 3D geologic models and their
418 corresponding dynamic predictions is key to scaling up this method for real-world applications. Similarly,
419 although the Stochastic pix2vid proxy model is only trained for GCS prediction, it is applicable for a
420 range of processes such as ground-water, compositional, geothermal, or conventional oil and gas systems.
421 Moreover, it is possible to extend the Stochastic pix2vid model from a data-driven mapping to a PINN
422 by including the discretized form of the governing PDE in the loss function and minimizing the residuals.
423 Another future opportunity is to test the performance of the Stochastic pix2vid model on unseen timesteps,
424 either interpolating the training timesteps or extrapolating beyond the training timesteps. Furthermore, the

425 Stochastic pix2vid model can be used as a proxy in workflows for history matching and closed-loop reservoir
426 management.

427 Reproducibility

428 The code will be made publicly available on the author's repository (github.com/misaelmmorales and
429 github.com/GeostatsGuy).

430 Funding

431 This research did not receive any specific grant from funding agencies in the public, or not-for-profit sectors.

432 Declarations

433 The authors declare no conflict of interests.

434 Acknowledgements

435 The authors thank the Digital Reservoir Characterization Technology (DIRECT) and Formation Evaluation
436 (FE) Industry Affiliate Program at the University of Texas at Austin for supporting this work.

437 References

- 438 [1] K. Michael, A. Golab, V. Shulakova, J. Ennis-King, G. Allinson, S. Sharma, and T. Aiken. Geological storage of co₂ in saline aquifers—a review of the experience from existing storage operations. *International Journal of Greenhouse Gas Control*, 4(4):659–667, 2010. ISSN 1750-5836. doi: <https://doi.org/10.1016/j.ijggc.2009.12.011>.
- 442 [2] A. Goodman, G. Bromhal, B. Strazisar, T. Rodosta, W.F. Guthrie, D. Allen, and G. Guthrie. Comparison of methods for geologic storage of carbon dioxide in saline formations. *International Journal of Greenhouse Gas Control*, 18:329–342, 2013. doi: 10.1016/j.ijggc.2013.07.016. cited By 48.
- 445 [3] J.S. Levine, I. Fukai, D.J. Soeder, G. Bromhal, R.M. Dilmore, G.D. Guthrie, T. Rodosta, S. Sanguinito, S. Frailey, C. Gorecki, W. Peck, and A.L. Goodman. U.s. doe netl methodology for estimating the

- 447 prospective co₂ storage resource of shales at the national and regional scale. *International Journal of*
448 *Greenhouse Gas Control*, 51:81–94, 2016. doi: 10.1016/j.ijggc.2016.04.028. cited By 81.
- 449 [4] Bert Metz, Ogunlade Davidson, HC De Coninck, Manuela Loos, and Leo Meyer. *IPCC special report*
450 *on carbon dioxide capture and storage*. Cambridge: Cambridge University Press, 2005.
- 451 [5] Energy 2020. European commission. In *A strategy for competitive, sustainable and secure energy*, 2010.
- 452 [6] United nations. Agreement, p. *United Nations Treaty Collect*, pages 1–27, 2015.
- 453 [7] S. Bachu. Review of co₂ storage efficiency in deep saline aquifers. *International Journal of Greenhouse*
454 *Gas Control*, 40:188–202, 2015. doi: 10.1016/j.ijggc.2015.01.007. cited By 277.
- 455 [8] J.F.D. Tapia, J.-Y. Lee, R.E.H. Ooi, D.C.Y. Foo, and R.R. Tan. Optimal co₂ allocation and scheduling
456 in enhanced oil recovery (eor) operations. *Applied Energy*, 184:337–345, 2016. doi: 10.1016/j.apenergy.
457 2016.09.093.
- 458 [9] N. Castelletto, P. Teatini, G. Gambolati, D. Bossie-Codreanu, O. Vincké, J.-M. Daniel, A. Battistelli,
459 M. Marcolini, F. Donda, and V. Volpi. Multiphysics modeling of co₂ sequestration in a faulted saline
460 formation in italy. *Advances in Water Resources*, 62:570–587, 2013. doi: 10.1016/j.advwatres.2013.04.
461 006. cited By 25.
- 462 [10] Elnara Rustamzade, Wen Pan, John T. Foster, and Michael Pyrcz. Comparison of commingled and
463 sequential production schemes by sensitivity analysis for gulf of mexico paleogene deepwater turbidite
464 oil fields: A simulation study. *Energy Exploration & Exploitation*, 0(0):01445987231195679, 2023. doi:
465 10.1177/01445987231195679. URL <https://doi.org/10.1177/01445987231195679>.
- 466 [11] K. Rashid, W. Bailey, B. Couët, and D. Wilkinson. An efficient procedure for expensive reservoir-
467 simulation optimization under uncertainty. *SPE Economics and Management*, 5(4):21–33, 2013. doi:
468 10.2118/167261-PA. cited By 16.
- 469 [12] C. Luo, S.-L. Zhang, C. Wang, and Z. Jiang. A metamodel-assisted evolutionary algorithm for expensive
470 optimization. *Journal of Computational and Applied Mathematics*, 236(5):759–764, 2011. doi: 10.1016/
471 j.cam.2011.05.047. cited By 29.
- 472 [13] Javier E. Santos, Bernard Chang, Alex Gigliotti, Eric Guiltinan, Mohamed Mehana, Arvind Mohan,
473 James McClure, Qinjun Kang, Hari Viswanathan, Nicholas Lubbers, Masa Prodanovic, and Michael
474 Pyrcz. Learning from a big dataset of digital rock simulations. In *AGU Fall Meeting Abstracts*, volume
475 2021, pages H25O–1207, December 2021.

- 476 [14] Bailian Chen, Dylan R. Harp, Youzuo Lin, Elizabeth H. Keating, and Rajesh J. Pawar. Geologic co2
477 sequestration monitoring design: A machine learning and uncertainty quantification based approach.
478 *Applied Energy*, 225:332–345, 9 2018. ISSN 03062619. doi: 10.1016/j.apenergy.2018.05.044.
- 479 [15] Wenyue Sun and Louis J. Durlofsky. Data-space approaches for uncertainty quantification of co2
480 plume location in geological carbon storage. *Advances in Water Resources*, 123:234–255, 1 2019. ISSN
481 03091708. doi: 10.1016/j.advwatres.2018.10.028. cited By 23.
- 482 [16] Bailian Chen, Dylan R. Harp, Zhiming Lu, and Rajesh J. Pawar. Reducing uncertainty in geologic
483 co2 sequestration risk assessment by assimilating monitoring data. *International Journal of Greenhouse
484 Gas Control*, 94, 3 2020. ISSN 17505836. doi: 10.1016/j.ijggc.2019.102926.
- 485 [17] B. Li and S.M. Benson. Influence of small-scale heterogeneity on upward co2plume migration in storage
486 aquifers. *Advances in Water Resources*, 83:389–404, 2015. doi: 10.1016/j.advwatres.2015.07.010. cited
487 By 84.
- 488 [18] Su Jiang and Louis J. Durlofsky. Use of multifidelity training data and transfer learning for efficient
489 construction of subsurface flow surrogate models. *Journal of Computational Physics*, 474, 2 2023. ISSN
490 10902716. doi: 10.1016/J.JCP.2022.111800.
- 491 [19] *Best Practices in Automatic Permeability Estimation: Machine-Learning Methods vs. Conventional
492 Petrophysical Models*, volume Day 4 Tue, June 13, 2023 of *SPWLA Annual Logging Symposium*, 06
493 2023. doi: 10.30632/SPWLA-2023-0084.
- 494 [20] H. Wu, N. Lubbers, H.S. Viswanathan, and R.M. Pollyea. A multi-dimensional parametric study of
495 variability in multi-phase flow dynamics during geologic co2 sequestration accelerated with machine
496 learning. *Applied Energy*, 287, 2021. doi: 10.1016/j.apenergy.2021.116580. cited By 14.
- 497 [21] Siddharth Misra, Yusuf Falola, Polina Churilova, Rui Liu, Chung-Kan Huang, and Jose F. Delgado.
498 Deep learning assisted extremely low-dimensional representation of subsurface earth. *SSRN Electronic
499 Journal*, 8 2022. doi: 10.2139/SSRN.4196705.
- 500 [22] Ademide O. Mabadeje and Michael J. Pyrcz. Rigid transformations for stabilized lower dimensional
501 space to support subsurface uncertainty quantification and interpretation, 2023.
- 502 [23] Mingliang Liu, Dario Grana, and Tapan Mukerji. Randomized tensor decomposition for large-scale
503 data assimilation problems for carbon dioxide sequestration. *Mathematical Geosciences*, 54:1139–1163,
504 5 2022. ISSN 18748953. doi: 10.1007/S11004-022-10005-1/FIGURES/17.

- 505 [24] S.W.A. Canchumuni, A.A. Emerick, and M.A.C. Pacheco. Towards a robust parameterization for
506 conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers and*
507 *Geosciences*, 128:87–102, 2019. doi: 10.1016/j.cageo.2019.04.006. cited By 80.
- 508 [25] Y. Zhang, P. Vouzis, and N.V. Sahinidis. Gpu simulations for risk assessment in co2 geologic sequestra-
509 tion. *Computers and Chemical Engineering*, 35(8):1631–1644, 2011. doi: 10.1016/j.compchemeng.2011.
510 03.023. cited By 20.
- 511 [26] Bicheng Yan, Dylan Robert Harp, Bailian Chen, and Rajesh J. Pawar. Improving deep learning per-
512 formance for predicting large-scale geological co2 sequestration modeling through feature coarsening.
513 *Scientific Reports*, 12:1–12, 11 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-24774-6.
- 514 [27] Zeeshan Tariq, Murtada Saleh Aljawad, Amjad Hasan, Mobeen Murtaza, Emad Mohammed, Ammar El-
515 Husseiny, Sulaiman A Alarifi, Mohamed Mahmoud, and Abdulazeez Abdulraheem. A systematic review
516 of data science and machine learning applications to the oil and gas industry. *Journal of Petroleum*
517 *Exploration and Production Technology*, pages 1–36, 2021.
- 518 [28] Mohammad Ali Mirza, Mahtab Ghoroori, and Zhangxin Chen. Intelligent petroleum engineering. *En-*
519 *gineering*, 18:27–32, 2022. ISSN 2095-8099. doi: <https://doi.org/10.1016/j.eng.2022.06.009>.
- 520 [29] Jean-Paul Chiles and Pierre Delfiner. *Geostatistics: modeling spatial uncertainty*, volume 713. John
521 Wiley & Sons, 2012.
- 522 [30] Michael J Pyrcz and Clayton V Deutsch. *Geostatistical reservoir modeling*. Oxford University Press,
523 USA, 2014.
- 524 [31] Proctor Joshua Brunton, Steve and Nathan Kutz. Discovering governing equations from data by sparse
525 identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the*
526 *United States of America*, 2016. doi: 10.1073/pnas.1517384113.
- 527 [32] He Xiaolong Fries, William and Youngsoo Choi. Lasdi: Parametric latent space dynamics identification.
528 *Computer Methods in Applied Mechanics and Engineering*, 2022. doi: 10.1016/j.cma.2022.115436.
- 529 [33] Choi Youngsoo Fries William Belof Jonathan He, Xiaolong and Jiun-Shyan Chen. glasdi: Parametric
530 physics-informed greedy latent space dynamics identification. *Journal of Computational Physics*, 2023.
- 531 [34] M. Liu and D. Grana. Time-lapse seismic history matching with an iterative ensemble smoother and
532 deep convolutional autoencoder. *Geophysics*, 85(1):M15–M31, 2020. cited By 2.

- 533 [35] Syamil Mohd Razak, Anyue Jiang, and Behnam Jafarpour. Latent-space inversion (lsi): a deep learning
 534 framework for inverse mapping of subsurface flow data. *Computational Geoscience*, 26:71–99, 11 2022.
 535 doi: 10.1007/s10596-021-10104-8.
- 536 [36] S. Oladyshkin, H. Class, and W. Nowak. Bayesian updating via bootstrap filtering combined with
 537 data-driven polynomial chaos expansions: Methodology and application to history matching for carbon
 538 dioxide storage in geological formations. *Computational Geosciences*, 17(4):671–687, 2013. doi: 10.
 539 1007/s10596-013-9350-6. cited By 36.
- 540 [37] Anqi Bao, Eduardo Gildin, Abhinav Narasingam, and Joseph S. Kwon. Data-driven model reduction
 541 for coupled flow and geomechanics based on dmd methods. *Fluids*, 4:138, 7 2019. ISSN 2311-5521. doi:
 542 10.3390/FLUIDS4030138.
- 543 [38] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang.
 544 Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
- 545 [39] Liu Yang, Dongkun Zhang, and George Em Karniadakis. Physics-informed generative adversarial net-
 546 works for stochastic differential equations, 2018.
- 547 [40] N. Wang, H. Chang, and D. Zhang. Efficient uncertainty quantification for dynamic subsurface flow with
 548 surrogate by theory-guided neural network. *Computer Methods in Applied Mechanics and Engineering*,
 549 373, 2021. doi: 10.1016/j.cma.2020.113492. cited By 33.
- 550 [41] Emilio Jose Rocha Coutinho, Marcelo Dall'Aqua, and Eduardo Gildin. Physics-aware deep-learning-
 551 based proxy reservoir simulation model equipped with state and well output prediction. *Frontiers in
 552 Applied Mathematics and Statistics*, 7:49, 9 2021. ISSN 22974687. doi: 10.3389/FAMS.2021.651178/
 553 BIBTEX.
- 554 [42] Yinhao Zhu, Nicholas Zabaras, Phaedon-Stelios Koutsourelakis, and Paris Perdikaris. Physics-
 555 constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification with-
 556 out labeled data. *Journal of Computational Physics*, 394:56–81, oct 2019. doi: 10.1016/j.jcp.2019.05.024.
 557 URL <https://doi.org/10.1016%2Fj.jcp.2019.05.024>.
- 558 [43] B Yegnanarayana. *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.
- 559 [44] Jeff Heaton. Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016,
 560 800 pp, isbn: 0262035618. *Genetic programming and evolvable machines*, 19(1-2):305–307, 2018.

- 561 [45] Yimin Liu and Louis J Durlofsky. 3d cnn-pca: A deep-learning-based parameterization for complex
562 geomodels. *Computers & Geosciences*, 148:104676, 2021.
- 563 [46] Zixiao Yang, Qiyu Chen, Zhesi Cui, Gang Liu, Shaoqun Dong, and Yiping Tian. Automatic recon-
564 struction method of 3d geological models based on deep convolutional generative adversarial networks.
565 *Computational Geosciences*, 26:1135–1150, 2022. doi: 10.1007/s10596-022-10152-8.
- 566 [47] Su Jiang and Louis J Durlofsky. Data-space inversion using a recurrent autoencoder for time-series
567 parameterization. *Computational Geosciences*, 25:411–432, 2021.
- 568 [48] Yanrui Ning, Hossein Kazemi, and Pejman Tahmasebi. A comparative machine learning study for time
569 series oil production forecasting: Arima, lstm, and prophet. *Computers and Geosciences*, 164:105126, 7
570 2022. ISSN 00983004. doi: 10.1016/j.cageo.2022.105126.
- 571 [49] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin
572 transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF*
573 *international conference on computer vision*, pages 10012–10022, 2021.
- 574 [50] Liuqing Yang, Sergey Fomel, Shoudong Wang, Xiaohong Chen, Wei Chen, Omar M. Saad, and Yangkang
575 Chen. Porosity and permeability prediction using a transformer and periodic long short-term network.
576 *Geophysics*, 88(1):WA293–WA308, 01 2023. ISSN 0016-8033. doi: 10.1190/geo2022-0150.1.
- 577 [51] Eduardo Maldonado Cruz and Michael J Pyrcz. Multi-horizon well performance forecasting with tem-
578 poral fusion transformers. *Available at SSRN 4403939*.
- 579 [52] Wen Pan, Carlos Torres-Verdín, and Michael J. Pyrcz. Stochastic pix2pix: A new machine learning
580 method for geophysical and well conditioning of rule-based channel reservoir models. *Natural Resources*
581 *Research*, 30:1319–1345, 4 2021. ISSN 15738981. doi: 10.1007/S11053-020-09778-1/FIGURES/24.
- 582 [53] Bogdan Sebacher and Stefan Adrian Toma. Bridging deep convolutional autoencoders and ensemble
583 smoothers for improved estimation of channelized reservoirs. *Mathematical Geosciences*, 54:903–939, 7
584 2022. ISSN 18748953. doi: 10.1007/S11004-022-09997-7/TABLES/3.
- 585 [54] Jichao Bao, Liangping Li, and Arden Davis. Variational autoencoder or generative adversarial networks?
586 a comparison of two deep learning methods for flow and transport data assimilation. *Mathematical*
587 *Geosciences*, 54:1017–1042, 8 2022. ISSN 18748953. doi: 10.1007/S11004-022-10003-3/FIGURES/17.
- 588 [55] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmen-
589 tation. *CoRR*, 2015. cited By 358.

- 590 [56] Eduardo Maldonado-Cruz and Michael J. Pyrcz. Fast evaluation of pressure and saturation predictions
591 with a deep learning surrogate flow model. *Journal of Petroleum Science and Engineering*, 212:110244,
592 5 2022. ISSN 0920-4105. doi: 10.1016/J.PETROL.2022.110244.
- 593 [57] Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. U-fno—an
594 enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water
595 Resources*, 163:104180, 2022. ISSN 0309-1708. doi: <https://doi.org/10.1016/j.advwatres.2022.104180>.
- 596 [58] Gege Wen, Zongyi Li, Qirui Long, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson.
597 Real-time high-resolution co 2 geological storage prediction using nested fourier neural operators. *Energy
598 & Environmental Science*, 2023. ISSN 1754-5692. doi: 10.1039/d2ee04204e.
- 599 [59] Honggeun Jo, Wen Pan, Javier E Santos, Hyungsik Jung, and Michael J Pyrcz. Machine learning
600 assisted history matching for a deepwater lobe system. *Journal of Petroleum Science and Engineering*,
601 207:109086, 2021.
- 602 [60] Feng Zhang, Long Nghiem, and Zhangxin Chen. Evaluating reservoir performance using a transformer
603 based proxy model. *Geoenergy Science and Engineering*, 226:211644, 2023.
- 604 [61] Daowei Zhang and Heng Li. Efficient surrogate modeling based on improved vision transformer neural
605 network for history matching. *SPE Journal*, pages 1–17, 2023.
- 606 [62] Yong Do Kim and Louis J. Durlofsky. Convolutional – recurrent neural network proxy for robust
607 optimization and closed-loop reservoir management. *Computational Geosciences*, pages 1–24, 1 2023.
608 ISSN 1420-0597. doi: 10.1007/S10596-022-10189-9/TABLES/1.
- 609 [63] Meng Tang, Yimin Liu, and Louis J. Durlofsky. A deep-learning-based surrogate model for data as-
610 similation in dynamic subsurface flow problems. *Journal of Computational Physics*, 413, 7 2020. ISSN
611 10902716. doi: 10.1016/J.JCP.2020.109456.
- 612 [64] M. Tang, Y. Liu, and L.J. Durlofsky. Deep-learning-based surrogate flow modeling and geological
613 parameterization for data assimilation in 3d subsurface flow. *Computer Methods in Applied Mechanics
614 and Engineering*, 376, 2021. doi: 10.1016/j.cma.2020.113636. cited By 39.
- 615 [65] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Generating videos with scene dynamics, 2016.
- 616 [66] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean
617 square error, 2016.

- 618 [67] Ruben Villegas, Jimei Yang, Seunghoon Hong, Xunyu Lin, and Honglak Lee. Decomposing motion and
619 content for natural video sequence prediction, 2018.
- 620 [68] Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. Mocogan: Decomposing motion and
621 content for video generation, 2017.
- 622 [69] Xingjian SHI, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO.
623 Convolutional lstm network: A machine learning approach for precipitation nowcasting. In C. Cortes,
624 N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing
625 Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf.
- 627 [70] Michael Iliadis, Leonidas Spinoulas, and Aggelos K. Katsaggelos. Deep fully-connected networks for
628 video compressive sensing, 2017.
- 629 [71] Kai Xu and Fengbo Ren. Csvideonet: A real-time end-to-end learning framework for high-frame-rate
630 video compressive sensing. In *2018 IEEE Winter Conference on Applications of Computer Vision
631 (WACV)*, pages 1680–1688. IEEE, 2018.
- 632 [72] Michael Dorkenwald, Timo Milbich, Andreas Blattmann, Robin Rombach, Konstantinos G. Derpanis,
633 and Björn Ommer. Stochastic image-to-video synthesis using cinns, 2021.
- 634 [73] Aleksander Holynski, Brian Curless, Steven M. Seitz, and Richard Szeliski. Animating pictures with
635 eulerian motion fields, 2020.
- 636 [74] Karsten Pruess, Curtis M Oldenburg, and GJ Moridis. Tough2 user’s guide version 2. Technical report,
637 Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 1999.
- 638 [75] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the
639 IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.
- 640 [76] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference
641 on computer vision and pattern recognition*, pages 7132–7141, 2018.
- 642 [77] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient
643 for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.
- 644 [78] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.
- 645 [79] D.S. Oliver, A.C. Reynolds, and N. Liu. *Inverse theory for petroleum reservoir characterization and
646 history matching*, volume 9780521881517. 2008. doi: 10.1017/CBO9780511535642. cited By 766.

- 647 [80] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from
648 error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 4 2004.
649 ISSN 1941-0042. doi: doi.org/10.1109/TIP.2003.819861.
- 650 [81] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint*
651 *arXiv:1711.05101*, 2017.
- 652 [82] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint*
653 *arXiv:1412.6980*, 2014.
- 654 [83] Nicolas Remy, Alexandre Boucher, and Jianbing Wu. *Applied Geostatistics with SGeMS: A User’s*
655 *Guide*. Cambridge University Press, 2009.
- 656 [84] G. W. Verly. *Sequential Gaussian Cosimulation: A Simulation Method Integrating Several Types of*
657 *Information*, pages 543–554. Springer Netherlands, Dordrecht, 1993. ISBN 978-94-011-1739-5. doi:
658 10.1007/978-94-011-1739-5_42.
- 659 [85] M.J. Pyrcz, J.B. Boisvert, and C.V. Deutsch. A library of training images for fluvial and deepwater
660 reservoirs and associated code. *Computers Geosciences*, 34(5):542–560, 2008. ISSN 0098-3004. doi:
661 <https://doi.org/10.1016/j.cageo.2007.05.015>.
- 662 [86] Misael M. Morales and Michael Pyrcz. GeostatsGuy/MLTrainingImages: MachineLearningTrainingIm-
663 ages_v1.0.0, March 2023. URL <https://doi.org/10.5281/zenodo.7702128>.
- 664 [87] Knut-Andreas Lie. *An introduction to reservoir simulation using MATLAB/GNU Octave: User guide*
665 *for the MATLAB Reservoir Simulation Toolbox (MRST)*. Cambridge University Press, 2019.
- 666 [88] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz
667 Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing*
668 *systems*, 30, 2017.
- 669 [89] Q. Li and G. Liu. *Risk assessment of the geological storage of CO2: A review*. 2016. doi: 10.1007/
670 978-3-319-27019-7_13. cited By 39.
- 671 [90] R.A. Chadwick, R. Arts, and O. Eiken. 4d seismic quantification of a growing co2 plume at sleipner,
672 north sea. *Petroleum Geology Conference Proceedings*, 6(0):1385–1399, 2005. doi: 10.1144/0061385.
673 cited By 188.

- 674 [91] R.A. Chadwick and D.J. Noy. History-matching flow simulations and timelapse seismic data from the
675 sleipner co₂ plume. *7th Petroleum Geology Conference [FROM MATURE BASINS to NEW FRON-*
676 *TIERS] (London, 3/30/2009-4/2/2009) Proceedings*, 2:1171–1182, 2010. cited By 31.
- 677 [92] Ismael Dawuda and Sanjay Srinivasan. Geologic modeling and ensemble-based history matching for
678 evaluating co₂ sequestration potential in point bar reservoirs. *Frontiers in Energy Research*, 10:867083,
679 2022.