

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-Verdin^{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 present 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 the
32 subsurface over and beyond the life of the project is a critical component to support optimum GCS project
33 decision making for safe and secure CO₂ sequestration. A schematic of typical GCS operations is shown in
34 Figure 1, including storage in depleted oil and gas reservoir and deep saline formations, and CO₂ enhanced
35 oil and coalbed methane recovery [7–9]. However, there are several technical challenges associated with
36 the subsurface models to support GCS operations. In order 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 [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 to calculate fast pre-
48 dictions of subsurface flow and transport behavior for real-time feedback on the impact of geological and
49 engineering controls on CO₂ behavior in the subsurface over time [19–21]. Dimensionality reduction tech-
50 niques are supervised or unsupervised machine learning methods that compress (or encode) the data, X , into
51 a lower-dimensional latent feature representation, z , and decompress (or decode) the latent representation
52 either: (1) back to the original data space, \hat{X} (unsupervised, AutoEncoder), or (2) to a new target feature
53 space, y (supervised, Encoder-Decoder) [22–24], as shown in Figure 2. These are enabled by recent advance-
54 ments in deep learning algorithms and in computing architecture and power, enabling GPU-enabled neural
55 network models that have accelerated the fields of forward and inverse modeling [25, 26]. Classical statistical
56 modeling methods are often hindered by the size of the models and their conditioning to big data, i.e., that
57 is data with volume, velocity, variety, value, and veracity [27, 28]. By analyzing big data sets, machine
58 learning techniques can uncover complex patterns and relationships in lower-dimensional, latent feature rep-
59 resentations that may not be discernible through traditional statistical and geostatistical methods [29–31].

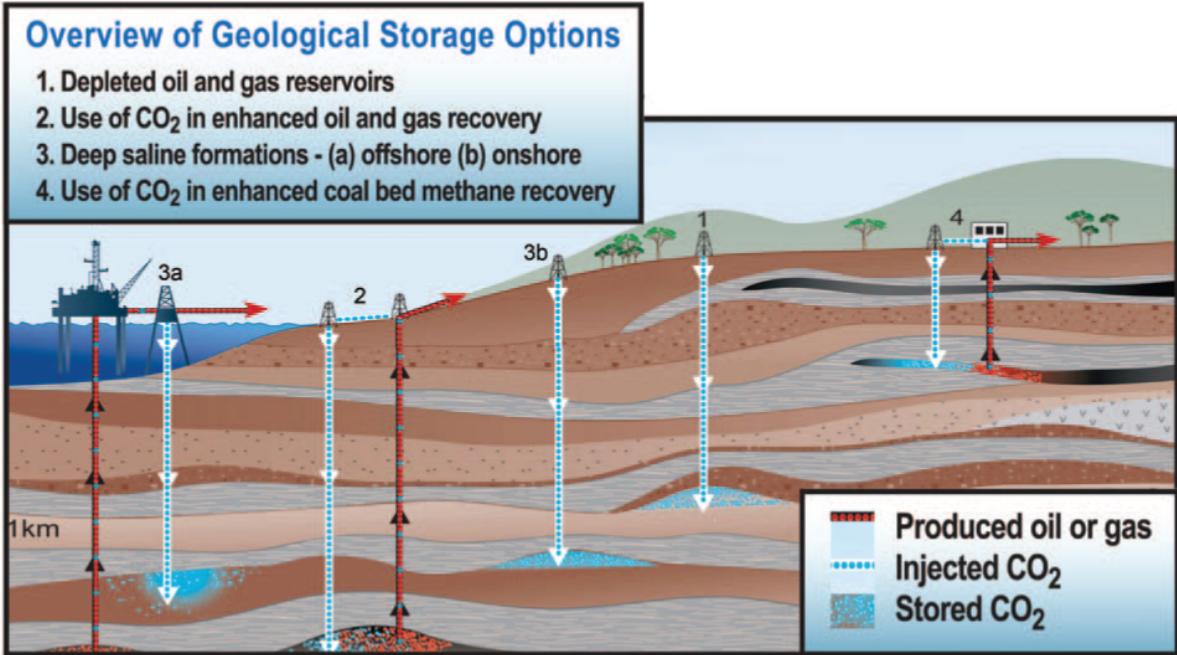


Figure 1: Types of geologic CO₂ storage operations and the geologic formations that can be used for sequestration. *Modified from the Carbon Dioxide Cooperative Research Center (CO2CRC), <http://www.co2crc.com.au/about/co2crc>*

60 When combined with a latent space modeling framework, machine learning approaches can efficiently and
 61 accurately exploit hidden features in the data, removing redundancies or noise, and decreasing the order of
 62 the problem significantly [32, 33].

63 Machine learning approaches applied to the subsurface are divided into two main categories, namely
 64 purely data-driven models or physics-informed models. Data-driven proxy models, or data mapping op-
 65 erators, are neural network architectures trained with labeled data that produce a mapping from input
 66 predictor features to output response features [34, 35]. On the other hand, the training process to match

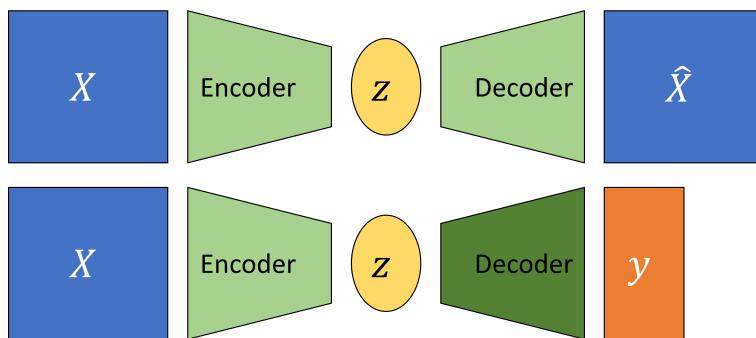


Figure 2: Dimensionality reduction model structures. Unsupervised AutoEncoder structure (top), and supervised Encoder-Decoder structure (bottom).

67 training data for PINNs is regularized with the minimization of the (physical) loss from the residual of the
68 governing partial differential equations (PDEs) along with the losses associated with the initial and boundary
69 conditions [36, 37]. However, other variants of PINNs such as physics-guided or physics-constrained neural
70 networks - where the PDE loss is not embedded in the training but the models have specific architectures or
71 parameters to mimic the physics in the system - have proven useful for subsurface energy resource engineer-
72 ing applications [38–40]. One disadvantage of machine learning techniques is that they require significant
73 amounts of training data but once trained, these prediction models can be applied over a wide variety of
74 settings and conditions for which they have been specifically trained [41, 42], but suffer from lack of gener-
75 alization, i.e., inability to provide accurate predictions away from the training data. For both data-driven
76 and physics-informed approaches, typically, spatial relationships are modeled through convolutional neural
77 networks (CNNs) [43, 44] and the temporal relationships through recurrent neural networks (RNNs) [45, 46],
78 but recent advancements in transformer-based architectures improve performance compared to the CNN and
79 RNN methods for spatial and temporal latent feature representations [47–49]. In general, efficient compres-
80 sion of the input features into a representative latent space is proven as an effective approach for spatial and
81 temporal parameterization of the forward or inverse problem.

82 A number of machine learning-based proxy (or surrogate) models have been developed to estimate the
83 reservoir behavior in subsurface energy resource applications. Most techniques rely on the concept of image
84 translation, or pix2pix, where a target image(s) is predicted from an input image(s) [50–53], as shown in
85 Figure 3. Maldonado-Cruz and Pyrcz [54] develop a convolutional U-Net model to predict pressure and
86 saturation states given an uncertain geologic realization. This work is an example of image-to-image static
87 forecasting, where the time state is given as an input, and the proxy model will predict a single response
88 state of pressure and saturation at the given time. Wen et al. [55] develop a Fourier Neural Operator (FNO)
89 architecture to predict image-to-image response states of pressure and saturation from an uncertain geologic
90 realization and was further extended for multi-scale and nested domains [56]. These methods are based on
91 a pix2pix, or image-to-image prediction, where a specific timestep is used as an input feature to predict the
92 relationship between the geologic model and the reservoir response at that specific timestep. This implies
93 that pix2pix or image-to-image methods are formulated as an even-determined or sometimes over-determined
94 estimation problem, where the number of input features is equal to or greater than the number of output
95 features. Moreover, numerous other proxy models have been developed for subsurface applications using
96 more complex architectures such as generative adversarial networks (GANs) [57] and transformers [58, 59].
97 Despite showing consistent results and significant speedups compared to traditional numerical simulation,
98 pix2pix models do not capture the spatiotemporal relationships and dynamic response of the subsurface
99 system.

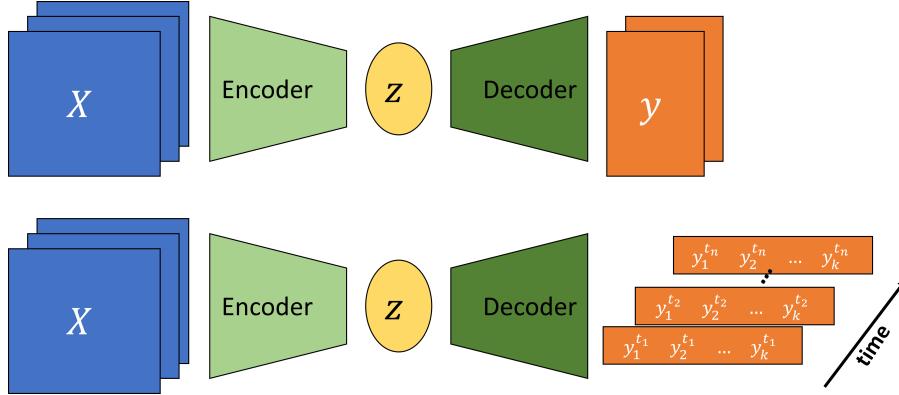


Figure 3: Image-to-image (pix2pix) (top) and image-to-timeseries (bottom) Encoder-Decoder structures.

Moving beyond image-to-image predictions, Kim and Durlofsky [60] 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 well flow rates and pressures over time from a static geologic realization. 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. [61, 62] 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 input. This type of proxy model is formulated to resolve the more complex under-determined 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 [63–67]. Iliadis et al. [68] are one of the first to develop a deep learning-based framework for video compressive sensing to reconstruct a video sequence from a single measured frame 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 [69] 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 [18, 61, 62] 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. [70] develop a conditional invertible neural network (cINN) as a bijective mapping between image and video domains using a dynamic latent representation. The cINN architecture allowed for video-to-image and image-to-video predictions, proving possible the generation of video frames from a static input image. Finally, Holynski et al. [71] implemented the idea of Eulerian motion fields to define the moving portions of the image and thus were able to accurately reconstruct a series of video frames from a static image using a spatiotemporal latent space parameterization. These advancements in the field of computer vision and video compressed sensing serve as a foundation for our image-to-video proxy model.

We propose the Stochastic pix2vid, a novel image-to-video spatiotemporal proxy model for the prediction of dynamic reservoir behavior over time from a suite of static geologic realizations representing a subsurface uncertainty model. Our model exploits the spatial and temporal structures in latent space to dynamically reconstruct the time-dependent pressure and saturation states from a static geologic realization. The encoder portion of the network receives as inputs the static geologic realization with geological depositional, high permeability channels representing the porosity, permeability, and facies spatial distributions, and the location of CO₂ injection well(s). The model then reconstructs the dynamic pressure and saturation distributions using a spatiotemporal decoder network with convolutional long short-term memory (ConvLSTM) layers, which are concatenated with the residuals of the spatial latent parameterizations from the encoder network. Thus, it is not an encoder-recurrent-decoder architecture, but instead a fully spatiotemporal convolutional-recurrent image-to-video synthesis model. Our stochastic pix2vid model shows significant advantages compared to image-to-image and encoder-recurrent-decoder models in terms of computational efficiency and prediction accuracy and can be used as a replacement for high-fidelity simulations (HFS) in GCS projects as an image-to-video mapping operator.

In the methodology section, we discuss the proposed spatiotemporal proxy model architecture as well as the geologic modeling and numerical reservoir simulation steps required to generate the training data. In the results and discussion sections, we evaluate the training and performance of the proposed proxy model and compare its efficiency and accuracy to high-fidelity numerical simulations using a 2D synthetic case for large-scale GCS operations.

152 2 Methodology

153 This section describes the governing equations, reservoir model and simulation specifications, and the archi-
 154 tecture and training strategy of the Stochastic pix2vid model.

155 **2.1 Governing equations** For the CO₂-water multiphase flow problem, the general form of the mass
 156 accumulation for component $\kappa = \text{CO}_2$ or water is given by [72]:

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

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

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

158 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
 159 fraction of component κ present in phase p . For each component κ , there is also the advective mass flux
 160 $F^\kappa|_{adv}$ obtained by summing over all phases p ,

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

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

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

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

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

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

167 2.2 Reservoir Model and Simulation

168 We use SGeMS [73] to construct an ensemble of realizations that is representative of various potential
 169 geologic scenarios for CO₂ storage in deep geological formations, e.g., fluvial, turbidite, and deepwater lobe
 170 systems. Using sequential Gaussian co-simulation [74], we generate a set of 1,000 random porosity (ϕ) and

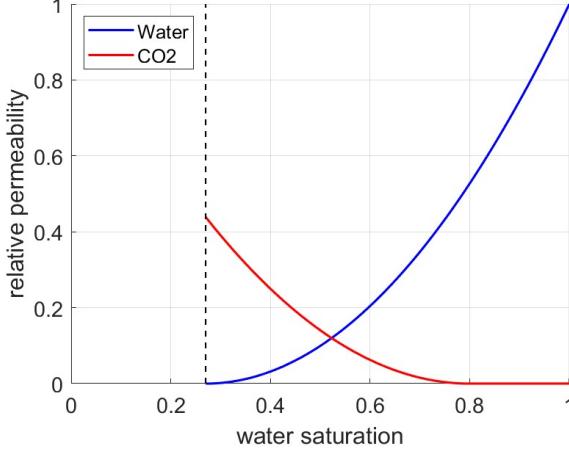


Figure 4: Relative permeability curves for the CO_2 -water system. The residual saturations are 0.27 and 0.2 for water and CO_2 , respectively.

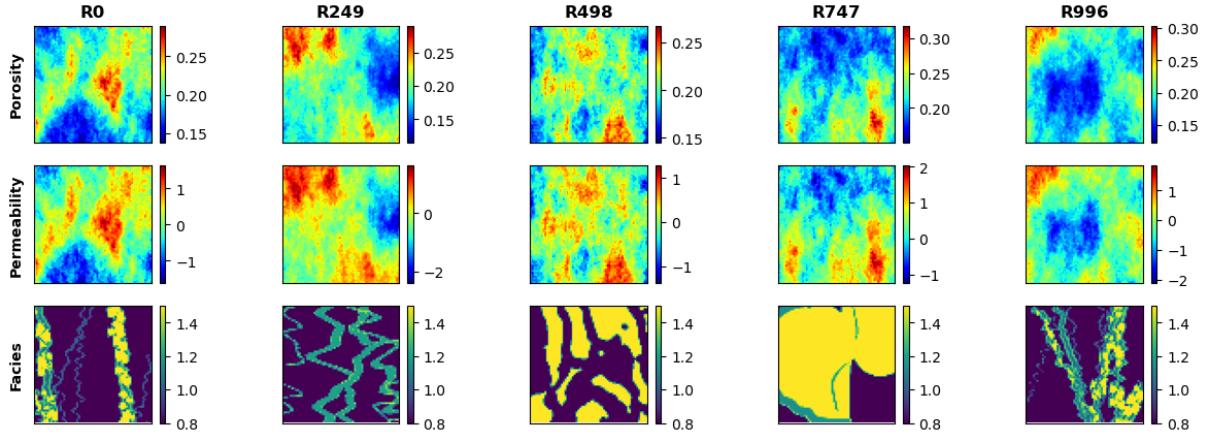


Figure 5: Spatial distribution of porosity (top), permeability (middle), and facies (bottom) for 5 random realizations.

171 permeability (k) distributions with a wide range of values, as shown in Figure 5. Facies distributions are
 172 obtained from a library of deepwater fluvial training images [75, 76]. These encompass a wide range of
 173 possible geologic scenarios including marked point (lobe, ellipse, and bar), FluvSim (channel, channel-levee,
 174 and channel-levee-splay), surface based (compensational cycles of lobes), and bank retreat (channel complex).
 175 To generate consistent porosity and permeability distributions with the facies-based geologic scenarios, we
 176 condition the original porosity and permeability distributions to the facies distributions. The resulting fluvial
 177 distributions are shown in Figure 6.

178 The conditioned fluvial porosity and permeability distributions are used as inputs for the numerical
 179 simulation of geologic CO_2 storage using MRST [77]. Specifically, the MRST-co2lab module is used as an
 180 automatic-differentiation framework for the compositional simulation of the two-phase CO_2 -water problem.
 181 The reservoir is initialized as a fully water saturated zone (i.e., aquifer) with an initial pressure of 4,000 psi.

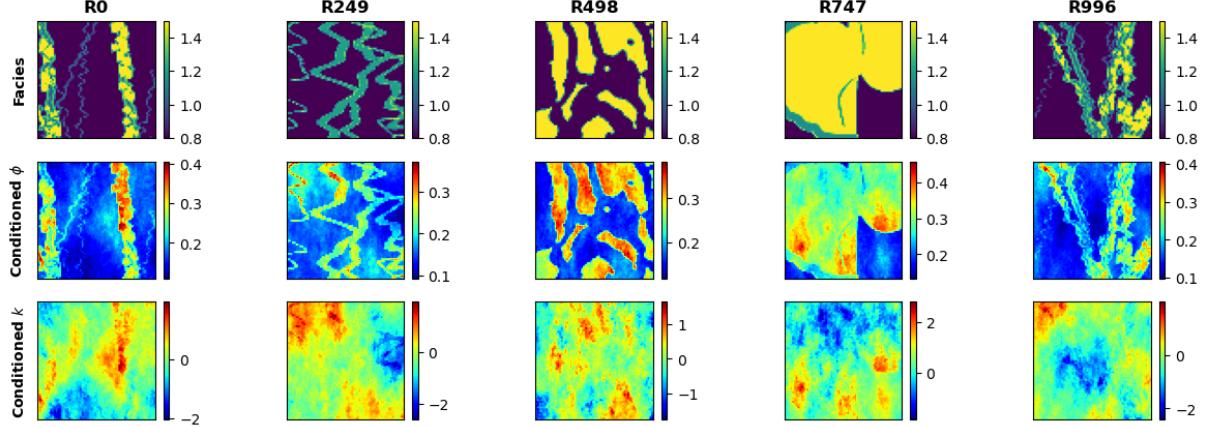


Figure 6: Spatial distribution conditioned to facies (top) for porosity (middle) and permeability (bottom) for 5 random realizations.

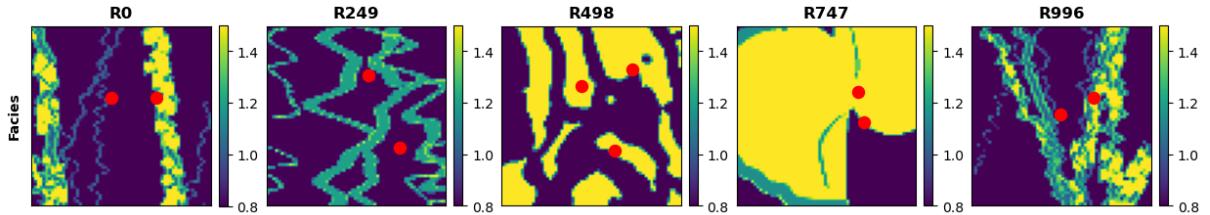


Figure 7: CO_2 injection well(s) location (red) overlaid over facies distributions for 5 random realizations.

182 The reservoir has constant isothermal conditions and constant pressure boundary conditions to represents a
 183 large-scale geologic CO_2 storage project with negligible dip, such as found in the Illinois Basin and parts of
 184 the North Sea and Gulf of Mexico.

185 The model has dimensions of 1km-1km-100m in the x-, y-, and z-directions, respectively. We use 64
 186 uniform grid cells in the x- and y-directions. The grid design is sufficiently refined to resolve the pressure
 187 and saturation plumes in highly heterogeneous reservoirs while remaining computationally tractable for the
 188 purpose of training deep learning models. A random number of injection wells, $w \in [1, 3]$, are placed randomly
 189 along the reservoir for each of the 1,000 realizations, no closer than 250m from the boundaries, as shown in
 190 Figure 7. The injection well(s) are randomly placed and not conditioned to zones of preferential porosity,
 191 permeability, nor facies. Each injection well has a constant radius of 0.1m and a single and continuous
 192 perforation that injects pure supercritical CO_2 at a constant rate such that the total injection rate of the w
 193 well(s) is 0.5 megatons per year.

194 The numerical simulation is run for 5 years, monitored monthly, for a total of 60 timesteps. At each
 195 grid cell and for each time step, we resolve the implicit pressure, explicit saturation (IMPES) formulation of
 196 Eq. (1) to obtain the corresponding dynamic pressure and saturation distributions over time (videos) from

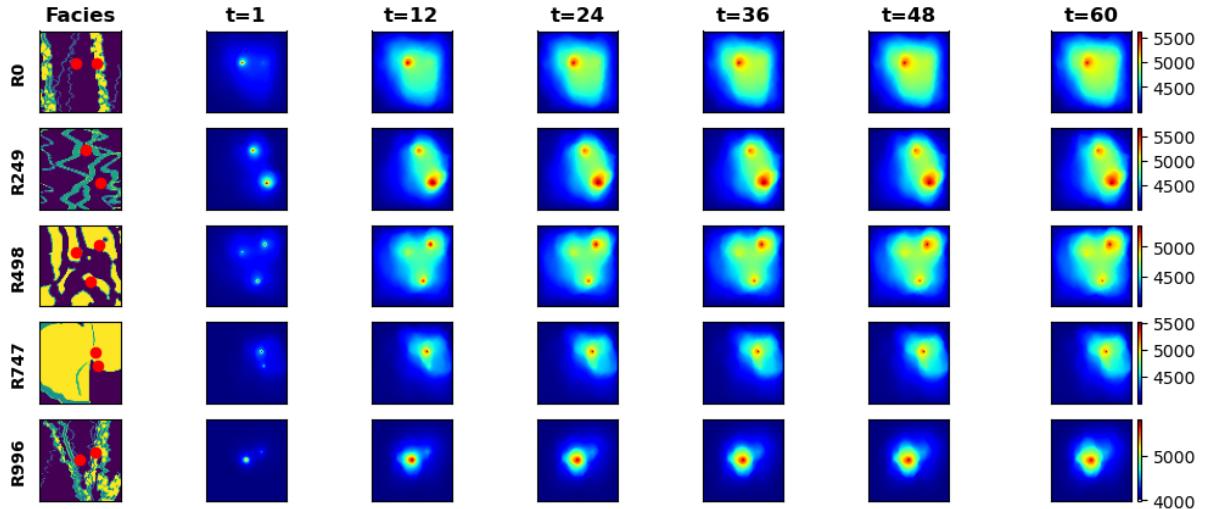


Figure 8: Pressure response distributions over time for 5 random realizations obtained from HFS (in psia).

197 the static geologic realizations of porosity and permeability conditioned to the fluvial facies (images) with
 198 random well(s) configuration. The pressure and saturation responses corresponding to the previously-shown
 199 geologic model realizations are shown in Figures 8 and 9, respectively.

200 2.3 Proxy Model Architecture

201 The Stochastic pix2vid model is designed as an image-to-video data-driven mapping operator from the
 202 static realizations of geologic distributions of porosity, permeability and facies as well as the injector well(s)
 203 distribution, to the dynamic responses of pressure and saturation distributions over time. A single model is
 204 trained to predict both pressure and saturation distributions over time as a multi-channel output from the
 205 multi-channel input features.

206 Let m represent a geologic model realization of porosity, permeability, facies, and injector well(s) distri-
 207 butions, such that $m = \{\phi, k, \text{facies}, w\}$. The dynamic responses of pressure and saturation over time are
 208 given by $d = f(m)$, such that $d = \{P(t), S(t)\}$ and f is the physics-based numerical reservoir simulation.
 209 Our aim is to replace f with a more efficient data-driven proxy by training the Stochastic pix2vid model.
 210 For this purpose, we exploit the latent structures in space and time of the static inputs and dynamic outputs
 211 through a spatiotemporal encoder-decoder architecture.

212 The encoder portion of the network is comprised of sequential convolutional layers to compress the
 213 spatial features of the model realizations into a latent parameterization z_m , given by $z_m = Enc(m)$. In their
 214 compressed representation, these features represent the salient characteristics of the geologic distributions.
 215 The decoder portion of the network is designed as a series of recursive residual convolutional-recurrent layers,
 216 such that the latent space z_m is recursively decoded into the dynamic distributions of pressure and saturation
 217 over time. The previous timestep latent representations, z_d^t , are used in the subsequent timestep to refine

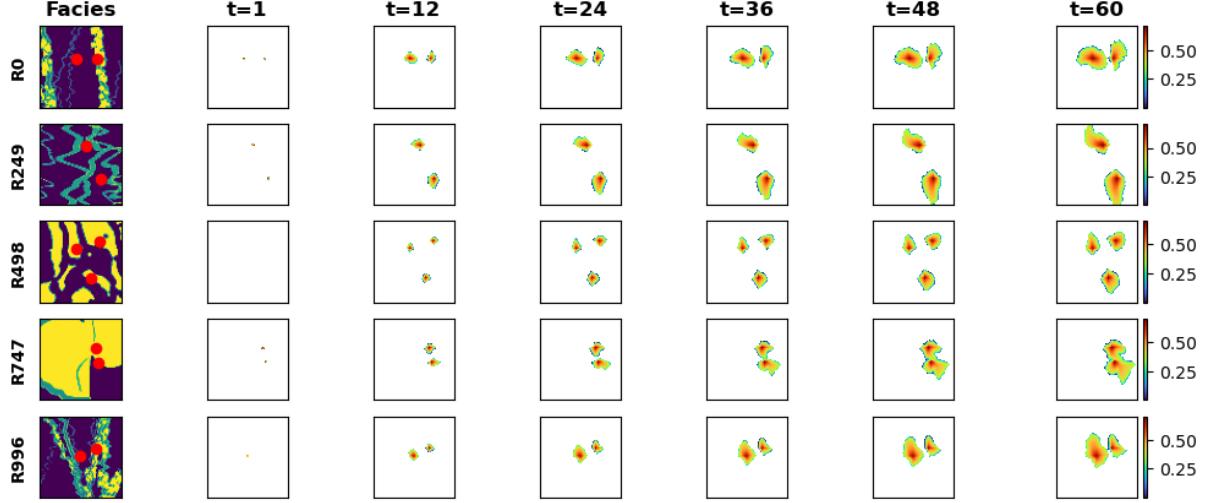


Figure 9: Saturation response distributions over time for 5 random realizations obtained from HFS.

218 the outputs and reduce systematic error propagation in time. Thus, the full architecture is represented as

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

219 The encoder portion compresses the geologic realizations, m , into a latent representation z_m through the
 220 use of depthwise separable convolutions [78]. This type of convolution learns the parameters for each channel
 221 in the image separately, avoiding mixing of variables or loss of resolution, as shown in Figure 10. This is
 222 especially important when dealing with Gaussian-distributed permeability and porosity in combination with
 223 binomial-distributed facies and binary well(s) location distributions. Each separable convolution layer is
 224 regularized with an l_1 -norm weight of 1×10^{-6} to control the null space in latent feature space. Moreover,
 225 we use a Squeeze-and-Excite layer to improve channel interdependence, and to avoid mixing and loss of
 226 resolution [79]. Each Squeeze-and-Excite layer will provide the optimal network weights for each channel
 227 independent of the other channels by passing the feature maps through a global pooling layer (squeeze) and a
 228 dense layer with nonlinear activation (excite), to add content aware mechanism for re-weighting each channel
 229 adaptively, as shown in Figure 11. Furthermore, by applying instance normalization, as opposed to the more
 230 common batch normalization, we achieve channel-independent normalization of the convolved features [80].
 231 Instance normalization is a special case of group normalization, where the numbers of channels per group is
 232 exactly 1, such that each channels gets its own normalization scheme, as shown in Figure 12. Parametric
 233 rectified linear units (PReLU) is used as the activation function, where at each minibatch iteration, the
 234 network learns the optimal leaky slope for activation in each layer, as shown in Figure 13. Finally, pooling
 235 and spatial dropout are applied, as the resulting feature map is reduced to half the input dimension. Through

- 236 3 convolutional encoding layers with filter size 3×3 , we obtain the latent parameterizations z_m^1 , z_m^2 , and z_m^3 .
 237 Table 1 summarizes the architecture and dimensions of each layer.



Figure 10: 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.

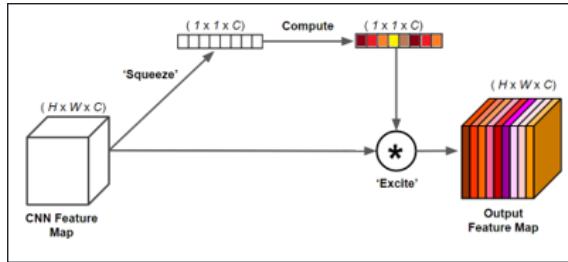


Figure 11: Schematic for a squeeze-and-excite layer. The "squeeze" layer takes the global average of the data for each channel, and the "excite" layer is a fully-connected layer with nonlinear activation to estimate the optimal weights for each channel in the data. The result is a weighted representation of the data based on their intrinsic global characteristics.

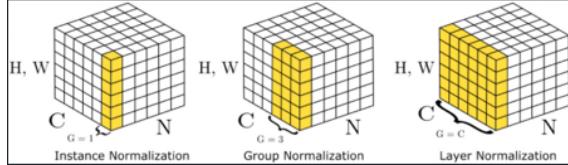


Figure 12: Schematic for instance normalization (left) compared to group normalization (center) and batch normalization (right). In an instance normalization layer, each channel will be normalized by themselves rather than normalizing the entire batch or a subset of channels (groups).

- 238 The decoder portion of the Stochastic pix2vid model extracts the spatiotemporal relationships from the
 239 latent representations of m to reconstruct the dynamic pressure and saturation distributions over time, d .
 240 To accurately reconstruct the spatiotemporal structure from the static latent space, z_m , we employ a series
 241 of convolutional-recurrent layers, namely a convolutional long-short term memory layer (ConvLSTM). The
 242 general form of a 2D ConvLSTM layer is shown in Figure 14. Through 3 convolutional-recurrent layers, we
 243 obtain the dynamic prediction of d as follows:
 244 Step 1: **Spatiotemporal decoding of z_m^3 :** The first ConvLSTM layer takes the smallest latent represen-
 245 tation, z_m^3 , and reconstructs the first decoded timestep z_d^3 .

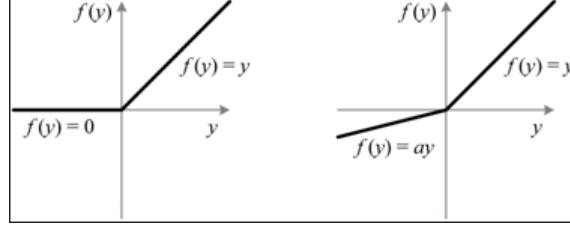


Figure 13: Schematic for the Parametric Rectified Linear Unit (PReLU) activation function (right) compared to the traditional ReLU activation function (left). The slope of the negative portion of the data, α , is learned for each batch.

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		
2	Spatial Dropout		$32 \times 32 \times 64 (z_m^1)$
	SeparableConv2D	$32 \times 32 \times 64$	
	Squeeze-and-Excite		
	Instance Norm		
3	PReLU + Pooling		
	Spatial Dropout		$16 \times 16 \times 128 (z_m^2)$
	SeparableConv2D	$16 \times 16 \times 128$	
	Squeeze-and-Excite		
4	Instance Norm		
	PReLU + Pooling		
	Spatial Dropout		$8 \times 8 \times 256 (z_m^3)$

- 246 Step 2: **Residual concatenation of z_m^2 :** The first decoded timestep, z_d^3 , is concatenated with the inter-
247 mediate static encoding z_m^2 to retain multi-scale features and improve prediction performance and
248 resolution.
- 249 Step 3: **Intermediate spatiotemporal decoding:** The second ConvLSTM layer takes the residual con-
250 catenation of the intermediate latent representations, $[z_m^2, z_d^3]$, to predict the next spatiotemporal
251 representation z_d^2 .
- 252 Step 4: **Residual concatenation of z_m^1 :** The intermediate decoded timestep, z_d^2 , is concatenated with the
253 largest static encoding z_m^1 .
- 254 Step 5: **Final spatiotemporal decoding:** The third ConvLSTM layer takes the residual concatenation of
255 the larger latent representations, $[z_m^1, z_d^2]$, to predict the full-scale dynamic output, d .
- 256 To enhance the performance of the spatiotemporal decoding, each ConvLSTM layer is followed by a batch

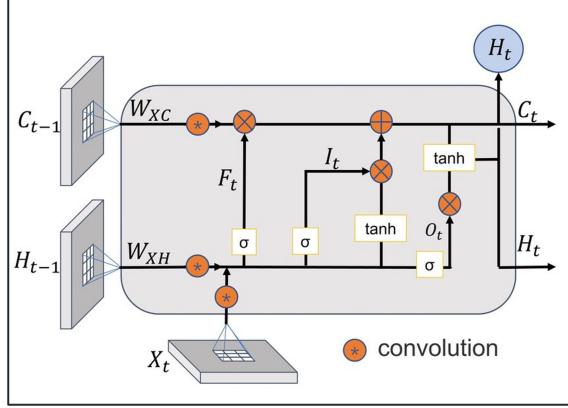


Figure 14: Schematic of a convolutional-LSTM (ConvLSTM) layer. The layer applies convolutional operations to the input data using a set of learnable filters to capture the spatial patterns. The recurrent part is a long short-term memory layer with memory and forget gates to capture the temporal patterns. LSTM units are applied to each spatial location separately allowing to capture both spatial and temporal dependencies in the data.

257 normalization, activation, and a transpose convolutional layer, the latter for downscaling the latent features
 258 to twice their dimension. Spatial dropout is applied, and the concatenated features are once more convolved
 259 and activated to obtain the layer prediction. Table 2 shows the architecture of the decoder network.

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

267 The complete Stochastic pix2vid architecture is shown in Figure 15. Here we observe the spatial com-
 268 pression of the geologic models, m , through the encoding portion of the network, and the spatiotemporal
 269 decoding and residual multi-scale concatenations through the decoder portion of the network. The result-
 270 ing architecture provides proxy model from static geologic models (images) to dynamic reservoir response
 271 (videos).

272 2.4 Training Strategy

273 The inputs to the Stochastic pix2vid are the geologic realizations, comprised of the distributions of
 274 porosity, permeability, facies, and injection well(s) location, represented as a matrix m of dimensions $64 \times$
 275 64×4 . The outputs are the results from the numerical reservoir simulation, namely pressure and saturation
 276 distributions over time, represented as a matrix d of dimensions $64 \times 64 \times 60 \times 2$. This yields an ill-posed

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		$t \times 64 \times 64 \times 2 (z_{d_t}^1)$

and under-determined estimation problem, which are known to be difficult to resolve [81, 82]. To improve the training efficiency and performance, we subsample in time from 60 timesteps to 11. In other words, instead of monthly monitoring, we predict the dynamic outputs at the initial step and every 6 months afterward; therefore the output matrix d has a final dimension of $64 \times 64 \times 11 \times 2$. We also perform min-max normalization so that the input and output features are in the range of $[0, 1]$, which greatly improves the performance of the nonlinear activation functions. Furthermore, we perform data augmentation by 90° rotation, making the network agnostic to orientation and effectively learning the flow physics in the system rather than memorizing spatial distribution patterns. The total amount of training data is therefore 2,000 realizations (after augmentation), which is split into 1,500 realizations for training and 500 realizations for testing. To improve model generalizability, at each epoch, each minibatch is split into 80/20 for training and validation sets, respectively.

A custom three-part loss function is used to accurately predict pixel-wise and perceptual information in the predictions. The mean squared error (MSE) is used to reconstruct the pixel-wise intensity values, while the mean absolute error (MAE) is used to optimize for the pressure and saturation plume edges. The third part is the structural similarity index metric (SSIM), which provides a perceptual image-to-image comparison of luminance, contrast, and structure [83]. For optimal training, the aim is to minimize the MSE and MAE while maximizing the SSIM for the true versus predicted outputs, d and \hat{d} , such that the total loss is given by:

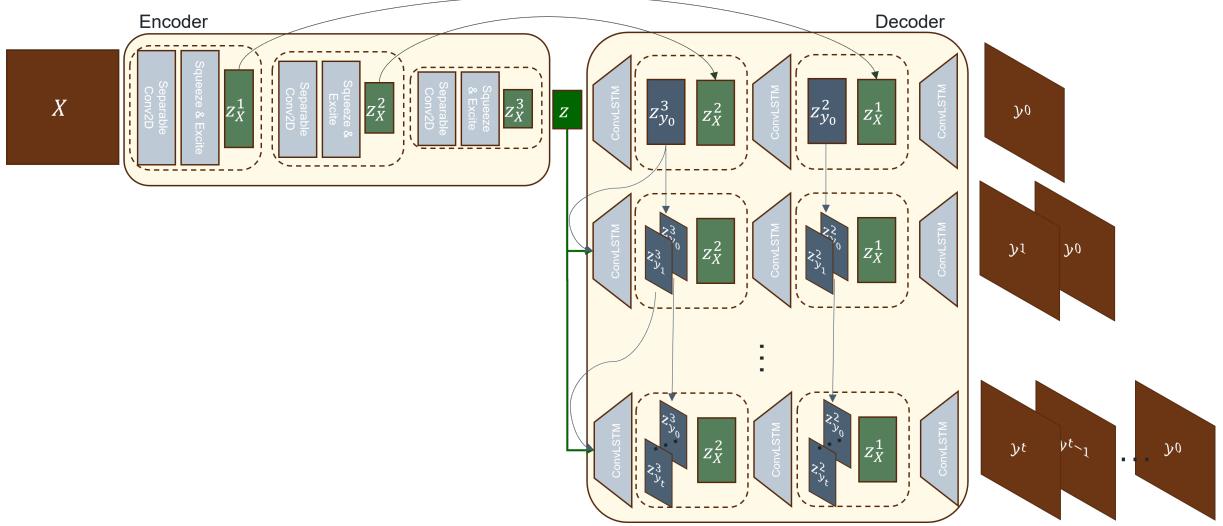


Figure 15: Architecture of the Stochastic pix2vid model. The input data, $X \equiv m$, is encoded through a series of convolutional layers to capture the spatial dependencies in the geologic models. The latent representation, z_m , is recursively passed through a spatiotemporal decoder with convolutional-recurrent layers, and concatenated with the residuals of the encoder to reconstruct iteratively the frames of the output (video) data, $y \equiv d$.

$$\mathcal{L} = \alpha(1 - SSIM) + (1 - \alpha)[\beta MSE + (1 - \beta) MAE] \quad (7)$$

where α and β are weighting coefficients obtained empirically as 0.33 and 0.66, respectively.

The model is trained using the AdamW optimizer [84]. This variant of the well-known adaptive momentum (Adam) optimizer [85] includes an added method to decay weights for the adaptive estimation of first-order and second-order moments. We implement a learning rate of 1×10^{-3} with a weight decay term of 1×10^{-5} .

3 Results

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

3.1 Training Performance

Using an NVIDIA Quadro M6000 GPU, we train for 100 epochs with a batch size of 50. The model has a total of 97,523,370 parameters, and the training time required is approximately 88 minutes for all 1,500 training realizations. The training and validation performance per epoch is shown in Figure 16. We observe minimal overfit in the validation set, corresponding to good model generalizability and prediction accuracy within the training data. Using physics-based numerical simulation, each realization requires approximately

309 30 seconds to obtain the dynamic pressure and saturation predictions from the static geologic models. Our
 310 Stochastic pix2vid model obtains the same results in approximately 4.59 milliseconds, corresponding to a
 311 $6,500\times$ speedup. The average MSE for the ensemble is 9.21×10^{-4} and 9.70×10^{-4} for training and testing,
 312 respectively. Similarly, the average SSIM for the ensemble is 98.97% and 97.91% for training and testing,
 313 respectively.

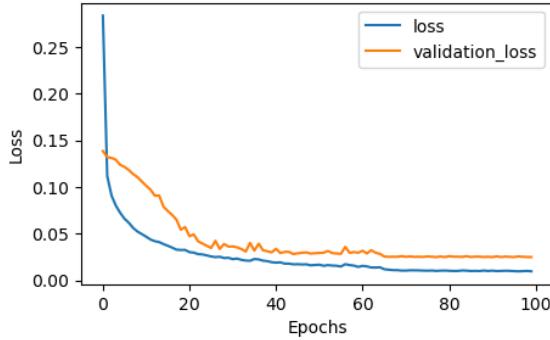


Figure 16: The total training and validation losses, \mathcal{L} , as a function of epoch number.

314 3.2 Prediction Results

315 After training the Stochastic pix2vid model with 1,500 realizations of static geologic models, $m =$
 316 $\{\phi, k, facies, w\}$, to predict the dynamic reservoir response, $d = \{P(t), S(t)\}$, we can compare the per-
 317 formance of the predictions for the training and unseen testing data.

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

323 Similarly, Figures 19 and 20 show the pressure and saturation distributions predictions along with the
 324 absolute difference to HFS for 3 testing realizations. We observe a similar performance, with an average MSE
 325 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% for
 326 saturation predictions. This indicates that the Stochastic pix2vid model has excellent generalization ability
 327 and achieves on par performance with HFS at a fraction of the computational cost.

328 It is interesting to note that the Stochastic pix2vid model is trained on a triple-loss function with MSE,
 329 MAE and SSIM. For both training and testing cases, we see that the average MSE for pressure is higher
 330 than that of saturation, while the opposite is true for the average SSIM. This can be attributed to the fact
 331 that there are more pixel-wise variations in pressure predictions, thus the loss focuses on matching those
 332 individual pixel-wise values. On the other hand, for saturation predictions, the contrast, luminance, and

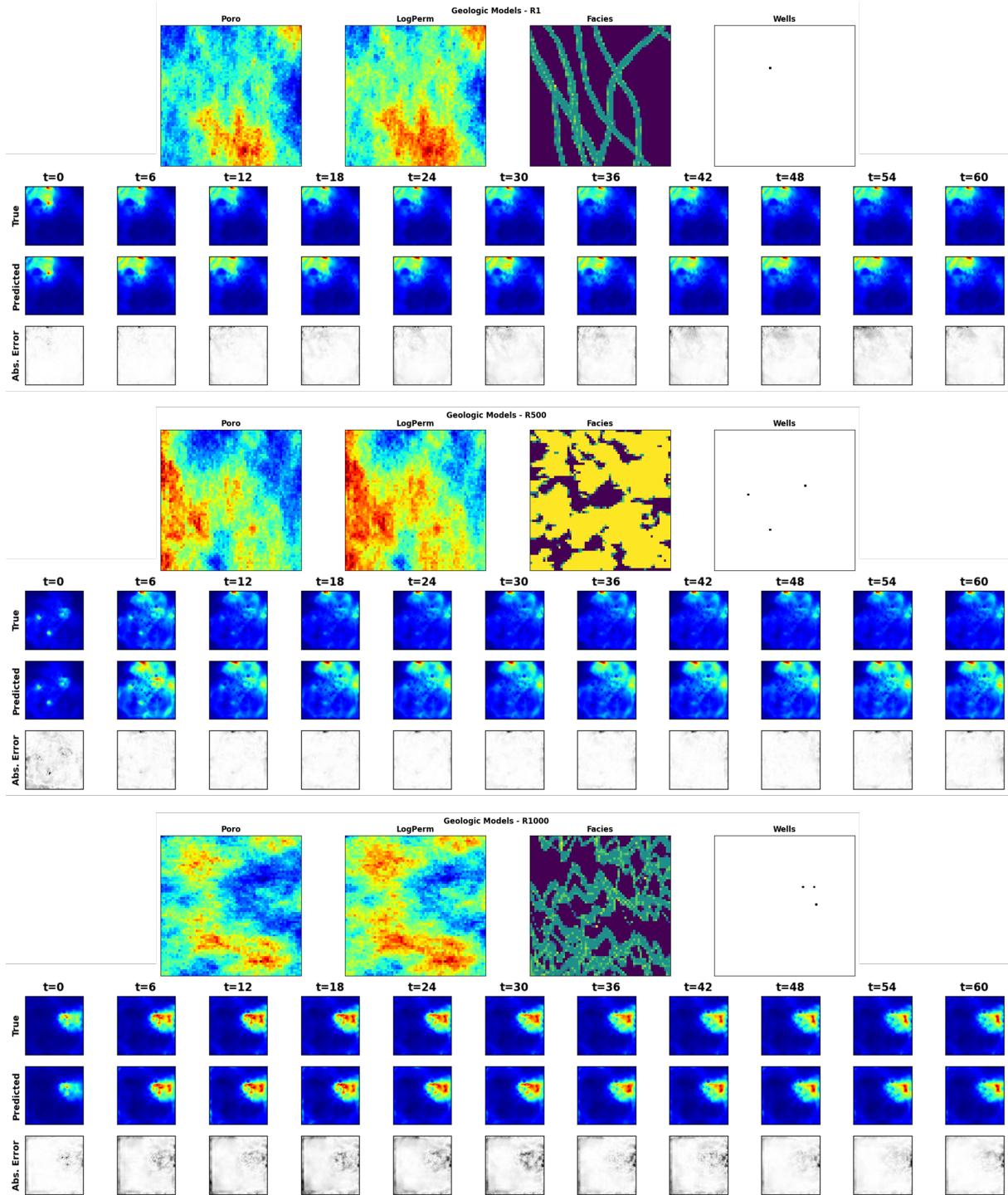


Figure 17: (Normalized) pressure distribution over time for 3 random training realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.

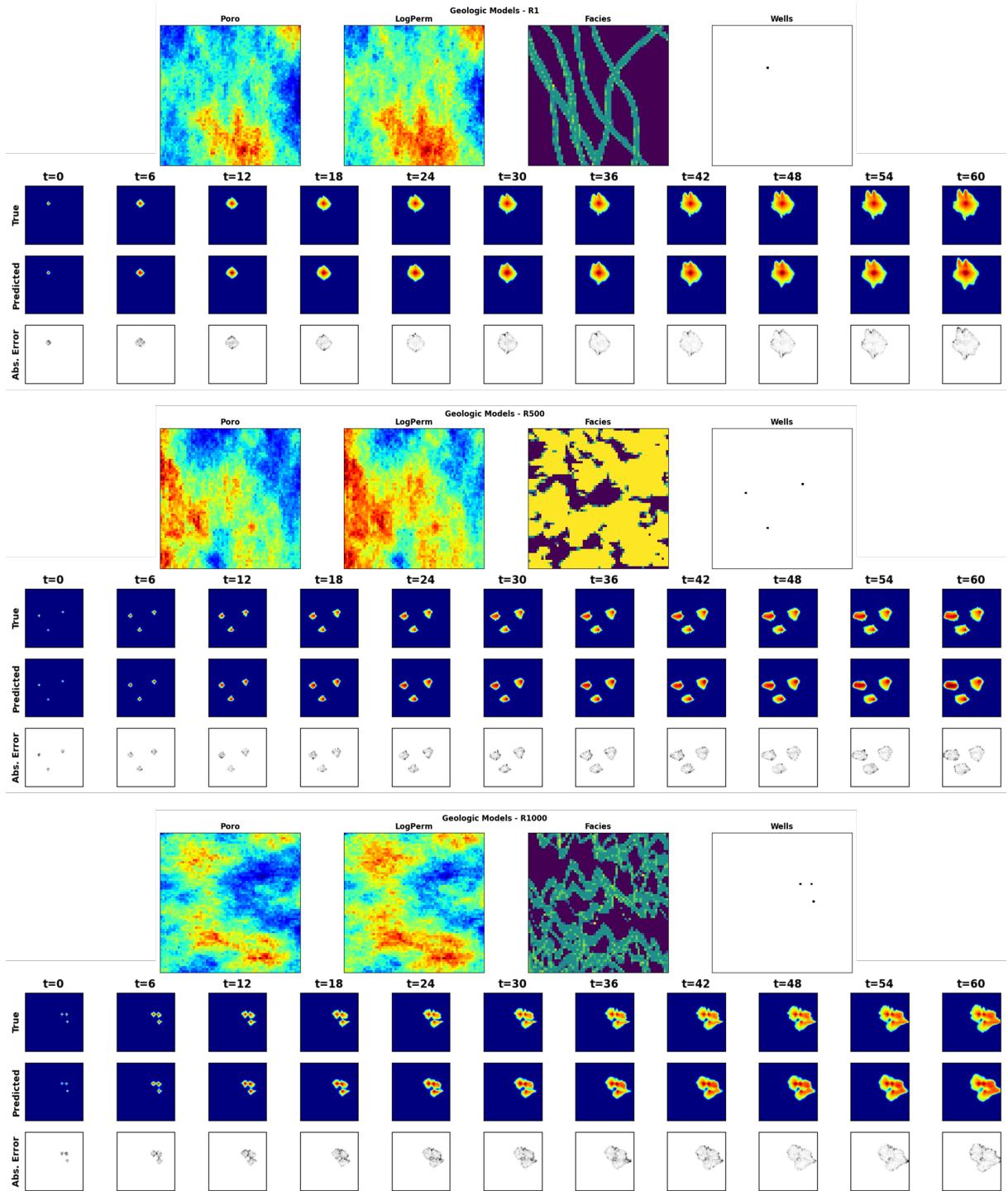


Figure 18: Saturation distribution over time for 3 random training realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.

333 structure play a bigger role in the prediction than the pixel-wise intensity values. Therefore, it is important
334 to take into account both metrics for training and validating spatiotemporal subsurface prediction models.

335 From **Section 2.3**, the first step of the Stochastic pix2vid model is to take the static geologic realizations,
336 m , and compresses them into a latent space representation, z_m , using the spatial encoder structure. Figure
337 21 show a random selection of latent feature maps, along with their superposition on the porosity and facies
338 distribution. This can be interpreted as an analog to the attention head mechanisms recently developed
339 in transformer-based architectures [86]. We observe that the latent feature maps are essentially learning
340 the injection location(s) and direction of flow based on the geologic distributions. Thus, proving that the
341 Stochastic pix2vid model is learning multiphase flow physics and dynamic reservoir behavior appropriately.

342 These results imply that our Stochastic pix2vid is capable of learning the spatiotemporal relationship be-
343 tween the static geologic models and the dynamic reservoir response. Thus, our image-to-video architecture
344 can outperform current image-to-image and encoder-recurrent-decoder architectures to provide improved
345 reservoir behavior prediction closer to that of conventional numerical simulation. To quantify the uncer-
346 tainty in predictions, a comparison of true (d) versus predicted (\hat{d}) response for pressure and saturation
347 distributions for the testing data is shown in Figure 22. The average R^2 over time is approximately 99%
348 with narrow 95% prediction bands that recursively narrow over time. From Figure 22 we observe the advan-
349 tage in implementing recursive refining of predictions over time with recurrent residual connections in the
350 spatiotemporal decoder network, thus reducing the spatiotemporal uncertainty in the predictions.

351 CO₂ saturation and pressure buildup fronts are important quantities for geologic CO₂ storage projects
352 and are often used for regulatory oversight [87, 88], monitoring metrics or history matching purposes [89, 90].
353 The distance between the injection well(s) and the saturation fronts represents the maximum extent of the
354 CO₂ plume; however, these are often very difficult to capture accurately with data-driven proxy models.
355 Our Stochastic pix2vid model shows greater absolute error on and around the plume fronts compared to
356 within the plumes. However, the overall shape and intensity of the pressure and saturation distributions over
357 time is very well captured for all realizations despite being highly heterogeneous. Therefore, the Stochastic
358 pix2vid model can be used as a reliable replacement for expensive numerical reservoir simulations, especially
359 in cases where large number of runs are required to obtain dynamic estimates (e.g., well placement and
360 control optimization, history matching, uncertainty quantification).

361 3.3 Discussion

362 In our Stochastic pix2vid model, the encoder block is composed of separable convolutions, squeeze and
363 excite layers, and instance normalization. These three particular implementations allow for precise param-
364 eterization of the geologic realization into a latent representation, without mixing the effects of Gaussian-
365 distributed properties against binary or binomial-distributed properties. Using recursive residual ConvLSTM

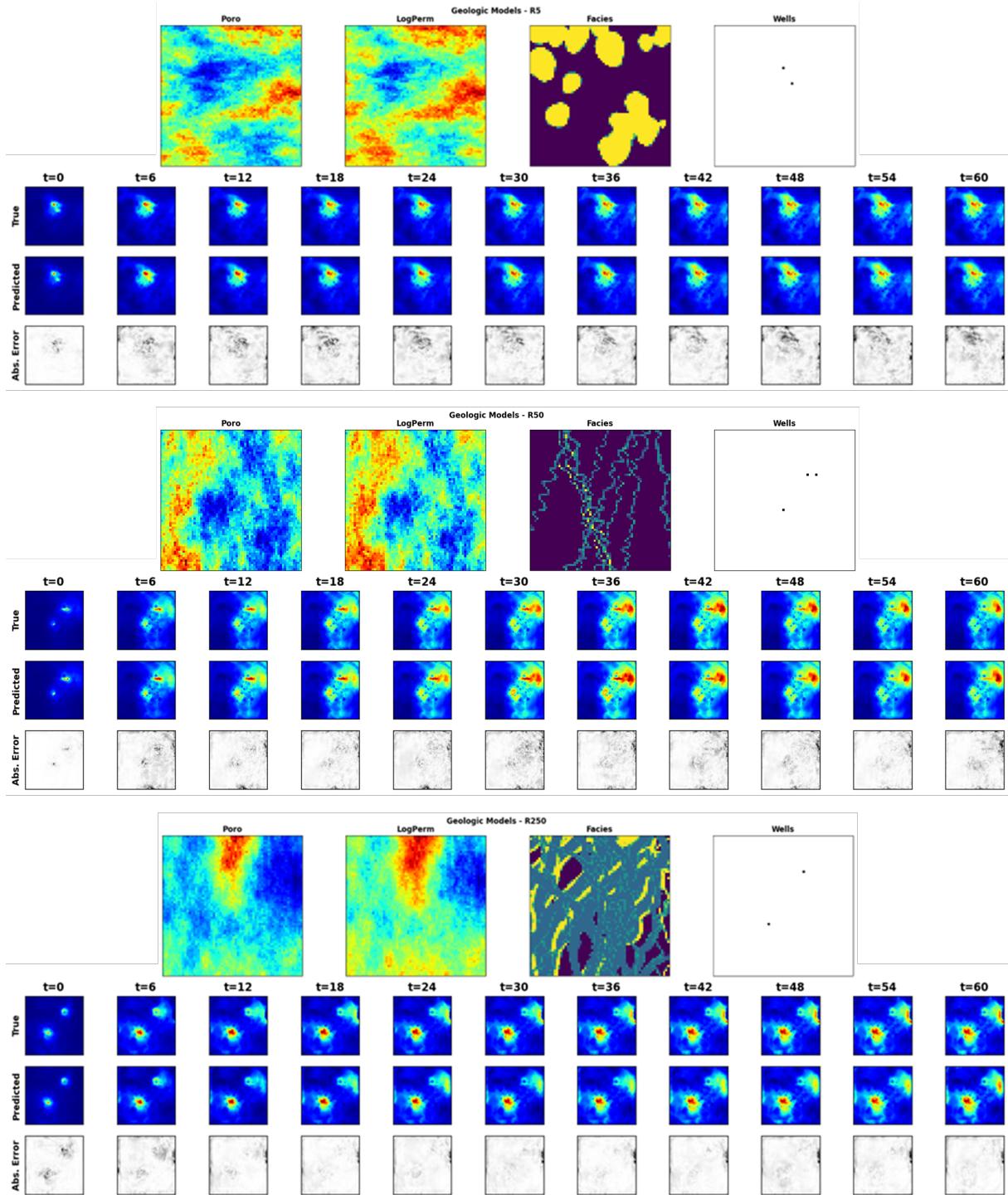


Figure 19: (Normalized) pressure distribution over time for 3 random testing realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.

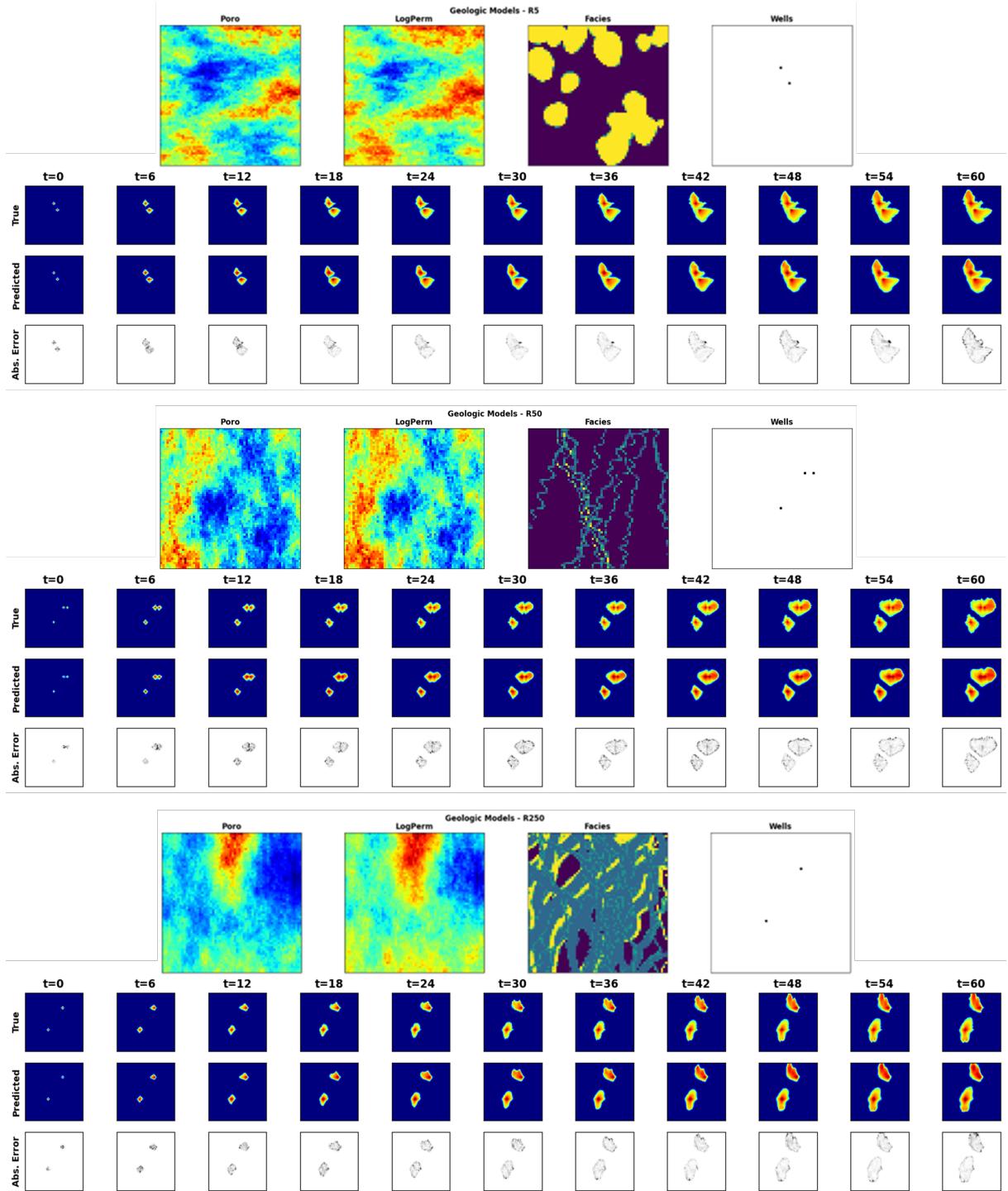


Figure 20: Saturation distribution over time for 3 random testing realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.

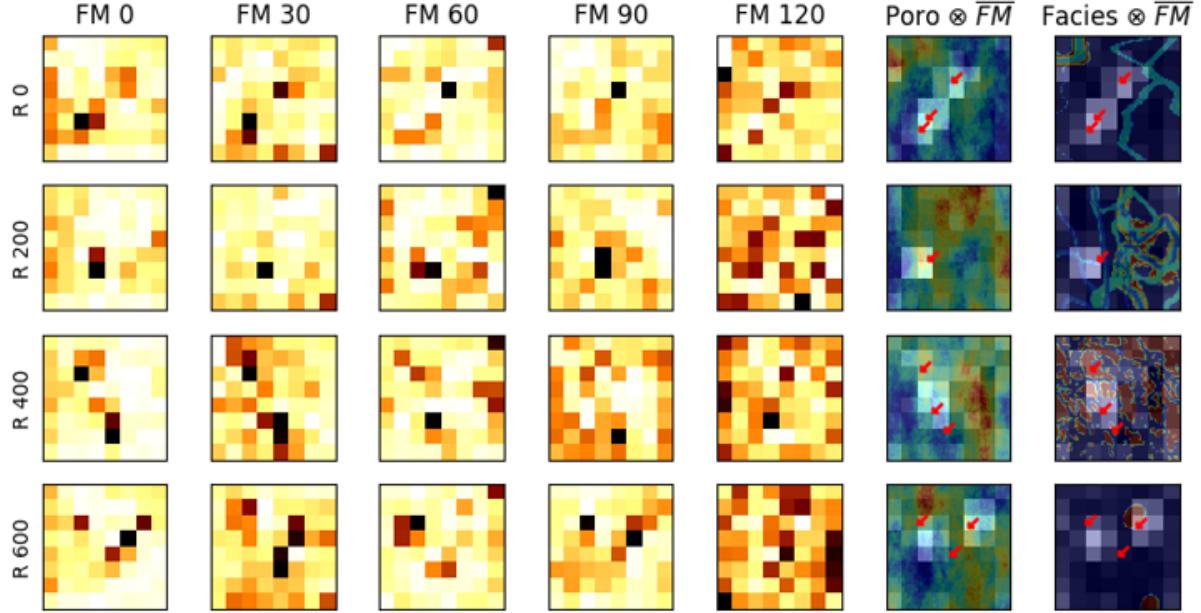


Figure 21: Five random feature maps (FM) of z_m^3 for 4 random realizations. Their average is overlaid on top of the porosity and facies distributions to show the attention mechanism of the encoder. Bright colors represent higher attention and dark colors represent lower attention.

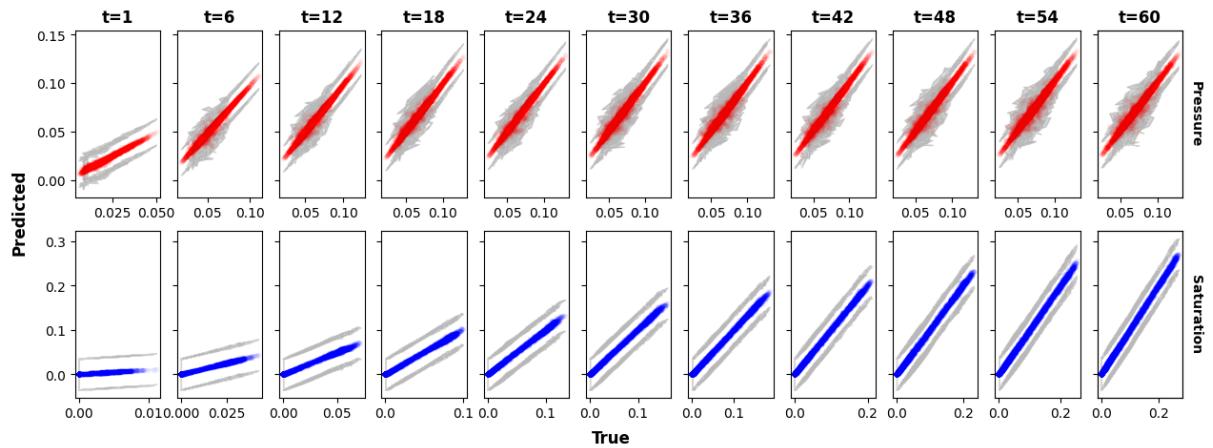


Figure 22: True versus predicted average (normalized) pressure (top) and saturation (bottom) over time for the testing data. The gray portion represents the 95% confidence bands, which become narrower over time.

366 layers, the decoder block iteratively predicts each dynamic state, or video frame, from the concatenation
367 of the previous dynamic latent representation and the intermediate encoding parameterizations. Thus, our
368 architecture makes the proxy model an image-to-video prediction formulation for dynamic reservoir states
369 from a static geologic realization.

370 By using GPU-enabled computations, we significantly accelerate the training and prediction time of the
371 Stochastic pix2vid model. Each HFS run was performed on an Intel ®i9-10900KF processor with 10 cores.
372 The 1,000 realizations are parallelized equally among all cores and the total simulation time accounting for
373 parallelization for all realizations is about 8.33 hours. Dynamic prediction of the 1,000 realizations using
374 the Stochastic pix2vid model on an NVIDIA Quadro M6000 GPU require a total of 4.6 seconds, or 0.001275
375 hours, with an accuracy of 99% and 98% for training and testing, respectively. This provides a sustainable
376 argument for the usage of our Stochastic pix2vid model as a replacement for HFS when computational time
377 is a constraint.

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

384 4 Conclusions

385 We develop a deep learning-based spatiotemporal proxy model to provide flow predictions for a large-scale
386 GCS operation. The key extension introduced is the use of a spatiotemporal convolutional-recurrent archi-
387 tecture for dynamic predictions of CO₂ pressure and saturation distributions over time from a static geologic
388 realization representing the subsurface uncertainty model. The framework is developed as an image-to-video
389 prediction, which is an under-determined estimation problem. Specifically, the implementation expands upon
390 the architectures of current encoder-recurrent-decoder models and provides a fast and accurate proxy as a
391 replacement for physics-based numerical reservoir simulation.

392 The spatiotemporal proxy is applied to a synthetic 2D GCS project with multiple uncertain geologic
393 scenarios and random number and location of injection well(s). A total of 1,000 geologic models are obtained
394 from a variety of possible geologic scenarios including fluvial, turbidite, and deepwater lobe systems. The
395 spatial distribution of porosity, permeability and facies, and the spatial location of the injector well(s) are used
396 as the input data. The proxy model is used to predict the dynamic reservoir response over time, namely the

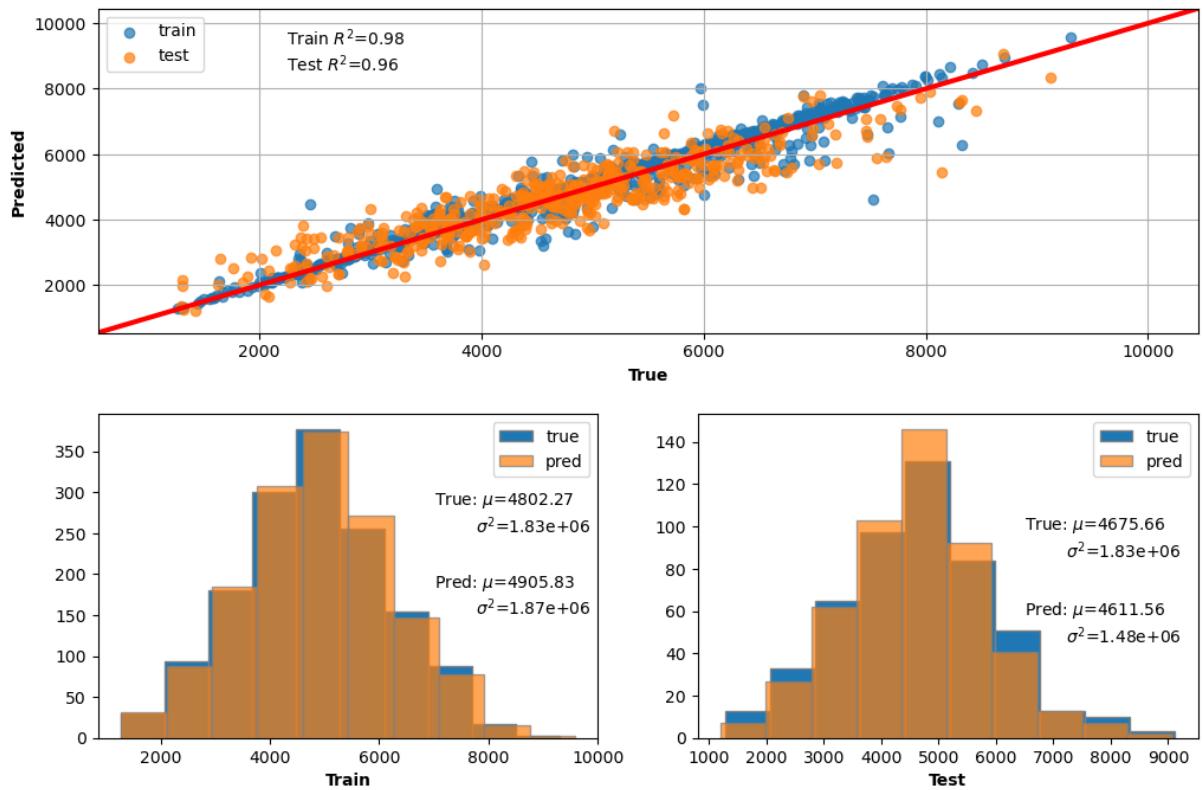


Figure 23: (Top) True vs. predicted cumulative CO₂ volume injected via pixel-wise saturation. (Bottom) True vs. predicted distributions of cumulative CO₂ saturation for training (left) and testing (right).

397 video frames, corresponding to the dynamic CO₂ pressure and saturation distributions, which are obtained
398 offline for training using HFS. The total training time is 88 minutes on a single NVIDIA Quadro M6000
399 GPU, and predictions are obtained with 98-99% accuracy within approximately 4.6 milliseconds, compared
400 to the approximate 30 seconds required for HFS – a 6,500× speedup.

401 There are several possible directions that could be considered for future work. First, an extension to 3D
402 geologic models and their corresponding dynamic predictions is key to scaling up this method for real-world
403 applications. Similarly, although the Stochastic pix2vid proxy model was only trained for GCS prediction, it
404 is applicable for a range of processes such as compositional, geothermal, or conventional oil and gas systems.
405 Moreover, it is possible to extend the Stochastic pix2vid model from a data-driven mapping operator to
406 a PINN by including the discretized form of the governing PDE in the loss function and minimizing the
407 residuals. Another future direction would be to test the performance of the Stochastic pix2vid model on
408 unseen timesteps, either interpolating the training timesteps or extrapolating beyond the training timesteps.
409 Furthermore, the Stochastic pix2vid model can be used as a proxy in workflows for history matching and
410 closed-loop reservoir management.

411 Reproducibility

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

414 Funding

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

416 Declarations

417 The authors declare no conflict of interests.

418 Acknowledgements

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

421 **References**

- 422 [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>.
- 426 [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.
- 429 [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 prospective co₂ storage resource of shales at the national and regional scale. *International Journal of Greenhouse Gas Control*, 51:81–94, 2016. doi: 10.1016/j.ijggc.2016.04.028. cited By 81.
- 433 [4] Bert Metz, Ogunlade Davidson, HC De Coninck, Manuela Loos, and Leo Meyer. *IPCC special report on carbon dioxide capture and storage*. Cambridge: Cambridge University Press, 2005.
- 435 [5] Energy 2020. European commission. In *A strategy for competitive, sustainable and secure energy*, 2010.
- 436 [6] United nations. Agreement, p. *United Nations Treaty Collect*, pages 1–27, 2015.
- 437 [7] S. Bachu. Review of co₂ storage efficiency in deep saline aquifers. *International Journal of Greenhouse Gas Control*, 40:188–202, 2015. doi: 10.1016/j.ijggc.2015.01.007. cited By 277.
- 439 [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 in enhanced oil recovery (eor) operations. *Applied Energy*, 184:337–345, 2016. doi: 10.1016/j.apenergy.2016.09.093.
- 442 [9] N. Castelletto, P. Teatini, G. Gambolati, D. Bossie-Codreanu, O. Vincké, J.-M. Daniel, A. Battistelli, M. Marcolini, F. Donda, and V. Volpi. Multiphysics modeling of co₂ sequestration in a faulted saline formation in italy. *Advances in Water Resources*, 62:570–587, 2013. doi: 10.1016/j.advwatres.2013.04.006. cited By 25.
- 446 [10] Elnara Rustamzade, Wen Pan, John T. Foster, and Michael Pyrcz. Comparison of commingled and sequential production schemes by sensitivity analysis for gulf of mexico paleogene deepwater turbidite oil fields: A simulation study. *Energy Exploration & Exploitation*, 0(0):01445987231195679, 2023. doi: 10.1177/01445987231195679. URL <https://doi.org/10.1177/01445987231195679>.

- 450 [11] K. Rashid, W. Bailey, B. Couët, and D. Wilkinson. An efficient procedure for expensive reservoir-
451 simulation optimization under uncertainty. *SPE Economics and Management*, 5(4):21–33, 2013. doi:
452 10.2118/167261-PA. cited By 16.
- 453 [12] C. Luo, S.-L. Zhang, C. Wang, and Z. Jiang. A metamodel-assisted evolutionary algorithm for expensive
454 optimization. *Journal of Computational and Applied Mathematics*, 236(5):759–764, 2011. doi: 10.1016/
455 j.cam.2011.05.047. cited By 29.
- 456 [13] Javier E. Santos, Bernard Chang, Alex Gigliotti, Eric Guiltinan, Mohamed Mehana, Arvind Mohan,
457 James McClure, Qinjun Kang, Hari Viswanathan, Nicholas Lubbers, Masa Prodanovic, and Michael
458 Pyrcz. Learning from a big dataset of digital rock simulations. In *AGU Fall Meeting Abstracts*, volume
459 2021, pages H25O–1207, December 2021.
- 460 [14] Bailian Chen, Dylan R. Harp, Youzuo Lin, Elizabeth H. Keating, and Rajesh J. Pawar. Geologic co2
461 sequestration monitoring design: A machine learning and uncertainty quantification based approach.
462 *Applied Energy*, 225:332–345, 9 2018. ISSN 03062619. doi: 10.1016/j.apenergy.2018.05.044.
- 463 [15] Wenyue Sun and Louis J. Durlofsky. Data-space approaches for uncertainty quantification of co2
464 plume location in geological carbon storage. *Advances in Water Resources*, 123:234–255, 1 2019. ISSN
465 03091708. doi: 10.1016/j.advwatres.2018.10.028. cited By 23.
- 466 [16] Bailian Chen, Dylan R. Harp, Zhiming Lu, and Rajesh J. Pawar. Reducing uncertainty in geologic
467 co2 sequestration risk assessment by assimilating monitoring data. *International Journal of Greenhouse
468 Gas Control*, 94, 3 2020. ISSN 17505836. doi: 10.1016/j.ijggc.2019.102926.
- 469 [17] B. Li and S.M. Benson. Influence of small-scale heterogeneity on upward co2plume migration in storage
470 aquifers. *Advances in Water Resources*, 83:389–404, 2015. doi: 10.1016/j.advwatres.2015.07.010. cited
471 By 84.
- 472 [18] Su Jiang and Louis J. Durlofsky. Use of multifidelity training data and transfer learning for efficient
473 construction of subsurface flow surrogate models. *Journal of Computational Physics*, 474, 2 2023. ISSN
474 10902716. doi: 10.1016/J.JCP.2022.111800.
- 475 [19] *Best Practices in Automatic Permeability Estimation: Machine-Learning Methods vs. Conventional
476 Petrophysical Models*, volume Day 4 Tue, June 13, 2023 of *SPWLA Annual Logging Symposium*, 06
477 2023. doi: 10.30632/SPWLA-2023-0084.

- 478 [20] H. Wu, N. Lubbers, H.S. Viswanathan, and R.M. Pollyea. A multi-dimensional parametric study of
479 variability in multi-phase flow dynamics during geologic co₂ sequestration accelerated with machine
480 learning. *Applied Energy*, 287, 2021. doi: 10.1016/j.apenergy.2021.116580. cited By 14.
- 481 [21] Siddharth Misra, Yusuf Falola, Polina Churilova, Rui Liu, Chung-Kan Huang, and Jose F. Delgado.
482 Deep learning assisted extremely low-dimensional representation of subsurface earth. *SSRN Electronic
483 Journal*, 8 2022. doi: 10.2139/SSRN.4196705.
- 484 [22] Ademide O. Mabadeje and Michael J. Pyrcz. Rigid transformations for stabilized lower dimensional
485 space to support subsurface uncertainty quantification and interpretation, 2023.
- 486 [23] Mingliang Liu, Dario Grana, and Tapan Mukerji. Randomized tensor decomposition for large-scale
487 data assimilation problems for carbon dioxide sequestration. *Mathematical Geosciences*, 54:1139–1163,
488 5 2022. ISSN 18748953. doi: 10.1007/S11004-022-10005-1/FIGURES/17.
- 489 [24] S.W.A. Canchumuni, A.A. Emerick, and M.A.C. Pacheco. Towards a robust parameterization for
490 conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers and
491 Geosciences*, 128:87–102, 2019. doi: 10.1016/j.cageo.2019.04.006. cited By 80.
- 492 [25] Y. Zhang, P. Vouzis, and N.V. Sahinidis. Gpu simulations for risk assessment in co₂ geologic sequestra-
493 tion. *Computers and Chemical Engineering*, 35(8):1631–1644, 2011. doi: 10.1016/j.compchemeng.2011.
494 03.023. cited By 20.
- 495 [26] Bicheng Yan, Dylan Robert Harp, Bailian Chen, and Rajesh J. Pawar. Improving deep learning per-
496 formance for predicting large-scale geological co₂ sequestration modeling through feature coarsening.
497 *Scientific Reports*, 12:1–12, 11 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-24774-6.
- 498 [27] Zeeshan Tariq, Murtada Saleh Aljawad, Amjad Hasan, Mobeen Murtaza, Emad Mohammed, Ammar El-
499 Husseiny, Sulaiman A Alarifi, Mohamed Mahmoud, and Abdulazeez Abdulraheem. A systematic review
500 of data science and machine learning applications to the oil and gas industry. *Journal of Petroleum
501 Exploration and Production Technology*, pages 1–36, 2021.
- 502 [28] Mohammad Ali Mirza, Mahtab Ghoroori, and Zhangxin Chen. Intelligent petroleum engineering. *En-
503 gineering*, 18:27–32, 2022. ISSN 2095-8099. doi: <https://doi.org/10.1016/j.eng.2022.06.009>.
- 504 [29] Proctor Joshua Brunton, Steve and Nathan Kutz. Discovering governing equations from data by sparse
505 identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the
506 United States of America*, 2016. doi: 10.1073/pnas.1517384113.

- 507 [30] He Xiaolong Fries, William and Youngsoo Choi. Lasdi: Parametric latent space dynamics identification.
 508 *Computer Methods in Applied Mechanics and Engineering*, 2022. doi: 10.1016/j.cma.2022.115436.
- 509 [31] Choi Youngsoo Fries William Belof Jonathan He, Xiaolong and Jiun-Shyan Chen. glasdi: Parametric
 510 physics-informed greedy latent space dynamics identification. *Journal of Computational Physics*, 2023.
- 511 [32] M. Liu and D. Grana. Time-lapse seismic history matching with an iterative ensemble smoother and
 512 deep convolutional autoencoder. *Geophysics*, 85(1):M15–M31, 2020. cited By 2.
- 513 [33] Syamil Mohd Razak, Anyue Jiang, and · Behnam Jafarpour. Latent-space inversion (lsi): a deep learning
 514 framework for inverse mapping of subsurface flow data. *Computational Geoscience*, 26:71–99, 11 2022.
 515 doi: 10.1007/s10596-021-10104-8.
- 516 [34] S. Oladyshkin, H. Class, and W. Nowak. Bayesian updating via bootstrap filtering combined with
 517 data-driven polynomial chaos expansions: Methodology and application to history matching for carbon
 518 dioxide storage in geological formations. *Computational Geosciences*, 17(4):671–687, 2013. doi: 10.
 519 1007/s10596-013-9350-6. cited By 36.
- 520 [35] Anqi Bao, Eduardo Gildin, Abhinav Narasingam, and Joseph S. Kwon. Data-driven model reduction
 521 for coupled flow and geomechanics based on dmd methods. *Fluids*, 4:138, 7 2019. ISSN 2311-5521. doi:
 522 10.3390/FLUIDS4030138.
- 523 [36] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang.
 524 Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
- 525 [37] Liu Yang, Dongkun Zhang, and George Em Karniadakis. Physics-informed generative adversarial net-
 526 works for stochastic differential equations, 2018.
- 527 [38] N. Wang, H. Chang, and D. Zhang. Efficient uncertainty quantification for dynamic subsurface flow with
 528 surrogate by theory-guided neural network. *Computer Methods in Applied Mechanics and Engineering*,
 529 373, 2021. doi: 10.1016/j.cma.2020.113492. cited By 33.
- 530 [39] Emilio Jose Rocha Coutinho, Marcelo Dall'Aqua, and Eduardo Gildin. Physics-aware deep-learning-
 531 based proxy reservoir simulation model equipped with state and well output prediction. *Frontiers in
 532 Applied Mathematics and Statistics*, 7:49, 9 2021. ISSN 22974687. doi: 10.3389/FAMS.2021.651178/
 533 BIBTEX.
- 534 [40] Yinhao Zhu, Nicholas Zabaras, Phaedon-Stelios Koutsourelakis, and Paris Perdikaris. Physics-
 535 constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification with-

- 536 out labeled data. *Journal of Computational Physics*, 394:56–81, oct 2019. doi: 10.1016/j.jcp.2019.05.024.
- 537 URL <https://doi.org/10.1016%2Fj.jcp.2019.05.024>.
- 538 [41] B Yegnanarayana. *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.
- 539 [42] Jeff Heaton. Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016,
540 800 pp, isbn: 0262035618. *Genetic programming and evolvable machines*, 19(1-2):305–307, 2018.
- 541 [43] Yimin Liu and Louis J Durlofsky. 3d cnn-pca: A deep-learning-based parameterization for complex
542 geomodels. *Computers & Geosciences*, 148:104676, 2021.
- 543 [44] Zixiao Yang, Qiyu Chen, Zhesi Cui, Gang Liu, Shaoqun Dong, and Yiping Tian. Automatic recon-
544 struction method of 3d geological models based on deep convolutional generative adversarial networks.
545 *Computational Geosciences*, 26:1135–1150, 2022. doi: 10.1007/s10596-022-10152-8.
- 546 [45] Su Jiang and Louis J Durlofsky. Data-space inversion using a recurrent autoencoder for time-series
547 parameterization. *Computational Geosciences*, 25:411–432, 2021.
- 548 [46] Yanrui Ning, Hossein Kazemi, and Pejman Tahmasebi. A comparative machine learning study for time
549 series oil production forecasting: Arima, lstm, and prophet. *Computers and Geosciences*, 164:105126, 7
550 2022. ISSN 00983004. doi: 10.1016/j.cageo.2022.105126.
- 551 [47] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin
552 transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF
553 international conference on computer vision*, pages 10012–10022, 2021.
- 554 [48] Liuqing Yang, Sergey Fomel, Shoudong Wang, Xiaohong Chen, Wei Chen, Omar M. Saad, and Yangkang
555 Chen. Porosity and permeability prediction using a transformer and periodic long short-term network.
556 *Geophysics*, 88(1):WA293–WA308, 01 2023. ISSN 0016-8033. doi: 10.1190/geo2022-0150.1.
- 557 [49] Eduardo Maldonado Cruz and Michael J Pyrcz. Multi-horizon well performance forecasting with tem-
558 poral fusion transformers. *Available at SSRN 4403939*.
- 559 [50] Wen Pan, Carlos Torres-Verdín, and Michael J. Pyrcz. Stochastic pix2pix: A new machine learning
560 method for geophysical and well conditioning of rule-based channel reservoir models. *Natural Resources
561 Research*, 30:1319–1345, 4 2021. ISSN 15738981. doi: 10.1007/S11053-020-09778-1/FIGURES/24.
- 562 [51] Bogdan Sebacher and Stefan Adrian Toma. Bridging deep convolutional autoencoders and ensemble
563 smoothers for improved estimation of channelized reservoirs. *Mathematical Geosciences*, 54:903–939, 7
564 2022. ISSN 18748953. doi: 10.1007/S11004-022-09997-7/TABLES/3.

- 565 [52] Jichao Bao, Liangping Li, and Arden Davis. Variational autoencoder or generative adversarial networks?
 566 a comparison of two deep learning methods for flow and transport data assimilation. *Mathematical*
 567 *Geosciences*, 54:1017–1042, 8 2022. ISSN 18748953. doi: 10.1007/S11004-022-10003-3/FIGURES/17.
- 568 [53] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmen-
 569 tation. *CoRR*, 2015. cited By 358.
- 570 [54] Eduardo Maldonado-Cruz and Michael J. Pyrcz. Fast evaluation of pressure and saturation predictions
 571 with a deep learning surrogate flow model. *Journal of Petroleum Science and Engineering*, 212:110244,
 572 5 2022. ISSN 0920-4105. doi: 10.1016/J.PETROL.2022.110244.
- 573 [55] Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. U-fno—an
 574 enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water*
 575 *Resources*, 163:104180, 2022. ISSN 0309-1708. doi: <https://doi.org/10.1016/j.advwatres.2022.104180>.
- 576 [56] Gege Wen, Zongyi Li, Qirui Long, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson.
 577 Real-time high-resolution co 2 geological storage prediction using nested fourier neural operators. *Energy*
 578 & *Environmental Science*, 2023. ISSN 1754-5692. doi: 10.1039/d2ee04204e.
- 579 [57] Honggeun Jo, Wen Pan, Javier E Santos, Hyungsik Jung, and Michael J Pyrcz. Machine learning
 580 assisted history matching for a deepwater lobe system. *Journal of Petroleum Science and Engineering*,
 581 207:109086, 2021.
- 582 [58] Feng Zhang, Long Nghiem, and Zhangxin Chen. Evaluating reservoir performance using a transformer
 583 based proxy model. *Geoenergy Science and Engineering*, 226:211644, 2023.
- 584 [59] Daowei Zhang and Heng Li. Efficient surrogate modeling based on improved vision transformer neural
 585 network for history matching. *SPE Journal*, pages 1–17, 2023.
- 586 [60] Yong Do Kim and Louis J. Durlofsky. Convolutional – recurrent neural network proxy for robust
 587 optimization and closed-loop reservoir management. *Computational Geosciences*, pages 1–24, 1 2023.
 588 ISSN 1420-0597. doi: 10.1007/S10596-022-10189-9/TABLES/1.
- 589 [61] Meng Tang, Yimin Liu, and Louis J. Durlofsky. A deep-learning-based surrogate model for data as-
 590 similation in dynamic subsurface flow problems. *Journal of Computational Physics*, 413, 7 2020. ISSN
 591 10902716. doi: 10.1016/J.JCP.2020.109456.
- 592 [62] M. Tang, Y. Liu, and L.J. Durlofsky. Deep-learning-based surrogate flow modeling and geological
 593 parameterization for data assimilation in 3d subsurface flow. *Computer Methods in Applied Mechanics*
 594 and *Engineering*, 376, 2021. doi: 10.1016/j.cma.2020.113636. cited By 39.

- 595 [63] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Generating videos with scene dynamics, 2016.
- 596 [64] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean
597 square error, 2016.
- 598 [65] Ruben Villegas, Jimei Yang, Seunghoon Hong, Xunyu Lin, and Honglak Lee. Decomposing motion and
599 content for natural video sequence prediction, 2018.
- 600 [66] Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. Mocogan: Decomposing motion and
601 content for video generation, 2017.
- 602 [67] Xingjian SHI, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO.
603 Convolutional lstm network: A machine learning approach for precipitation nowcasting. In C. Cortes,
604 N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing*
605 *Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf.
- 606 [68] Michael Iliadis, Leonidas Spinoulas, and Aggelos K. Katsaggelos. Deep fully-connected networks for
607 video compressive sensing, 2017.
- 608 [69] Kai Xu and Fengbo Ren. Csvideonet: A real-time end-to-end learning framework for high-frame-rate
609 video compressive sensing. In *2018 IEEE Winter Conference on Applications of Computer Vision*
610 (*WACV*), pages 1680–1688. IEEE, 2018.
- 611 [70] Michael Dorkenwald, Timo Milbich, Andreas Blattmann, Robin Rombach, Konstantinos G. Derpanis,
612 and Björn Ommer. Stochastic image-to-video synthesis using cinns, 2021.
- 613 [71] Aleksander Holynski, Brian Curless, Steven M. Seitz, and Richard Szeliski. Animating pictures with
614 eulerian motion fields, 2020.
- 615 [72] Karsten Pruess, Curtis M Oldenburg, and GJ Moridis. Tough2 user’s guide version 2. Technical report,
616 Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 1999.
- 617 [73] Nicolas Remy, Alexandre Boucher, and Jianbing Wu. *Applied Geostatistics with SGEMS: A User’s*
618 *Guide*. Cambridge University Press, 2009.
- 619 [74] G. W. Verly. *Sequential Gaussian Cosimulation: A Simulation Method Integrating Several Types of*
620 *Information*, pages 543–554. Springer Netherlands, Dordrecht, 1993. ISBN 978-94-011-1739-5. doi:
621 10.1007/978-94-011-1739-5_42.

- 623 [75] M.J. Pyrcz, J.B. Boisvert, and C.V. Deutsch. A library of training images for fluvial and deepwater
624 reservoirs and associated code. *Computers & Geosciences*, 34(5):542–560, 2008. ISSN 0098-3004. doi:
625 <https://doi.org/10.1016/j.cageo.2007.05.015>.
- 626 [76] Misael M. Morales and Michael Pyrcz. GeostatsGuy/MLTrainingImages: MachineLearningTrainingIm-
627 ages_v1.0.0, March 2023. URL <https://doi.org/10.5281/zenodo.7702128>.
- 628 [77] Knut-Andreas Lie. *An introduction to reservoir simulation using MATLAB/GNU Octave: User guide*
629 for the MATLAB Reservoir Simulation Toolbox (MRST). Cambridge University Press, 2019.
- 630 [78] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the*
631 *IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.
- 632 [79] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference*
633 *on computer vision and pattern recognition*, pages 7132–7141, 2018.
- 634 [80] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient
635 for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.
- 636 [81] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.
- 637 [82] D.S. Oliver, A.C. Reynolds, and N. Liu. *Inverse theory for petroleum reservoir characterization and*
638 *history matching*, volume 9780521881517. 2008. doi: 10.1017/CBO9780511535642. cited By 766.
- 639 [83] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from
640 error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 4 2004.
641 ISSN 1941-0042. doi: doi.org/10.1109/TIP.2003.819861.
- 642 [84] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint*
643 *arXiv:1711.05101*, 2017.
- 644 [85] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint*
645 *arXiv:1412.6980*, 2014.
- 646 [86] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz
647 Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing*
648 *systems*, 30, 2017.
- 649 [87] Q. Li and G. Liu. *Risk assessment of the geological storage of CO2: A review*. 2016. doi: 10.1007/
650 978-3-319-27019-7-13. cited By 39.

- 651 [88] R.A. Chadwick, R. Arts, and O. Eiken. 4d seismic quantification of a growing co₂ plume at sleipner,
652 north sea. *Petroleum Geology Conference Proceedings*, 6(0):1385–1399, 2005. doi: 10.1144/0061385.
653 cited By 188.
- 654 [89] R.A. Chadwick and D.J. Noy. History-matching flow simulations and timelapse seismic data from the
655 sleipner co₂ plume. *7th Petroleum Geology Conference [FROM MATURE BASINS to NEW FRON-*
656 *TIERS] (London, 3/30/2009-4/2/2009) Proceedings*, 2:1171–1182, 2010. cited By 31.
- 657 [90] Ismael Dawuda and Sanjay Srinivasan. Geologic modeling and ensemble-based history matching for
658 evaluating co₂ sequestration potential in point bar reservoirs. *Frontiers in Energy Research*, 10:867083,
659 2022.