

# Stochastic pix2vid: A new spatiotemporal deep learning method for image-to-video synthesis in geologic CO<sub>2</sub> storage prediction

Misael M. Morales<sup>1\*</sup>, Carlos Torres-Verdín<sup>1,2</sup> and Michael J. Pyrcz<sup>1,2</sup>

<sup>1</sup>Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, TX, USA.

<sup>2</sup>Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, USA.

\*Corresponding author(s). E-mail(s): [misaelmorales@utexas.edu](mailto:misaelmorales@utexas.edu);

## Abstract

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

**Keywords:** Image-to-video synthesis, Spatiotemporal prediction, Convolutional neural network, Recurrent neural network, Proxy model

## <sup>1</sup> 1 Introduction

<sup>2</sup> Geologic CO<sub>2</sub> sequestration (GCS) has emerged  
<sup>3</sup> as a potential technology solution to reduce  
<sup>4</sup> anthropogenic greenhouse gas emissions to the  
<sup>5</sup> atmosphere [1–3], and has become increasingly

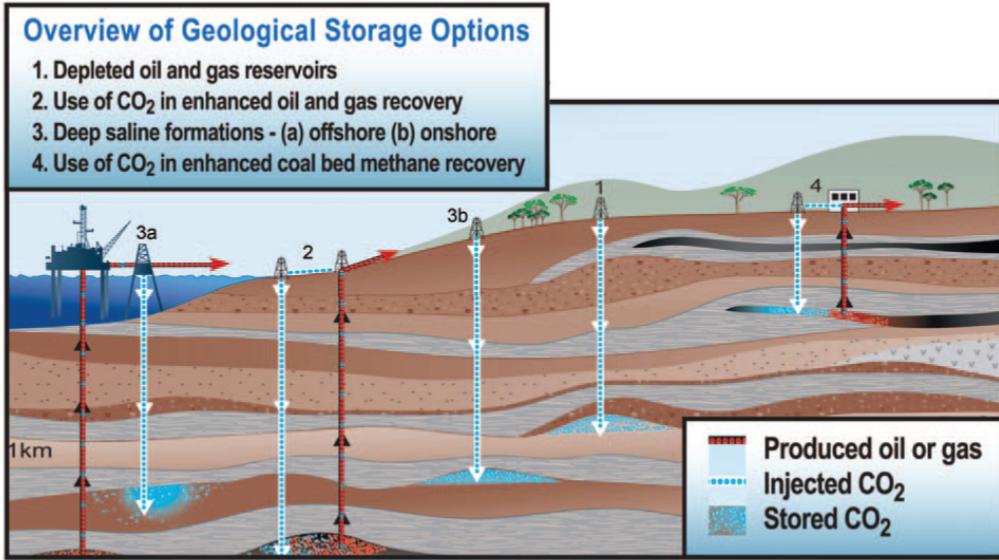
<sup>6</sup> popular worldwide due to the need to meet  
<sup>7</sup> international climate protection agreements [4–6].  
<sup>8</sup> Modeling injected CO<sub>2</sub> movement in the subsur-  
<sup>9</sup> face over and beyond the life of the project is  
<sup>10</sup> a critical component to support optimum GCS

11 project decision making for safe and secure CO<sub>2</sub>  
12 sequestration. A schematic of typical GCS oper-  
13 ations is shown in Figure 1, including storage  
14 in depleted oil and gas reservoir and deep saline  
15 formations, and CO<sub>2</sub> enhanced oil and coalbed  
16 methane recovery [7–9]. However, there are several  
17 technical challenges associated with the subsur-  
18 face modeling to support GCS operations. To  
19 accurately forecast and monitor subsurface multi-  
20 phase flow, physics-based high-fidelity numerical  
21 simulations are required. These numerical simu-  
22 lations are computationally intensive and time-  
23 consuming since they require iterative solutions  
24 of nonlinear systems of equations applied over  
25 large volumes of the subsurface at sufficient resolu-  
26 tion to represent heterogeneity [10–13]. Also, due  
27 to the large degree of uncertainty in subsurface  
28 data, and the spatial distribution of the proper-  
29 ties of heterogeneous porous media between the  
30 sparsely sampled data, GCS operations require a  
31 robust probabilistic-based uncertainty assessment  
32 for improved engineering decision-making [14–16].  
33 In order to capture the fine-scale multiphase flow  
34 behavior given an uncertain spatial distribution of  
35 subsurface properties, a large number of numeri-  
36 cal simulations are required, leading to very high  
37 computational costs and delayed feedback unable  
38 to support timely decision making [17, 18].

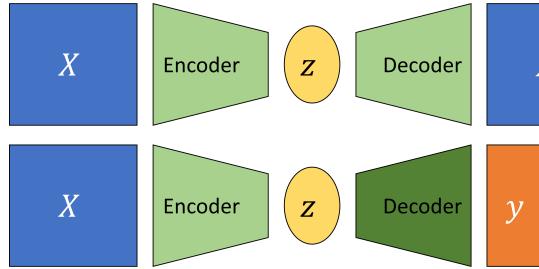
39 To overcome this, machine learning techniques  
40 have emerged as candidate proxy models due  
41 to their ability to perform dimensionality reduc-  
42 tion for efficient problem parameterization and  
43 model complicated systems to calculate fast pre-  
44 dictions of subsurface flow and transport behavior  
45 for real-time feedback on the impact of geologi-  
46 cal and engineering controls on CO<sub>2</sub> behavior in  
47 the subsurface over time [19–21]. Dimensionality  
48 reduction techniques are supervised or unsuper-  
49 vised machine learning methods that compress  
50 (or encode) the data,  $X$ , into a lower-dimensional  
51 latent feature representation,  $z$ , and decompress  
52 (or decode) the latent representation either: (1)  
53 back to the original data space,  $\hat{X}$  (unsupervised,  
54 AutoEncoder), or (2) to a new response feature  
55 space,  $y$  (supervised, Encoder-Decoder) [22–24],  
56 as shown in Figure 2. The recent advancements in  
57 deep learning algorithms and in computing archi-  
58 tecture and power, enable GPU-enabled neural  
59 network models that have accelerated the fields of  
60 forward and inverse modeling [25, 26]. Classical  
61 statistical modeling methods are often hindered

62 by the size of the models and their conditioning to  
63 big data, i.e., that is data with volume, velocity,  
64 variety, value, and veracity [27, 28], and fail to gen-  
65 eralize beyond fit-for-purpose frameworks [29, 30].  
66 By analyzing big data sets, machine learning  
67 techniques can uncover complex patterns and rela-  
68 tionships in lower-dimensional, latent feature rep-  
69 resentations that may not be discernible through  
70 traditional statistical and geostatistical methods  
71 [31–33]. When combined with a latent space mod-  
72 eling framework, machine learning approaches effi-  
73 ciently and accurately exploit hidden patterns and  
74 features in the data, remove redundancies or noise,  
75 and decrease the mathematical and computational  
76 complexity of the problem significantly [34, 35].

77 Supervised machine learning approaches  
78 applied to the subsurface are divided into two  
79 main categories, namely purely data-driven  
80 models or physics-informed models. Data-driven  
81 proxy models are neural network architectures  
82 trained with labeled data that produce a map-  
83 ping from input predictor feature to output  
84 response features [36, 37]. On the other hand,  
85 the training process to match training data for  
86 PINNs is regularized with the minimization of the  
87 (physical) loss from the residual of the governing  
88 partial differential equations (PDEs) along with  
89 the losses associated with the initial and bound-  
90 ary conditions [38, 39]. However, other variants  
91 of PINNs such as physics-guided or physics-  
92 constrained neural networks where the PDE loss  
93 is not embedded in the training step, instead the  
94 models have specific architectures or parameters  
95 to mimic the physics in the system, have proven  
96 useful for subsurface energy resource engineering  
97 applications [40–42]. One disadvantage of machine  
98 learning techniques is that they require significant  
99 amounts of training data, but once trained these  
100 prediction models suffer from lack of generaliza-  
101 tion, i.e., inability to provide accurate predictions  
102 away from the training data beyond which they  
103 have been specifically trained [43, 44]. For both  
104 data-driven and physics-informed approaches,  
105 typically, spatial relationships are modeled  
106 through convolutional neural networks (CNNs)  
107 [45, 46] and the temporal relationships through  
108 recurrent neural networks (RNNs) [47, 48], but  
109 recent advancements in transformer-based archi-  
110 tectures improve performance compared to the



**Fig. 1** Types of geologic CO<sub>2</sub> storage operations and the geologic formations that can be used for sequestration. Modified from the Carbon Dioxide Cooperative Research Center (CO<sub>2</sub>CRC), <http://www.co2crc.com.au/about/co2crc>



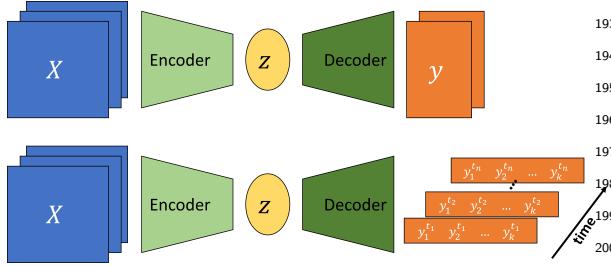
**Fig. 2** Dimensionality reduction model structures. Unsupervised AutoEncoder structure (top), and supervised Encoder-Decoder structure (bottom).

111 CNN and RNN methods for spatial and temporal 140  
112 latent feature representations [49–51]. 141

113 A number of machine learning-based proxy 142  
114 models have been developed to estimate the reser- 143  
115 voir behavior in subsurface energy resource appli- 144  
116 cations. Most techniques rely on the concept of 145  
117 image translation, or pix2pix, where a target 146  
118 image(s) is predicted from an input image(s) 147  
119 [52– 148  
120 55], as shown in Figure 3. Maldonado-Cruz and 149  
121 Pyrcz [56] develop a convolutional U-Net model 150  
122 to predict pressure and saturation states given 151  
123 an uncertain geologic realization. This work is 152  
124 an example of image-to-image static forecasting, 153  
125 where the time state is given as an input, and 154  
126 the proxy model will predict a single response 155  
state of pressure and saturation at the given time.

127 Wen et al. [57] develop a Fourier Neural Operator 128 (FNO) architecture to predict image-to-image 129 response states of pressure and saturation from 130 an uncertain geologic realization and is further 131 extended for multi-scale and nested domains [58]. 132 These methods are based on a pix2pix, or image- 133 to-image prediction, where a specific timestep is 134 used as an input feature to predict the rela- 135 tionship between the geologic model and the 136 reservoir response at that specific timestep. 137 This implies that pix2pix or image-to-image methods 138 are formulated as an even-determined or some- 139 times over-determined estimation problem, where 140 the number of input features is equal to or greater 141 than the number of output features. Moreover, 142 numerous other proxy models have been devel- 143 oped for subsurface applications using more 144 complex architectures such as generative adversarial 145 networks (GANs) [59] and transformers [60, 61]. 146 Despite showing consistent results and signifi- 147 cant speedups compared to traditional numerical 148 simulation, pix2pix models do not capture the spa- 149 tiotemporal relationships and dynamic response of 150 the subsurface system.

151 Moving beyond image-to-image predictions, 152 Kim and Durlofsky [62] develop a convolutional- 153 recurrent proxy for pix2time, or image-to- 154 timeseries, forecasting and discuss its advantages



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

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 hydrocarbon 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 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 [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 between image and video domains using a dynamic latent representation. The cINN architecture allows for video-to-image and image-to-video predictions, demonstrating possible the generation of video frames from a static input image. Finally, Holynski et al. [73] implemented the idea of Eulerian motion fields to define the moving portions of the image 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 are the foundation for our image-to-video proxy model.

We propose a novel image-to-video spatiotemporal proxy model, Stochastic pix2vid, for the prediction of dynamic reservoir behavior over time from a subsurface uncertainty model suite of static geologic realizations. Our model exploits the spatial and temporal structures in latent space to dynamically reconstruct the time-dependent pressure and multiphase saturation states from a static geologic realization. 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

241 stochastic pix2vid model has significant advantages compared to image-to-image and encoder-decoder models in terms of computational efficiency and prediction accuracy and can be used as a replacement for physics-based numerical reservoir simulations, or high-fidelity simulations (HFS), in GCS projects as an image-to-video mapping operator.

249 In the methodology section, we describe the governing equations of multiphase flow in GCS, 250 and the proposed spatiotemporal proxy model 251 architecture. In the results and discussion sections, 252 we describe the geologic modeling and numerical 253 reservoir simulation steps required to generate 254 the training data, and evaluate the training and 255 performance of the proposed proxy model and 256 compare its efficiency and accuracy to high-fidelity 257 numerical simulations using a 2D synthetic case 258 for large-scale GCS operations.

## 260 2 Methodology

261 This section describes the governing equations, 262 and the architecture and training strategy of the 263 Stochastic pix2vid model.

### 264 2.1 Governing equations

265 For the CO<sub>2</sub>-water multiphase flow problem, the 266 general form of the mass accumulation for component 267 κ = CO<sub>2</sub> or water is given by [74]:

$$278 \frac{\partial M^k}{\partial t} = -\nabla \bullet F^\kappa + q^\kappa. \quad (1)$$

279 For each component κ, the mass accumulation 280 term  $M^\kappa$  is summed over all phases  $p$ ,

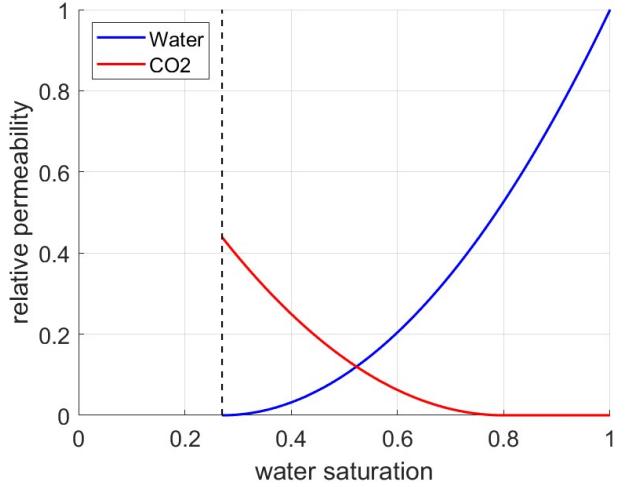
$$281 M^k = \phi \sum_p S_p \rho_p X_p^\kappa \quad (2)$$

282 where  $\phi$  is the porosity,  $S_p$  is the saturation of 283 phase  $p$ ,  $\rho_p$  is the density of phase  $p$ , and  $X_p^\kappa$  is the 284 mass fraction of component  $\kappa$  present in phase  $p$ .

285 For each component  $\kappa$ , there is also the advective 286 mass flux  $F^\kappa|_{adv}$  obtained by summing over all 287 phases  $p$ ,

$$288 F^\kappa|_{adv} = \sum_p X_p^\kappa F_p \quad (3)$$

289 where each individual phase flux  $F_p$  is given by 290 Darcy's equation:



291 **Fig. 4** Relative permeability curves for the CO<sub>2</sub>-water 292 system. The residual saturations are 0.27 and 0.2 for water 293 and CO<sub>2</sub>, respectively.

$$294 F_p = \rho_p u_p = -k \frac{k_{r,p} \rho_p}{\mu_p} (\nabla P_p - \rho_p g) \quad (4)$$

295 where  $u_p$  is the Darcy velocity of phase  $p$ , 296  $k$  is the absolute permeability,  $k_{r,p}$  is the relative 297 permeability of phase  $p$ ,  $\mu_p$  is the viscosity 298 of phase  $p$ , and  $g$  is the gravitational acceleration 299 constant. The relative permeability curves for the 300 CO<sub>2</sub>-water system are shown in Figure 4. The fluid 301 pressure of phase  $p$ ,

$$302 P_p = P + P_c \quad (5)$$

303 is given by the sum of the reference phase 304 pressure  $P$  and the capillary pressure  $P_c$ . The 305 numerical simulation does not include molecular 306 diffusion or hydrodynamic dispersion effects for 307 practical purposes.

### 308 2.2 Proxy Model Architecture

309 Our proposed Stochastic pix2vid image-to-video 310 data-driven method, is mapping operator from 311 the static realizations of geologic distributions 312 of porosity, permeability and facies as well as 313 the injector well(s) distribution, to the dynamic 314 responses of pressure and saturation distributions 315 over time.

316 Let  $m$  represent a geologic model realization of 317 porosity, permeability, facies, and injector well(s)

300 distributions, such that  $m = \{\phi, k, facies, w\}$ .  
 301 The dynamic responses of pressure and satura-  
 302 tion over time are given by  $d = f(m)$ , such that  
 303  $d = \{P(t), S(t)\}$  and  $f$  is the physics-based numer-  
 304 ical reservoir simulation. Our aim is to replace  $f$   
 305 with a more efficient data-driven proxy by train-  
 306 ing the Stochastic pix2vid model, which is trained  
 307 as a single model to predict both pressure and sat-  
 308 uration distributions over time as a multi-channel  
 309 output from the multi-channel input features,  $m$ .  
 310 For this purpose, we exploit the latent struc-  
 311 tures in space and time of the static inputs  
 312 and dynamic outputs through a spatiotemporal  
 313 encoder-decoder architecture.  
 314

315 The encoder portion of the network is com-  
 316 prised of sequential convolutional layers to com-  
 317 press the spatial features of the subsurface real-  
 318 izations into a latent parameterization  $z_m$ , given  
 319 by  $z_m = Enc(m)$ . In their compressed rep-  
 320 resentation, these features represent the salient  
 321 characteristics of the geologic distributions. The  
 322 decoder portion of the network is designed as a  
 323 series of recursive residual convolutional-recurrent  
 324 layers, such that the latent space  $z_m$  is recursively  
 325 decoded into the dynamic distributions of pressure  
 326 and saturation over time. The previous timestep  
 327 latent representations,  $z_d^t$ , are used in the sub-  
 328 sequent timesteps of the decoder, such that the  
 329 subsequent timesteps will predict the current and  
 330 previous timestep(s) jointly and iteratively, pro-  
 331 viding a reduction of systematic error in time as  
 332 subsequent frames of the dynamic output video  
 333 are predicted. The full architecture is represented  
 334 as:  
 335

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

336 The encoder,  $Enc(\cdot)$ , compresses the geologic  
 337 realizations,  $m$ , into a latent representation  $z_m$   
 338 through the use of depthwise separable convo-  
 339 lutions [75]. This type of convolution learns the  
 340 parameters for each channel in the image sep-  
 341 arately, avoiding mixing of variables or loss of  
 342 resolution, as shown in Figure 5. This is especially  
 343 important when dealing with discrete, non-smooth  
 344 porosity and permeability spatial distributions  
 345 due to discrete facies and binary well(s) location  
 346 distributions. Each separable convolution layer is  
 347 regularized with an  $l_1$ -norm weight of  $1 \times 10^{-6}$ .  
 348 Moreover, we use a Squeeze-and-Excite layer to  
 349 improve channel interdependence, and to avoid  
 350

351 mixing and loss of resolution [76]. Each Squeeze-  
 352 and-Excite layer will provide the optimal network  
 353 weights for each channel independent of the other  
 354 channels by passing the feature maps through a  
 355 global pooling layer (squeeze) and a dense layer  
 356 with nonlinear activation (excite), to add content  
 357 aware mechanism for re-weighting each channel  
 358 adaptively, as shown in Figure 6. Furthermore,  
 359 by applying instance normalization, as opposed  
 360 to the more common batch normalization, we  
 361 achieve channel-independent normalization of the  
 362 convolved features [77]. Instance normalization is  
 363 a special case of group normalization, where the  
 364 numbers of channels per group is exactly 1, such  
 365 that each channels gets its own normalization  
 366 scheme, as shown in Figure 7. Parametric recti-  
 367 fied linear units (PReLU) is used as the activation  
 368 function, where at each minibatch iteration, the  
 369 network learns the optimal leaky slope for activa-  
 370 tion in each layer, as shown in Figure 8. Finally,  
 371 pooling and spatial dropout are applied to reduce  
 372 in half the input dimension of each layer and to  
 373 provide a means of spatial regularization, respec-  
 374 tively. Through 3 convolutional encoding layers  
 375 with filter size  $3 \times 3$ , we obtain the latent param-  
 376 eterizations  $z_m^1$ ,  $z_m^2$ , and  $z_m^3$ . Table 1 summarizes  
 377 the architecture and dimensions of each layer.  
 378

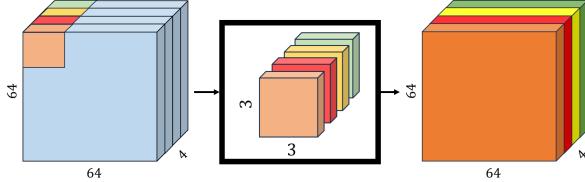
Step 1: **Depthwise Separable encoding:** The  
 379 first layer of  $Enc$  takes the geologic  
 380 model realization,  $m$ , and computes the  
 381 depthwise separable convolutional fea-  
 382 tures channel-by-channel.  
 383

Step 2: **Squeeze-and-Excite encoding:** By  
 384 taking the channel-wise global average  
 385 of the feature space from Step 1, a  
 386 fully-connected layer predicts the appro-  
 387 priate weighting coefficients to best  
 388 parameterize the features.  
 389

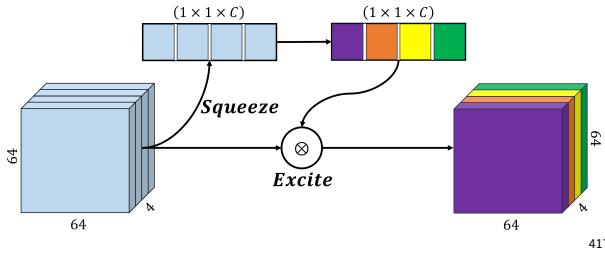
Step 3: **Instance Normalization of the fea-  
 390 ture space:** Feature normalization is  
 391 applied on a channel-by-channels basis for  
 392 each batch of the encoded feature space,  
 393 avoiding mixing and blurring.  
 394

Step 4: **Activation, Pooling, and Spatial  
 395 Dropout:** The PReLU nonlinear activa-  
 396 tion function is used, and for each batch,  
 397 an optimal leaky slope is learned. Pooling  
 398 is used to reduce the feature space in half,  
 399 and Spatial Dropout of 5% is used to reg-  
 400 ularize the learning process and increase  
 401 robustness in prediction.  
 402

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

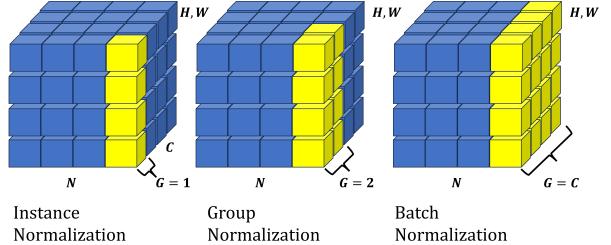


**Fig. 5** Schematic for a separable convolutional layer. Each channel is convolved with its own set of convolutional filters to obtain the best representation, as opposed to traditional convolutions where the same filter is applied to all channels in the data.

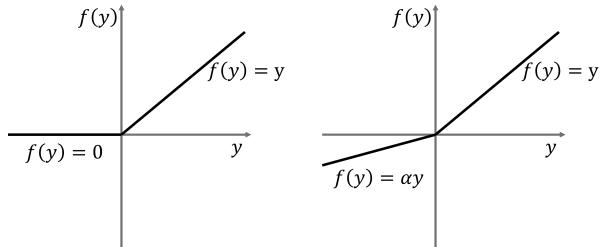


**Fig. 6** Schematic for a squeeze-and-excite layer. The "squeeze" layer takes the global average of the data for each channel, and the "excite" layer is a fully-connected layer with nonlinear activation to estimate the optimal weights for each channel in the data. The result is a weighted representation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

405 The decoder,  $Dec^t(\cdot)$ , of the Stochastic pix2vid  
406 model extracts the spatiotemporal relationships  
407 from the latent representations of  $m$  to reconstruct  
408 the dynamic pressure and saturation distributions  
409 over time,  $d$ . To accurately reconstruct the spa-  
410 tiotemporal structure from the static latent space,  
411  $z_m$ , we employ a series of convolutional-recurrent  
412 layers, namely a convolutional long-short term  
413 memory layer (ConvLSTM). The general form  
414 of a 2D ConvLSTM layer is shown in Figure  
415 9. Through 3 convolutional-recurrent layers, we  
416 obtain the dynamic prediction of  $\hat{d}$  as follows:



**Fig. 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).



**Fig. 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,  $\alpha$ , is learned for each batch.

418 Step 6: **Spatiotemporal decoding of  $z_m^3$ :** The  
419 first ConvLSTM layer takes the smallest  
420 latent representation,  $z_m^3$ , and recon-  
421 structs the first decoded timestep  $z_d^3$ .

422 Step 7: **Residual concatenation of  $z_m^2$ :** The  
423 first decoded timestep,  $z_d^3$ , is concatenated  
424 with the intermediate static encoding  $z_m^2$   
425 to retain multi-scale features and improve  
426 prediction performance and resolution.

427 Step 8: **Intermediate spatiotemporal decod-  
428 ing:** The second ConvLSTM layer takes  
429 the residual concatenation of the interme-  
430 diate latent representations,  $[z_m^2, z_d^3]$ , to  
431 predict the next spatiotemporal represen-  
432 tation  $z_d^2$ .

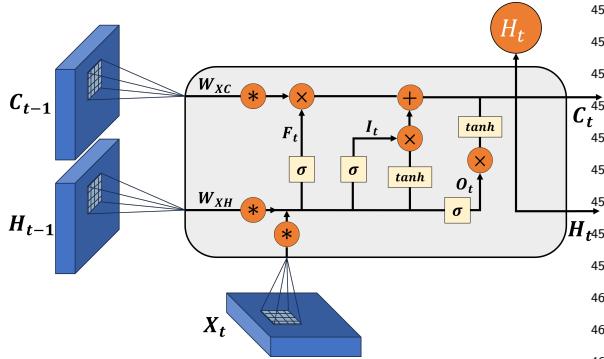
433 Step 9: **Residual concatenation of  $z_m^1$ :** The  
434 intermediate decoded timestep,  $z_d^2$ , is con-  
435 catenated with the largest static encoding  
436  $z_m^1$ .

437 Step 10: **Final spatiotemporal decoding:** The  
438 third ConvLSTM layer takes the resi-  
439 dual concatenation of the larger latent

**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		$8 \times 8 \times 256 (z_m^3)$

representations,  $[z_m^1, z_d^2]$ , to predict the full-scale dynamic output,  $d$ .



**Fig. 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.

To enhance the performance of the spatiotemporal decoding, each ConvLSTM layer is followed by a batch normalization, activation, and a transpose convolutional layer, the latter for downscaling the latent features to twice their dimension. Spatial dropout is applied, and the concatenated features are once more convolved and activated

to obtain the layer prediction. Table 2 shows the architecture of the decoder network.

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

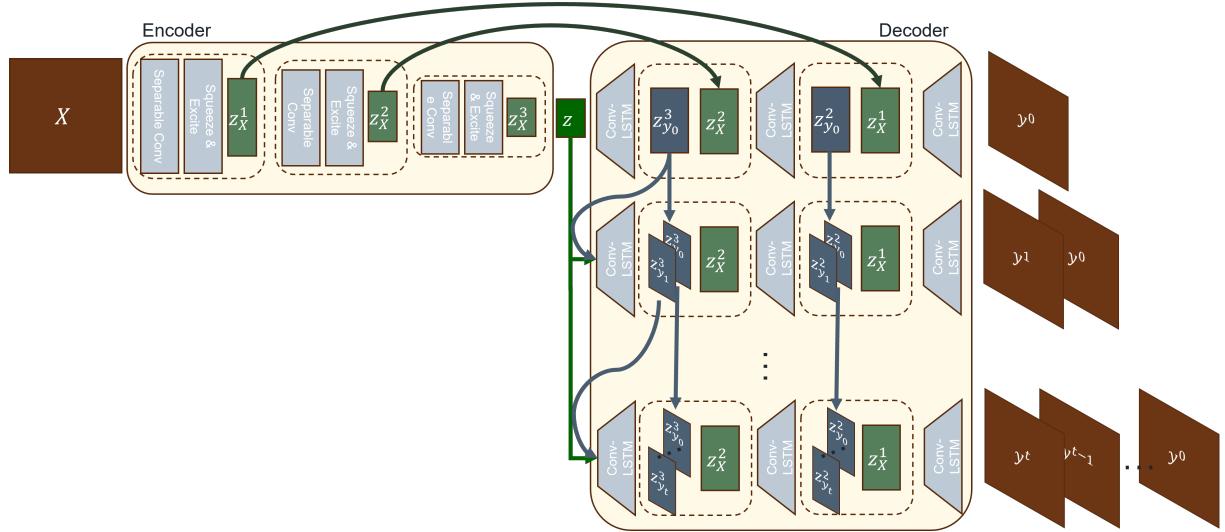
The complete Stochastic pix2vid architecture is shown in Figure 10. Here we observe the spatial compression of the geologic models,  $m$ , through the encoding portion of the network, and the spatiotemporal decoding and residual multi-scale concatenations through the decoder portion of the network. The resulting architecture provides proxy model from a subsurface static uncertainty model (images) to subsurface dynamic response (videos).

### 2.3 Training Strategy

The inputs to the Stochastic pix2vid are the geologic realizations, comprised of the distributions of

**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		
4	Spatial Dropout		
	Concatenate ( $z_m^1$ )		
	Conv2D + Sigmoid		$t \times 64 \times 64 \times 2 (z_{d_t}^1)$
	ConvLSTM2D		
	BatchNorm + LeakyReLU		



**Fig. 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,  $y \equiv d$ .

porosity, permeability, facies, and injection well(s) location, represented as a matrix  $m$  of dimensions  $64 \times 64 \times 4$ . The outputs are the results from the numerical reservoir simulation, namely pressure and saturation distributions over time, represented as a matrix  $d$  of dimensions  $64 \times 64 \times 60 \times 2$ . This yields an ill-posed and under-determined estimation problem, which are difficult

485 to resolve [78, 79]. To improve the training effi-  
 486 ciency and performance, we subsample in time  
 487 from 60 timesteps to 11. In other words, instead of  
 488 monthly monitoring, we predict the dynamic out-  
 489 puts at the initial step and every 6 months after-  
 490 ward; therefore the output matrices,  $(d, \hat{d})$ , have a  
 491 final dimension of  $64 \times 64 \times 11 \times 2$ . This is done to  
 492 make the problem more tractable and speed up the  
 493 training and prediction process, while retaining  
 494 majority of the temporal information.

495 We also perform min-max normalization so  
 496 that the input and output features are in the  
 497 range of  $[0, 1]$ , which greatly improves the per-  
 498 formance of the nonlinear activation functions.  
 499 Furthermore, we perform data augmentation by  
 500  $90^\circ$  image rotation, making the network agnostic  
 501 to orientation and encourage effectively learning  
 502 the flow physics in the system rather than mem-  
 503 orizing spatial distribution patterns. The total  
 504 amount of training data is therefore 2,000 real-  
 505 alizations (after augmentation), which is split into  
 506 1,500 realizations for training and 500 realizations  
 507 for testing. To improve model generalizability, at  
 508 each epoch, each training set minibatch is further  
 509 split into a training and validation subset using  
 510 an 80/20 split. The validation set is only used to  
 511 adjust the trainable model parameters for each  
 512 batch at each epoch and is randomly partitioned  
 513 from the training batch at every epoch, while the  
 514 testing data remains unseen to quantify the model  
 515 performance after training.

516 A custom three-part loss function is used to  
 517 accurately predict pixel-wise and perceptual infor-  
 518 mation in the predictions. The mean squared error  
 519 (MSE) is used to reconstruct the pixel-wise inten-  
 520 sity values, while the mean absolute error (MAE)  
 521 is used to optimize for the pressure and satura-  
 522 tion plume edges. The third part is the structural  
 523 similarity index metric (SSIM), which provides  
 524 a perceptual image-to-image comparison of lumi-  
 525 nance, contrast, and structure [80]. For optimal  
 526 training, the aim is to minimize the MSE and  
 527 MAE while maximizing the SSIM for the true  
 528 versus predicted outputs,  $d$  and  $\hat{d}$ , such that the  
 529 total loss is given by  $\mathcal{L} = \alpha(1 - SSIM) + (1 -$   
 530  $\alpha)[\beta MSE + (1 - \beta)MAE]$ , where  $\alpha$  and  $\beta$  are  
 531 weighting coefficients obtained empirically as 0.33  
 532 and 0.66, respectively.

533 The model is trained using the AdamW opti-  
 534 mizer [81]. This variant of the well-known adaptive

535 momentum (Adam) optimizer [82] includes an  
 536 added method to decay weights for the adap-  
 537 tive estimation of first-order and second-order  
 538 moments. We implement a learning rate of  $1 \times 10^{-3}$   
 539 with a weight decay term of  $1 \times 10^{-5}$ .

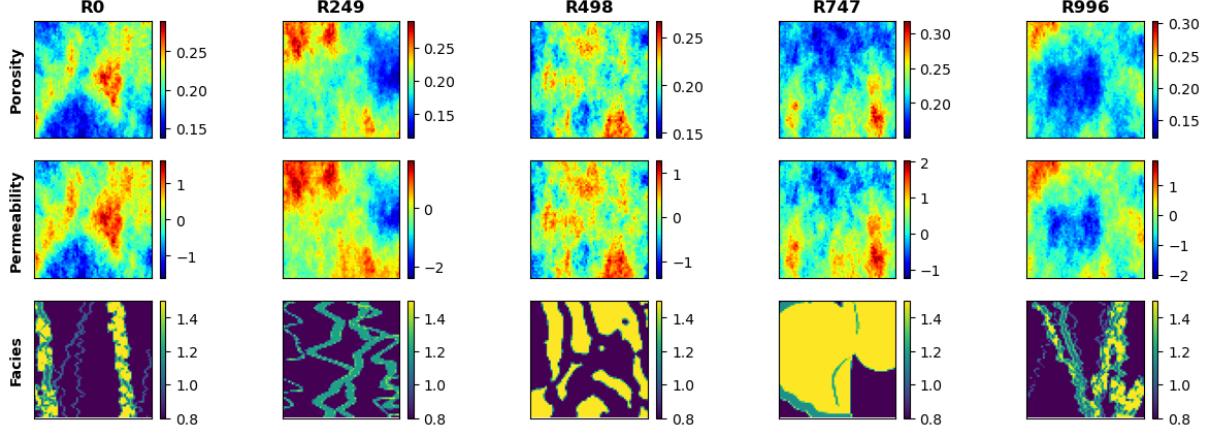
## 3 Results

This section describes the geologic model genera-  
 541 tion, training performance and discusses the appli-  
 542 cation of the Stochastic pix2vid proxy to rapidly  
 543 forecast CO<sub>2</sub> plume migration for a large-scale  
 544 GCS operation.

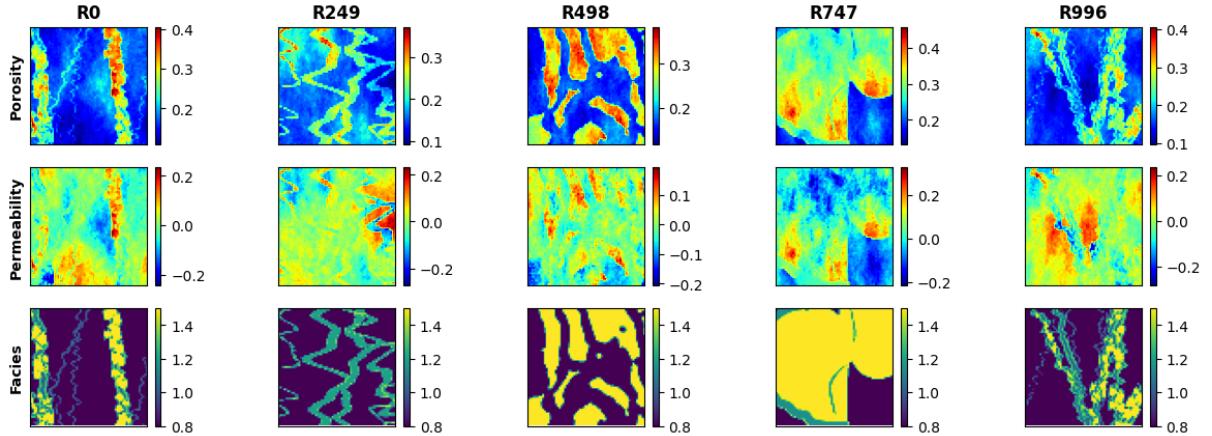
### 3.1 Reservoir Model and Simulation

We use SGeMS [83] to construct the subsurface  
 547 uncertainty model, an ensemble of static feature  
 548 realizations that is representative of various poten-  
 549 tial geologic scenarios for CO<sub>2</sub> storage. Using  
 550 sequential Gaussian co-simulation [84], we gen-  
 551 erate a set of 1,000 random porosity ( $\phi$ ) and  
 552 permeability ( $k$ ) distributions with a wide range  
 553 of values, as shown in Figure 11. Facies distri-  
 554 butions are obtained from a library of deepwater  
 555 fluvial training images [85, 86]. These encompass  
 556 a wide range of possible geologic scenarios includ-  
 557 ing marked point (lobe, ellipse, and bar), FluvSim  
 558 (channel, channel-levee, and channel-levee-splay),  
 559 surface based (compensational cycles of lobes),  
 560 and bank retreat (channel complex). To generate  
 561 consistent porosity and permeability distributions  
 562 with the facies-based geologic scenarios, we condi-  
 563 tion the original porosity and permeability dis-  
 564 tributions to the facies distributions. The resulting  
 565 fluvial distributions are shown in Figure 12.

The model has dimensions of 1km-1km-100m  
 566 in the x-, y-, and z-directions, respectively. We use  
 567 64 uniform grid cells in the x- and y-directions.  
 568 The grid design is sufficiently refined to resolve the  
 569 pressure and saturation plumes in highly heteroge-  
 570 neous reservoirs while remaining computationally  
 571 tractable for the purpose of training deep learn-  
 572 ing models. A random number of injection wells,  
 573  $w \in [1, 3]$ , are placed randomly along the reservoir  
 574 for each of the 1,000 realizations, no closer than  
 575 250m from the boundaries, as shown in Figure  
 576 13. The injection well(s) are randomly placed and  
 577 not conditioned to zones of preferential porosity,  
 578 permeability, nor facies. Each injection well has a



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



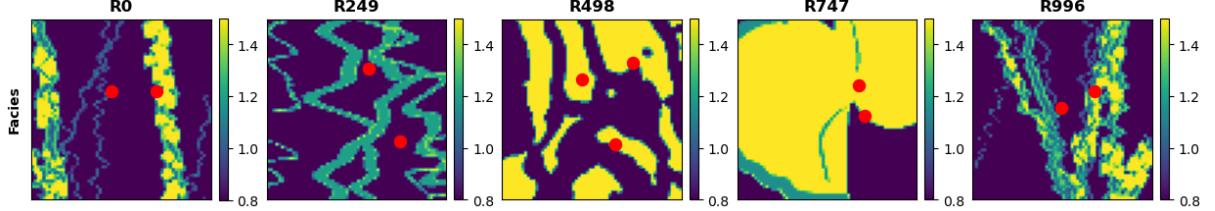
**Fig. 12** Spatial distribution conditioned to facies (top) for porosity (middle) and permeability (bottom) for 5 random realizations.

constant radius of 0.1m and a single and continuous perforation that injects pure supercritical CO<sub>2</sub> at a constant rate such that the total injection rate of the  $w$  well(s) is 0.5 megatons per year.

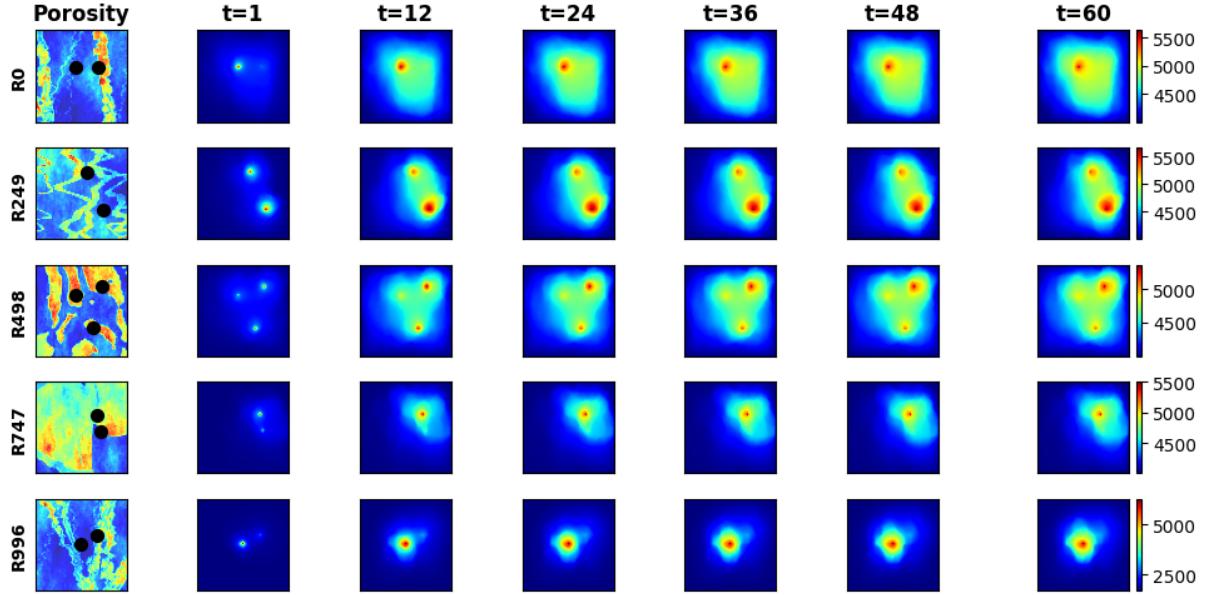
The conditional fluvial porosity and permeability distributions are used as input models for the numerical simulation of geologic CO<sub>2</sub> storage using MRST [87] to calculate the response models for training our proposed model. The reservoir is initialized as a fully water saturated zone (i.e., 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 CO<sub>2</sub> 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.



**Fig. 13**  $\text{CO}_2$  injection well(s) location (red) overlaid over facies distributions for 5 random realizations.



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

### 3.2 Training Performance

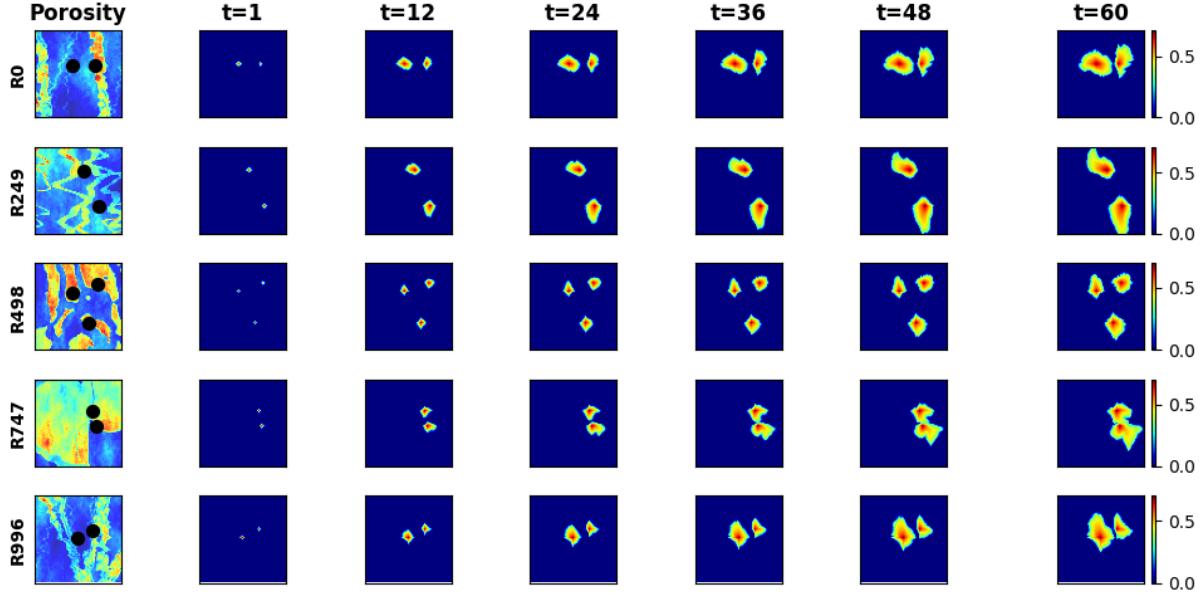
Using an NVIDIA Quadro M6000 GPU, we train for 100 epochs with batch size of 50. The model has in total 97,523,370 parameters, and the training time required is approximately 88 minutes for all 1,500 training realizations. The training and validation performance per epoch is shown in Figure 16. We observe 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 30 seconds to obtain the dynamic pressure and saturation predictions from the static geologic models. Our Stochastic pix2vid model obtains the same results in approximately 4.59 milliseconds, corresponding to a  $6,500 \times$  speedup. The average MSE

for the ensemble is  $9.21 \times 10^{-4}$  and  $9.70 \times 10^{-4}$  for training and testing, respectively. Similarly, the average SSIM for the ensemble is 98.97% and 97.91% for training and testing, respectively.

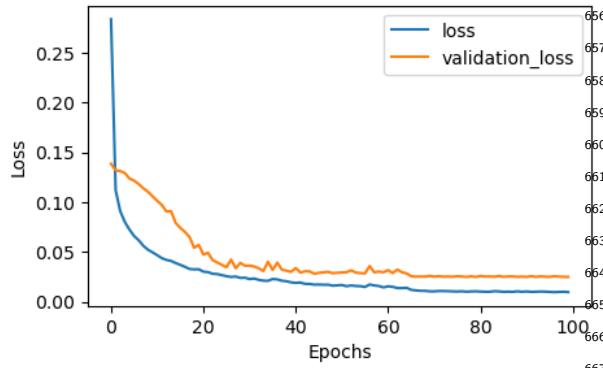
### 3.3 Prediction Results

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

Figures 17 and 18 show the predicted dynamic pressure and saturation distributions, respectively, along with the absolute difference to HFS for 3 training realizations. We observe reasonable agreement between the true and predicted  $\text{CO}_2$



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



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

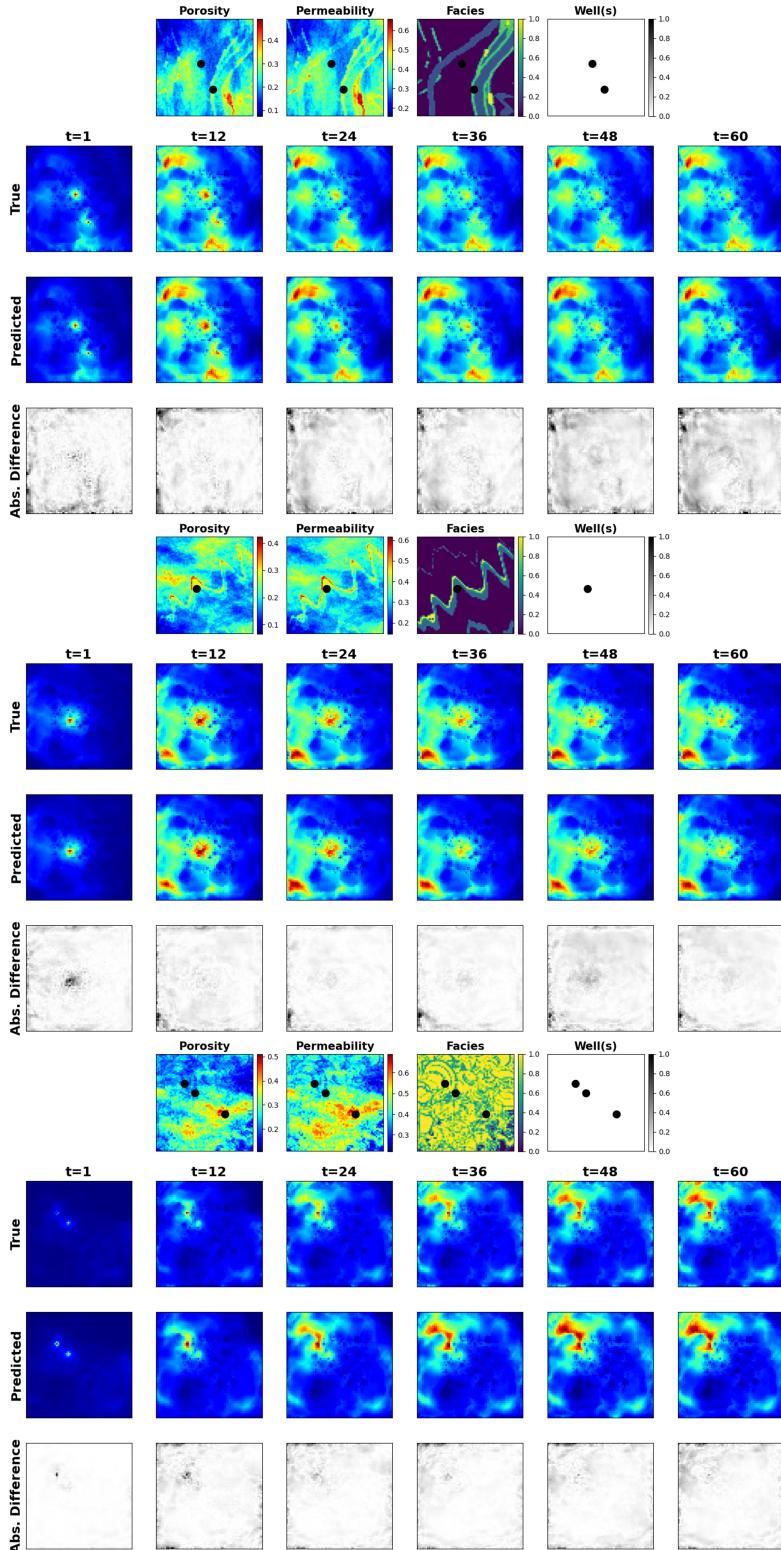
predictions. This indicates that the Stochastic pix2vid model is generalizable and achieves on par performance with HFS at a fraction of the computational cost.

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

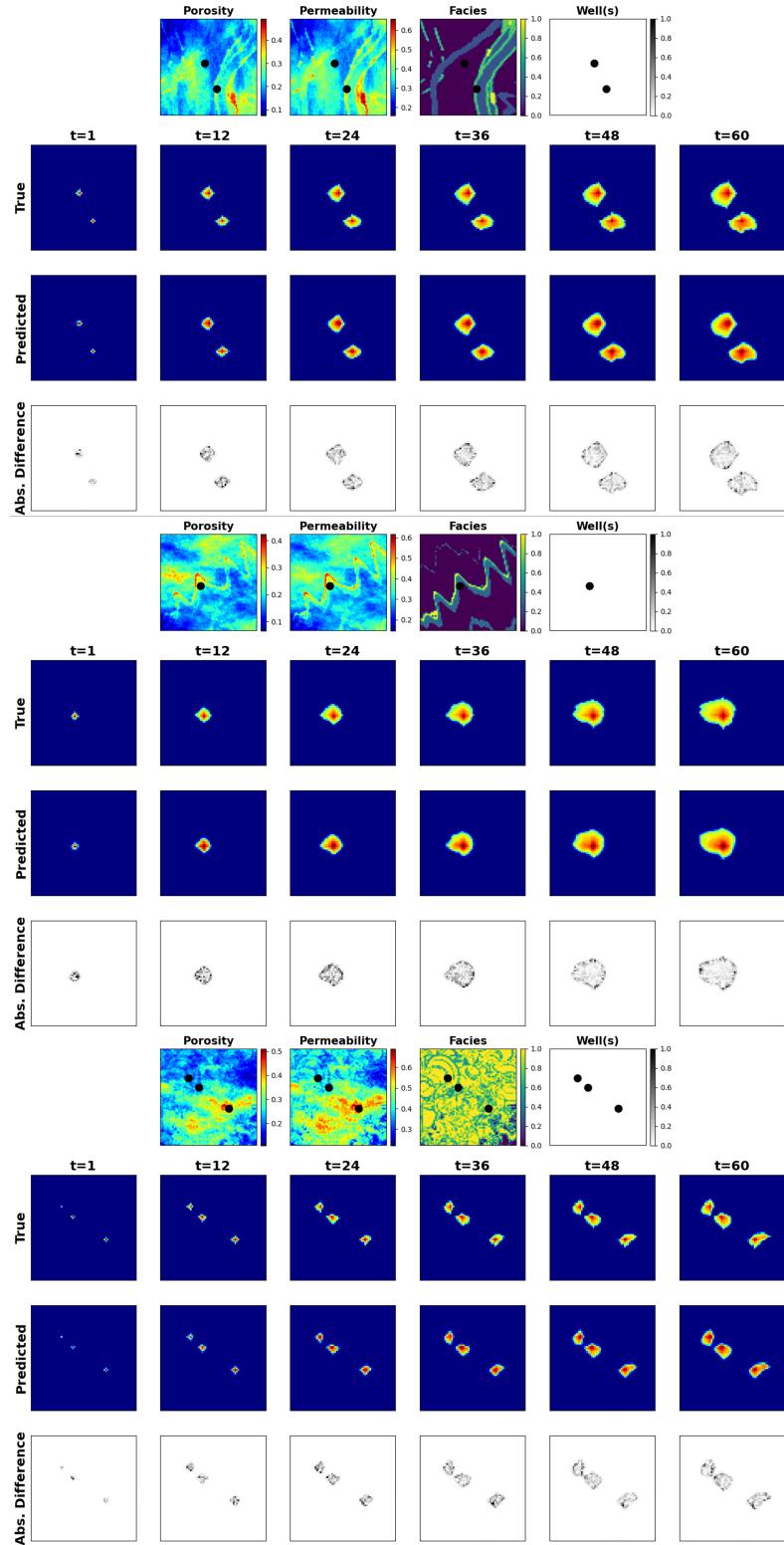
These results imply that our Stochastic pix2vid is capable of learning the spatiotemporal relationship between the static geologic models and the dynamic reservoir response. Thus, our image-to-video architecture can outperform current image-to-image and encoder-recurrent-decoder architectures for improved reservoir

pressure and saturation plumes over time, pixel-wise with an average MSE of  $3.25 \times 10^{-4}$  and perceptually with SSIM of 98.59% for pressure predictions and MSE of  $1.50 \times 10^{-4}$  and SSIM of 97.31% for saturation predictions.

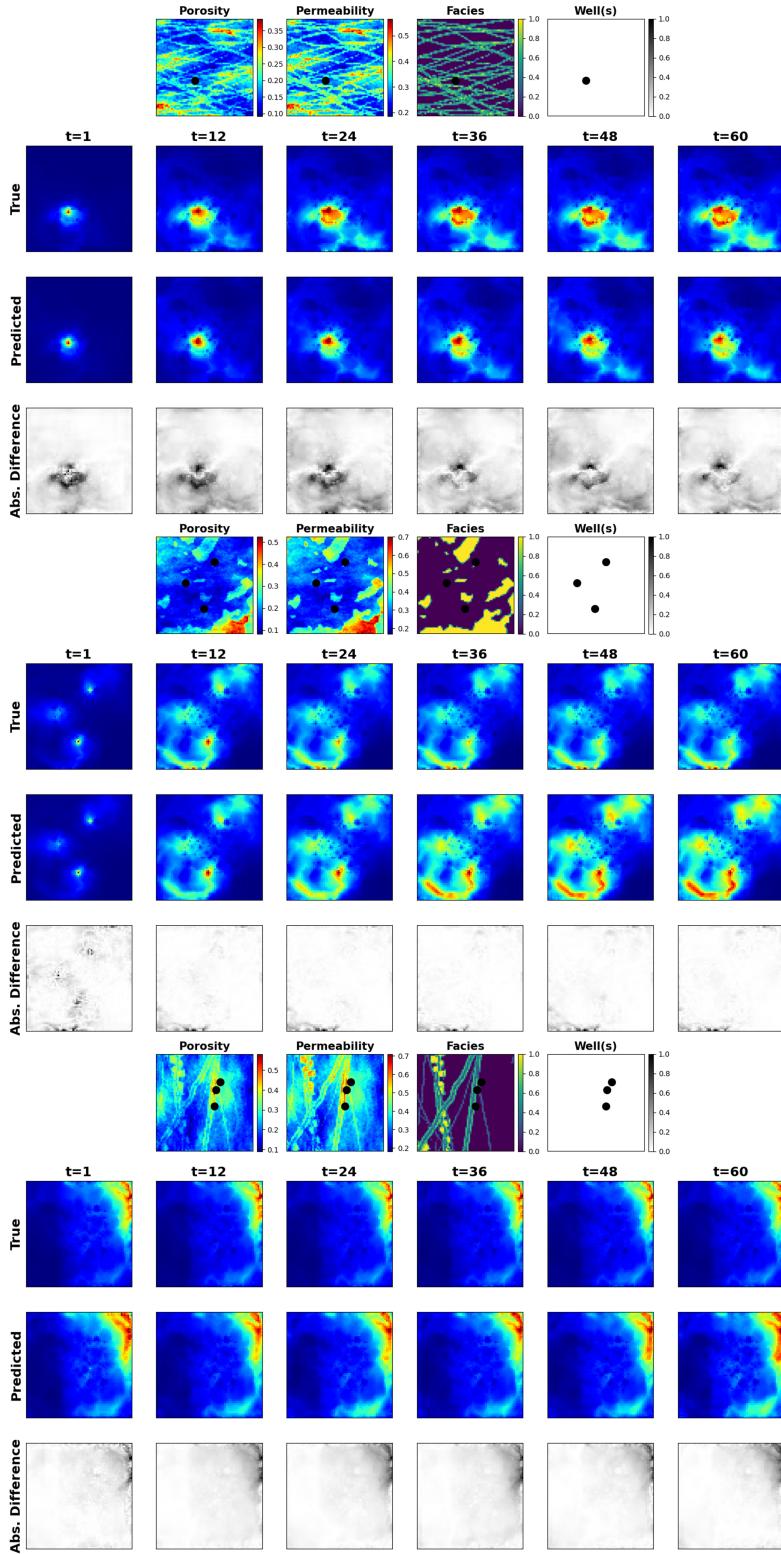
Similarly, Figures 19 and 20 show the pressure and saturation distributions predictions along with the absolute difference to HFS for 3 testing realizations. We observe a similar performance, with an average MSE of  $3.71 \times 10^{-4}$  and SSIM of 97.55% for pressure predictions and MSE of  $1.61 \times 10^{-3}$  and SSIM of 96.19% for saturation



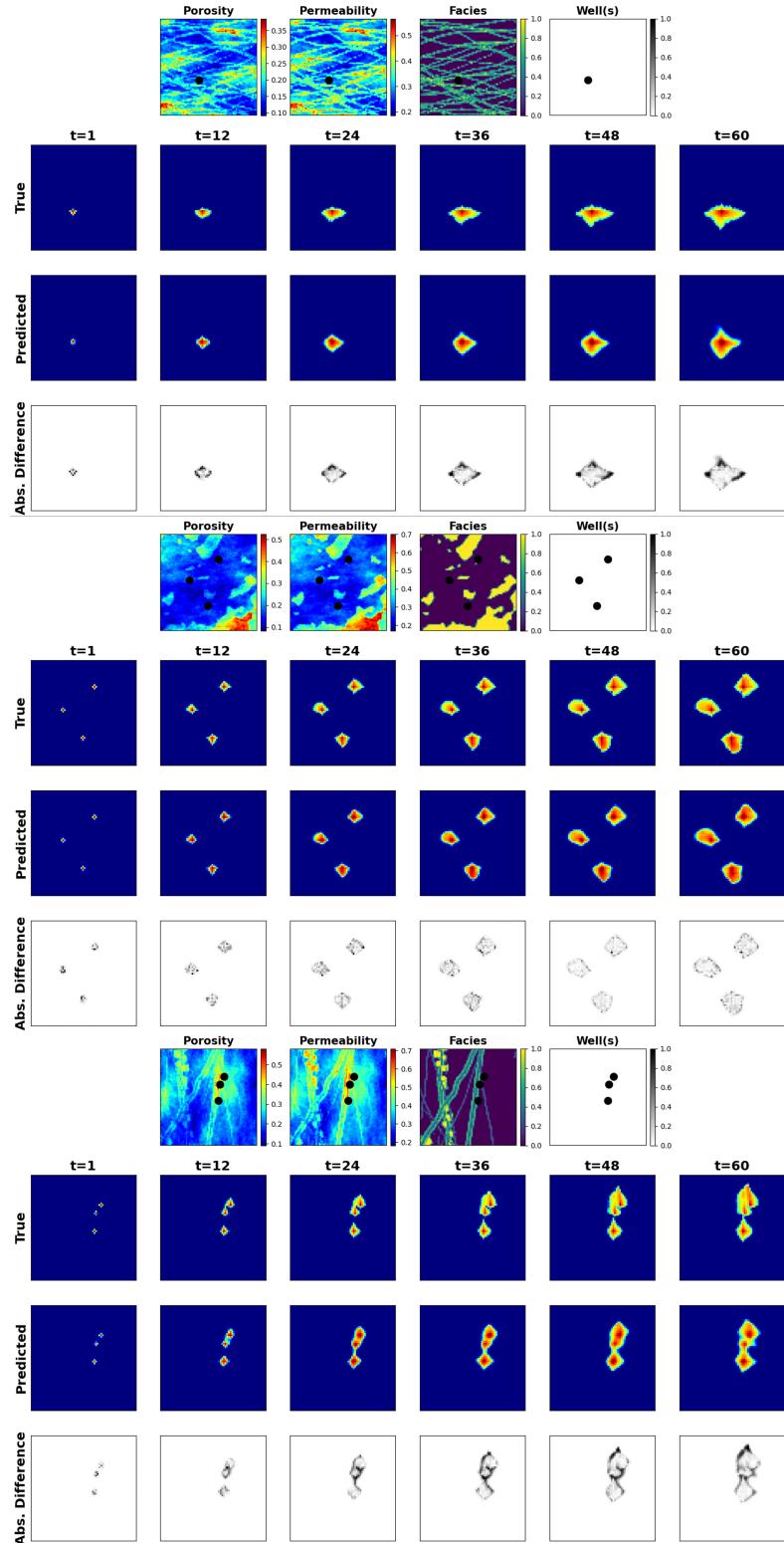
**Fig. 17** Normalized pressure distribution over time for 3 random training realization. For each panel, the top row is the ground truth from the HFS, the middle row is the Stochastic pix2vid prediction, and the bottom row is the absolute difference to HFS.



**Fig. 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.



**Fig. 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.



**Fig. 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.

behavior prediction. A comparison of true versus predicted results for pressure and saturation responses for the testing data is shown in Figure 22. For the pressure and saturation predictions, the average  $R^2$  over time is approximately 99% with narrow 95% prediction bands that recursively narrow over time. From Figure 22 we observe the Stochastic pix2vid model's performance at recursively refining the predictions over time due to the residual connections in the spatiotemporal decoder network.

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

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

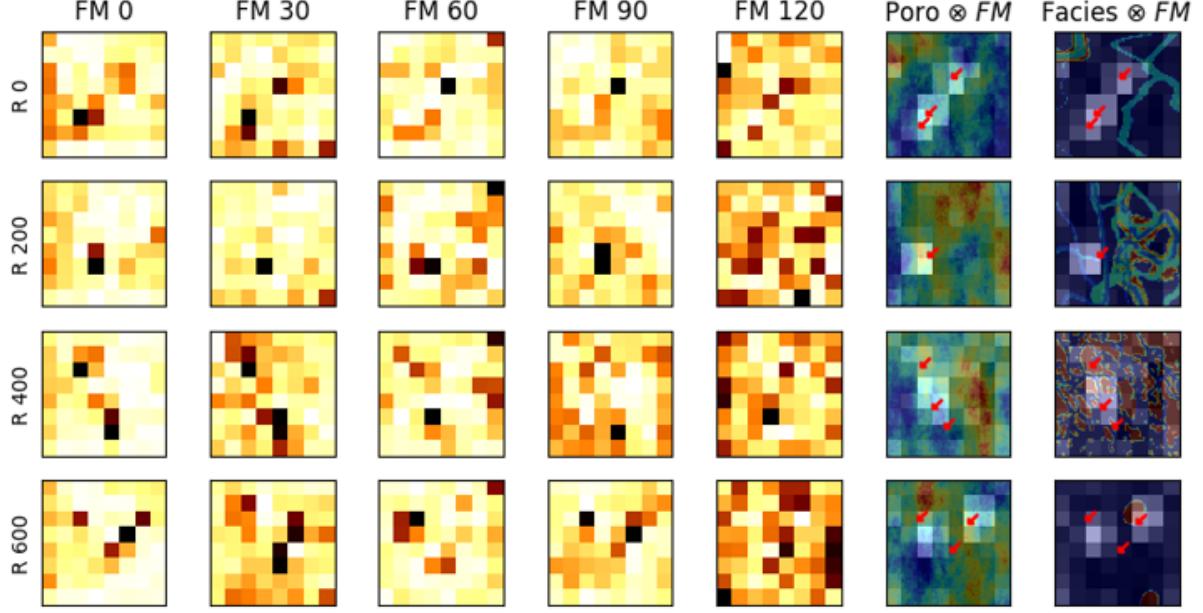
$\text{CO}_2$  saturation and pressure buildup fronts are important quantities for geologic  $\text{CO}_2$  storage projects and are often used for regulatory

oversight [89, 90], monitoring metrics or history matching purposes [91, 92]. The distance between the injection well(s) and the saturation fronts represents the maximum extent of the  $\text{CO}_2$  plume; however, these are often very difficult to capture accurately with data-driven proxy models. Our Stochastic pix2vid method shows greater absolute error on and around the plume fronts compared to within the plumes. However, the overall shape and intensity of the pressure and saturation distributions over time is very well captured for all realizations despite being highly heterogeneous. Therefore, the Stochastic pix2vid model can be used as a reliable replacement for expensive numerical reservoir simulations, especially in cases where large number of runs are required to obtain dynamic estimates (e.g., well placement and control optimization, history matching, uncertainty quantification).

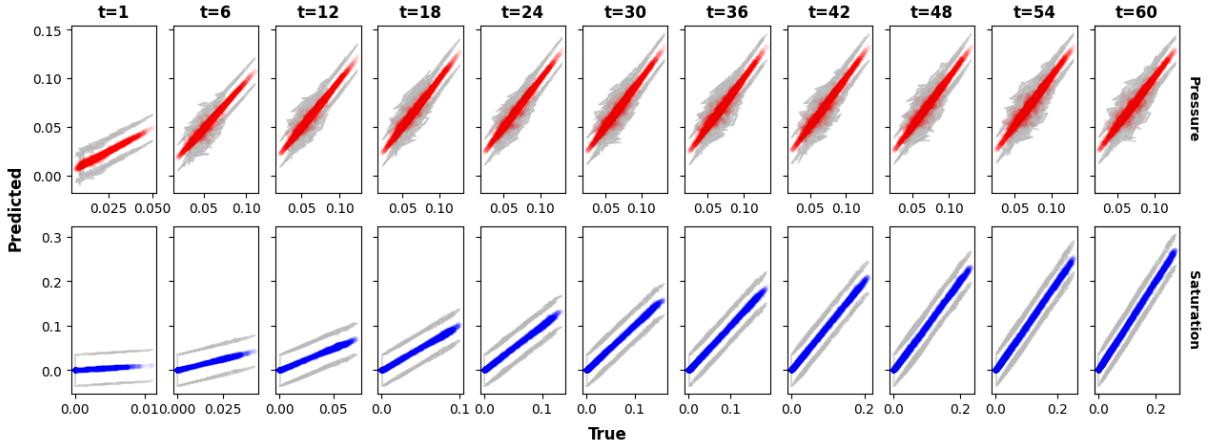
### 3.4 Discussion

In our Stochastic pix2vid model, the encoder block is composed of separable convolutions, squeeze and excite layers, and instance normalization. These three particular implementations allow for precise parameterization of the geologic realization into a latent representation, without mixing the effects of Gaussian-distributed properties against binary or binomial-distributed properties. Using recursive residual ConvLSTM layers, the decoder block iteratively predicts each dynamic state, or video frame, from the concatenation of the previous dynamic latent representation and the intermediate encoding parameterizations. Thus, our architecture makes the proxy model an image-to-video prediction formulation for dynamic reservoir states from a static geologic realization.

To further demonstrate the effectiveness of our Stochastic pix2vid model for geologic  $\text{CO}_2$  storage operations, we plot the cumulative pixel-wise  $\text{CO}_2$  saturation as a surrogate for the cumulative  $\text{CO}_2$  volume injected. For all training and testing realizations, Figure 23 shows the sum of pixel-wise  $\text{CO}_2$  saturation and the probability density function (PDF) of the true versus predicted saturations. We observe an  $R^2$  of 98% for training and 96% for testing in the cumulative  $\text{CO}_2$  saturation of true versus predicted results, and a conformable PDFs for both training and testing.



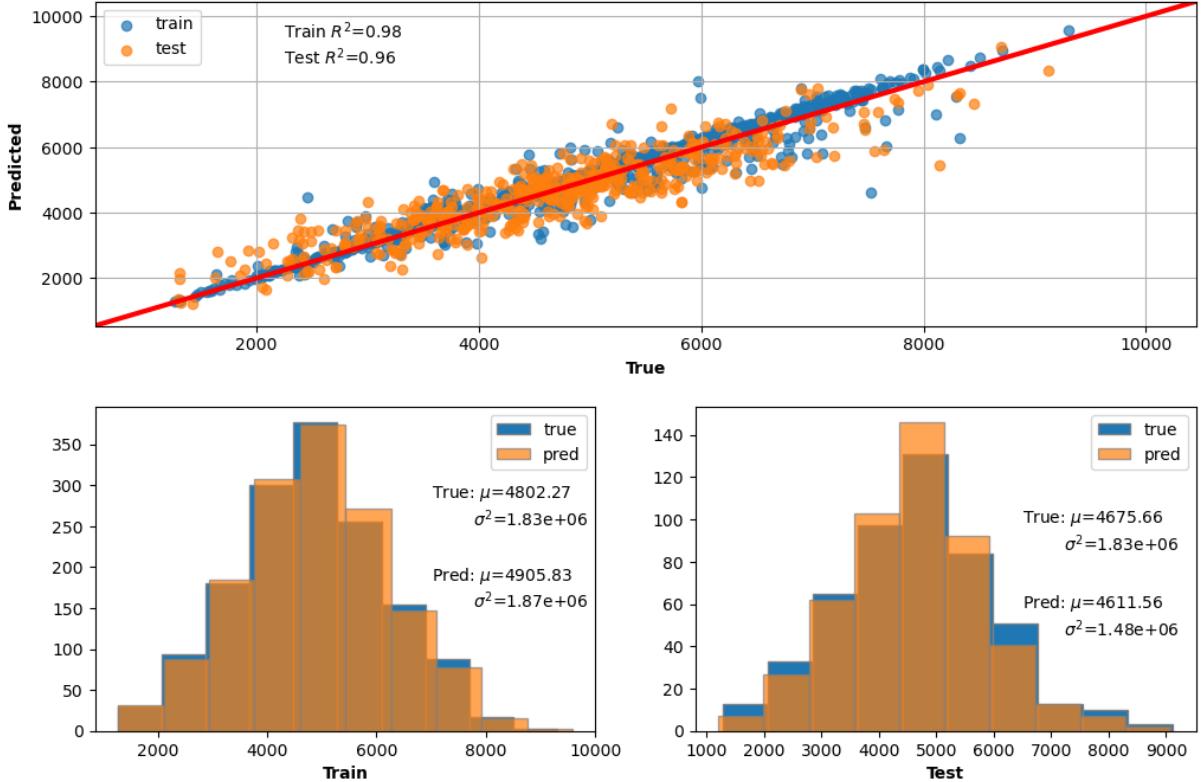
**Fig. 21** Five random feature maps (FM) of  $z_m^3$  for 4 random realizations. Their average is superimposed on top of the porosity and facies distributions to show the attention mechanism of the encoder. Bright colors represent higher attention and dark colors represent lower attention.



**Fig. 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.

Our Stochastic pix2vid method has several limitations. In order to learn the spatiotemporal relationships between input images and output videos, the model requires substantial amounts of training data, which in turn require expensive physics-based numerical simulation runs. Moreover, the method would require retraining in order

to apply to a different subsurface flow and transport problem, increasing the time required for generating the training data and the time required to retrain the model. One major limitation is the inability to predict for timesteps beyond those present in the training data. The architecture of the Stochastic pix2vid is designed to reconstruct



**Fig. 23** (Top) True vs. predicted cumulative CO<sub>2</sub> volume injected via pixel-wise saturation. (Bottom) True vs. predicted distributions of cumulative CO<sub>2</sub> saturation for training (left) and testing (right).

only the 11 timesteps present in  $d$ , therefore it  
 is capable of interpolation for steps in between  
 the training timesteps, but incapable to fore-  
 cast beyond  $t = 5$  years (60 months). Lastly, the  
 method is designed for images at the resolution  
 of  $64 \times 64$  pixels, and preprocessing is required to  
 reshape training data of other dimensions to this  
 size.

## 4 Conclusions

We develop a deep learning-based spatiotemporal proxy model to provide efficient flow predictions for a large-scale GCS operations to support optimum decision making. Our proposed method, Stochastic pix2vid, introduces the use of a spatiotemporal convolutional-recurrent architecture for dynamic predictions of CO<sub>2</sub> pressure and saturation distributions over time from a static geologic realization representing the subsurface uncertainty model. The framework is developed as

an image-to-video prediction, which is an under-determined estimation problem. Specifically, the implementation expands upon the architectures of current encoder-recurrent-decoder models and provides a fast and accurate proxy as a replacement for physics-based numerical reservoir simulation.

The spatiotemporal proxy is applied to a synthetic 2D GCS project with multiple uncertain geologic scenarios and random number and location of injection well(s). A total of 1,000 geologic models are obtained from a variety of possible geologic scenarios including fluvial, turbidite, and deepwater lobe systems. The spatial distribution of porosity, permeability and facies, and the spatial location of the injector well(s) are used as the input data. The proxy model is used to predict the dynamic reservoir response over time, namely the video frames, corresponding to the dynamic CO<sub>2</sub> pressure and saturation distributions, which are obtained offline for training using HFS. The total

- 836 training time is 88 minutes on a single NVIDIA Quadro M6000 GPU, and predictions are obtained with 98-99% accuracy within approximately 4.6 milliseconds, compared to the approximate 30 seconds required for HFS, a  $6,500\times$  speedup.
- 837 There are several opportunities for future work. First, an extension to 3D geologic models and their corresponding dynamic predictions is key to scaling up this method for real-world applications. Similarly, although the Stochastic pix2vid proxy model is only trained for GCS prediction, it is applicable for a range of processes such as ground-water, compositional, geothermal, or conventional oil and gas systems. Moreover, it is possible to extend the Stochastic pix2vid model from a data-driven mapping to a PINN by including the discretized form of the governing PDE in the loss function and minimizing the residuals. Another future opportunity is to test the performance of the Stochastic pix2vid model on unseen timesteps, either interpolating the training timesteps or extrapolating beyond the training timesteps. Furthermore, the Stochastic pix2vid model can be used as a proxy in workflows for history matching and closed-loop reservoir management.
- 838 **Reproducibility.** The code will be made publicly available on the author's repository ([github.com/misaelmmorales](https://github.com/misaelmmorales) and [github.com/GeostatsGuy](https://github.com/GeostatsGuy)).
- 839 **Funding.** This research did not receive any specific grant from funding agencies in the public, or not-for-profit sectors.
- 840 **Declarations.** The authors declare no conflict of interests.
- 841 **Acknowledgments.** The authors thank the Digital Reservoir Characterization Technology (DIRECT) and Formation Evaluation (FE) Industry Affiliate Program at the University of Texas at Austin for supporting this work.
- 842 **References**
- 843 [1] K. Michael, A. Golab, V. Shulakova, J. Ennis-King, G. Allinson, S. Sharma, and T. Aiken. Geological storage of co<sub>2</sub> 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>.
- 844 [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.
- 845 [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 co<sub>2</sub> 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.
- 846 [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.
- 847 [5] Energy 2020. European commission. In *A strategy for competitive, sustainable and secure energy*, 2010.
- 848 [6] United nations. Agreement, p. *United Nations Treaty Collect*, pages 1–27, 2015.
- 849 [7] S. Bachu. Review of co<sub>2</sub> 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.
- 850 [8] J.F.D. Tapia, J.-Y. Lee, R.E.H. Ooi, D.C.Y. Foo, and R.R. Tan. Optimal co<sub>2</sub> allocation and scheduling in enhanced oil recovery (eor) operations. *Applied Energy*, 184:337–345, 2016. doi: 10.1016/j.apenergy.2016.09.093.
- 851 [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 co<sub>2</sub> sequestration in a faulted saline formation in

- italy. *Advances in Water Resources*, 62:570– 973  
587, 2013. doi: 10.1016/j.advwatres.2013.04. 974  
006. cited By 25. 975
- [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>. 980
- [11] K. Rashid, W. Bailey, B. Couët, and D. Wilkinson. An efficient procedure for expensive reservoir-simulation optimization under uncertainty. *SPE Economics and Management*, 5(4):21–33, 2013. doi: 10.2118/ 985  
167261-PA. cited By 16. 986
- [12] C. Luo, S.-L. Zhang, C. Wang, and Z. Jiang. A metamodel-assisted evolutionary algorithm for expensive optimization. *Journal of Computational and Applied Mathematics*, 236(5): 991  
759–764, 2011. doi: 10.1016/j.cam.2011.05. 992  
047. cited By 29. 993
- [13] Javier E. Santos, Bernard Chang, Alex Gigliotti, Eric Guiltinan, Mohamed Mehana, Arvind Mohan, James McClure, Qinjun Kang, Hari Viswanathan, Nicholas Lubbers, Masa Prodanovic, and Michael Pyrcz. Learning from a big dataset of digital rock simulations. In *AGU Fall Meeting Abstracts*, volume 1001, pages H25O–1207, December 2021. 1004
- [14] Bailian Chen, Dylan R. Harp, Youzuo Lin, Elizabeth H. Keating, and Rajesh J. Pawar. Geologic co<sub>2</sub> sequestration monitoring design: A machine learning and uncertainty quantification based approach. *Applied Energy*, 225:332–345, 9 2018. ISSN 03062619. doi: 10.1016/j.apenergy.2018.05.044. 1008
- [15] Wenyue Sun and Louis J. Durlofsky. Data-space approaches for uncertainty quantification of co<sub>2</sub> plume location in geological carbon storage. *Advances in Water Resources*, 123:234–255, 1 2019. ISSN 03091708. doi: 10.1016/j.advwatres.2018.10.028. cited By 23. 1012
- [16] Bailian Chen, Dylan R. Harp, Zhiming Lu, and Rajesh J. Pawar. Reducing uncertainty in geologic co<sub>2</sub> sequestration risk assessment by assimilating monitoring data. *International Journal of Greenhouse Gas Control*, 94, 3 2020. ISSN 17505836. doi: 10.1016/j.ijggc.2019.102926. 976
- [17] B. Li and S.M. Benson. Influence of small-scale heterogeneity on upward co<sub>2</sub>plume migration in storage aquifers. *Advances in Water Resources*, 83:389–404, 2015. doi: 10.1016/j.advwatres.2015.07.010. cited By 84. 982
- [18] Su Jiang and Louis J. Durlofsky. Use of multifidelity training data and transfer learning for efficient construction of subsurface flow surrogate models. *Journal of Computational Physics*, 474, 2 2023. ISSN 10902716. doi: 10.1016/J.JCP.2022.111800. 988
- [19] *Best Practices in Automatic Permeability Estimation: Machine-Learning Methods vs. Conventional Petrophysical Models*, volume Day 4 Tue, June 13, 2023 of *SPWLA Annual Logging Symposium*, 06 2023. doi: 10.30632/SPWLA-2023-0084. 994
- [20] H. Wu, N. Lubbers, H.S. Viswanathan, and R.M. Polleya. A multi-dimensional parametric study of variability in multi-phase flow dynamics during geologic co<sub>2</sub> sequestration accelerated with machine learning. *Applied Energy*, 287, 2021. doi: 10.1016/j.apenergy. 2021.116580. cited By 14. 1002
- [21] Siddharth Misra, Yusuf Falola, Polina Churilova, Rui Liu, Chung-Kan Huang, and Jose F. Delgado. Deep learning assisted extremely low-dimensional representation of subsurface earth. *SSRN Electronic Journal*, 8 2022. doi: 10.2139/SSRN.4196705. 1009
- [22] Ademide O. Mabadeje and Michael J. Pyrcz. Rigid transformations for stabilized lower dimensional space to support subsurface uncertainty quantification and interpretation, 2023. 1011
- [23] Mingliang Liu, Dario Grana, and Tapan Mukerji. Randomized tensor decomposition for large-scale data assimilation problems for 1017

- 1018 carbon dioxide sequestration. *Mathematical Geosciences*, 54:1139–1163, 5 2022. ISSN 1062  
 1019 18748953. doi: 10.1007/S11004-022-10005-1/ 1063  
 1020 FIGURES/17. 1064  
 1021 1065
- 1022 [24] S.W.A. Canchumuni, A.A. Emerick, and 1066  
 1023 M.A.C. Pacheco. Towards a robust parameterization for conditioning facies models using 1067  
 1024 deep variational autoencoders and ensemble 1068  
 1025 smoother. *Computers and Geosciences*, 128: 1069  
 1026 87–102, 2019. doi: 10.1016/j.cageo.2019.04. 1070  
 1027 006. cited By 80. 1071  
 1028 1072
- 1029 [25] Y. Zhang, P. Vouzis, and N.V. Sahinidis. Gpu 1073  
 1030 simulations for risk assessment in co2 geo- 1074  
 1031 logic sequestration. *Computers and Chemical 1075  
 1032 Engineering*, 35(8):1631–1644, 2011. doi: 10. 1076  
 1033 1016/j.compchemeng.2011.03.023. cited By 1077  
 1034 20. 1078
- 1035 [26] Bicheng Yan, Dylan Robert Harp, Bailian 1079  
 1036 Chen, and Rajesh J. Pawar. Improving 1080  
 1037 deep learning performance for predicting 1081  
 1038 large-scale geological co2 sequestration mod- 1082  
 1039eling through feature coarsening. *Scientific 1083  
 1040 Reports*, 12:1–12, 11 2022. ISSN 2045-2322.  
 1041 doi: 10.1038/s41598-022-24774-6. 1084
- 1042 [27] Zeeshan Tariq, Murtada Saleh Aljawad, 1086  
 1043 Amjed Hasan, Mobeen Murtaza, Emad 1087  
 1044 Mohammed, Ammar El-Husseiny, 1088  
 1045 Sulaiman A Alarifi, Mohamed Mahmoud, 1089  
 1046 and Abdulazeez Abdulraheem. A systematic 1090  
 1047 review of data science and machine learning 1091  
 1048 applications to the oil and gas industry. *Journal 1092  
 1049 of Petroleum Exploration and Production 1093  
 1050 Technology*, pages 1–36, 2021. 1094
- 1051 [28] Mohammad Ali Mirza, Mahtab Ghoroori, 1095  
 1052 and Zhangxin Chen. Intelligent petroleum 1096  
 1053 engineering. *Engineering*, 18:27–32, 2022. 1097  
 1054 ISSN 2095-8099. doi: <https://doi.org/10.1016/j.eng.2022.06.009>. 1098
- 1099
- 1056 [29] Jean-Paul Chiles and Pierre Delfiner. *Geo- 1100  
 1057 statistics: modeling spatial uncertainty*, vol- 1101  
 1058 ume 713. John Wiley & Sons, 2012. 1102
- 1103
- 1059 [30] Michael J Pyrcz and Clayton V Deutsch. 1104  
 1060 *Geostatistical reservoir modeling*. Oxford 1105  
 1061 University Press, USA, 2014. 1106
- [31] Proctor Joshua Brunton, Steve and Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 2016. doi: 10.1073/pnas. 1517384113.
- [32] He Xiaolong Fries, William and Youngsoo Choi. Lasdi: Parametric latent space dynamics identification. *Computer Methods in Applied Mechanics and Engineering*, 2022. doi: 10.1016/j.cma.2022.115436.
- [33] Choi Youngsoo Fries William Belof Jonathan He, Xiaolong and Jiun-Shyan Chen. glasdi: Parametric physics-informed greedy latent space dynamics identification. *Journal of Computational Physics*, 2023.
- [34] M. Liu and D. Grana. Time-lapse seismic history matching with an iterative ensemble smoother and deep convolutional autoencoder. *Geophysics*, 85(1):M15–M31, 2020. cited By 2.
- [35] Syamil Mohd Razak, Anyue Jiang, and Behnam Jafarpour. Latent-space inversion (lsi): a deep learning framework for inverse mapping of subsurface flow data. *Computational Geoscience*, 26:71–99, 11 2022. doi: 10.1007/s10596-021-10104-8.
- [36] S. Oladyshkin, H. Class, and W. Nowak. Bayesian updating via bootstrap filtering combined with data-driven polynomial chaos expansions: Methodology and application to history matching for carbon dioxide storage in geological formations. *Computational Geosciences*, 17(4):671–687, 2013. doi: 10.1007/s10596-013-9350-6. cited By 36.
- [37] Anqi Bao, Eduardo Gildin, Abhinav Narasingam, and Joseph S. Kwon. Data-driven model reduction for coupled flow and geomechanics based on dmd methods. *Fluids*, 4:138, 7 2019. ISSN 2311-5521. doi: 10.3390/FLUIDS4030138.
- [38] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed

- 1107 machine learning. *Nature Reviews Physics*, 3 1151  
 1108 (6):422–440, 2021. 1152  
 1109 [39] Liu Yang, Dongkun Zhang, and George Em 1154  
 1110 Karniadakis. Physics-informed generative 1155  
 1111 adversarial networks for stochastic differen- 1156  
 1112 tial equations, 2018. 1157  
 1113 [40] N. Wang, H. Chang, and D. Zhang. Effi- 1158  
 1114 cient uncertainty quantification for dynamic 1159  
 1115 subsurface flow with surrogate by theory- 1160  
 1116 guided neural network. *Computer Methods 1161*  
*in Applied Mechanics and Engineering*, 373, 1161  
 1117 2021. doi: 10.1016/j.cma.2020.113492. cited 1162  
 1118 By 33. 1163  
 1119 [41] Emilio Jose Rocha Coutinho, Marcelo 1164  
 1120 Dall'Aqua, and Eduardo Gildin. Physics- 1165  
 1121 aware deep-learning-based proxy reservoir 1166  
 1122 simulation model equipped with state 1167  
 1123 and well output prediction. *Frontiers 1168*  
*in Applied Mathematics and Statistics*, 1169  
 1124 7:49, 9 2021. ISSN 22974687. doi: 1170  
 1125 10.3389/FAMS.2021.651178/BIBTEX. 1171  
 1126 [42] Yinhao Zhu, Nicholas Zabaras, Phaedon- 1172  
 1127 Stelios Koutsourelakis, and Paris Perdikaris. 1173  
 1128 Physics-constrained deep learning for high- 1174  
 1129 dimensional surrogate modeling and uncer- 1175  
 1130 tainty quantification without labeled data. 1176  
 1131 *Journal of Computational Physics*, 394:56– 1177  
 1132 81, oct 2019. doi: 10.1016/j.jcp.2019.05. 1178  
 1133 024. URL <https://doi.org/10.1016%2Fj.jcp.2019.05.024>. 1179  
 1134 [43] B Yegnanarayana. *Artificial neural networks*. 1181  
 1135 PHI Learning Pvt. Ltd., 2009. 1182  
 1136 [44] Jeff Heaton. Ian goodfellow, yoshua 1183  
 1137 bengio, and aaron courville: Deep learn- 1184  
 1138 ing: The mit press, 2016, 800 pp, isbn: 1185  
 0262035618. *Genetic programming and evolv- 1186*  
*able machines*, 19(1-2):305–307, 2018. 1187  
 1139 [45] Yimin Liu and Louis J Durlofsky. 3d cnn- 1188  
 1140 pca: A deep-learning-based parameterization 1189  
 1141 for complex geomodels. *Computers & Geo- 1190*  
*sciences*, 148:104676, 2021. 1191  
 1142 [46] Zixiao Yang, Qiyu Chen, Zhesi Cui, Gang 1192  
 1143 Liu, Shaoqun Dong, and Yiping Tian. Auto- 1193  
 1144 matic reconstruction method of 3d geological 1194  
 1145 1195  
 1146 models based on deep convolutional gener- 1151  
 1147 ative adversarial networks. *Computational 1152*  
*Geosciences*, 26:1135–1150, 2022. doi: 10. 1153  
 1007/s10596-022-10152-8.  
 1148 [47] Su Jiang and Louis J Durlofsky. Data-space 1154  
 1149 inversion using a recurrent autoencoder for 1155  
 1150 time-series parameterization. *Computational 1156*  
*Geosciences*, 25:411–432, 2021.  
 1151 [48] Yanrui Ning, Hossein Kazemi, and Pejman 1157  
 1152 Tahmasebi. A comparative machine learning 1158  
 1153 study for time series oil production forecast- 1159  
 1154 ing: Arima, lstm, and prophet. *Computers 1160*  
*and Geosciences*, 164:105126, 7 2022. ISSN 1161  
 00983004. doi: 10.1016/j.cageo.2022.105126.  
 1155 [49] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yix- 1162  
 1156 uan Wei, Zheng Zhang, Stephen Lin, and 1163  
 1157 Baining Guo. Swin transformer: Hierarchi- 1164  
 1158 cal vision transformer using shifted windows. 1165  
 In *Proceedings of the IEEE/CVF interna- 1166  
 tional conference on computer vision*, pages 1167  
 10012–10022, 2021.  
 1159 [50] Liuqing Yang, Sergey Fomel, Shoudong 1170  
 1160 Wang, Xiaohong Chen, Wei Chen, Omar M. 1171  
 1161 Saad, and Yangkang Chen. Porosity and 1172  
 1162 permeability prediction using a transformer 1173  
 1163 and periodic long short-term network. *Geo- 1174*  
*physics*, 88(1):WA293–WA308, 01 2023. ISSN 1175  
 0016-8033. doi: 10.1190/geo2022-0150.1.  
 1164 [51] Eduardo Maldonado Cruz and Michael J 1176  
 1165 Pyrcz. Multi-horizon well performance fore- 1177  
 1166 casting with temporal fusion transformers. 1178  
 Available at SSRN 4403939.  
 1167 [52] Wen Pan, Carlos Torres-Verdín, and 1179  
 1168 Michael J. Pyrcz. Stochastic pix2pix: A new 1180  
 1169 machine learning method for geophysical and 1181  
 1170 well conditioning of rule-based channel reser- 1182  
 1171 voir models. *Natural Resources Research*, 1183  
 30:1319–1345, 4 2021. ISSN 15738981. doi: 1184  
 10.1007/S11053-020-09778-1/FIGURES/24.  
 1172 [53] Bogdan Sebacher and Stefan Adrian Toma. 1185  
 1173 Bridging deep convolutional autoencoders 1186  
 1174 and ensemble smoothers for improved esti- 1187  
 1175 mation of channelized reservoirs. *Mathematical 1188*  
*Geosciences*, 54:903–939, 7 2022. ISSN 1189  
 18748953. doi: 10.1007/S11004-022-09997-7/

- 1196 TABLES/3.
- 1197 [54] Jichao Bao, Liangping Li, and Arden Davis. Variational autoencoder or generative adversarial networks? a comparison of two deep learning methods for flow and transport data assimilation. *Mathematical Geosciences*, 54: 1017–1042, 8 2022. ISSN 18748953. doi: 10.1007/S11004-022-10003-3/FIGURES/17.
- 1198 [55] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, 2015. cited By 358.
- 1199 [56] Eduardo Maldonado-Cruz and Michael J. Pyrcz. Fast evaluation of pressure and saturation predictions with a deep learning surrogate flow model. *Journal of Petroleum Science and Engineering*, 212:110244, 5 2022. ISSN 0920-4105. doi: 10.1016/J.PETROL.2022.110244.
- 1200 [57] Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022. ISSN 0309-1708. doi: <https://doi.org/10.1016/j.advwatres.2022.104180>.
- 1201 [58] Gege Wen, Zongyi Li, Qirui Long, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. Real-time high-resolution co 2 geological storage prediction using nested fourier neural operators. *Energy & Environmental Science*, 2023. ISSN 1754-5692. doi: 10.1039/d2ee04204e.
- 1202 [59] Honggeun Jo, Wen Pan, Javier E Santos, Hyungsik Jung, and Michael J Pyrcz. Machine learning assisted history matching for a deepwater lobe system. *Journal of Petroleum Science and Engineering*, 207: 109086, 2021.
- 1203 [60] Feng Zhang, Long Nghiem, and Zhangxin Chen. Evaluating reservoir performance using a transformer based proxy model. *Geoenergy Science and Engineering*, 226: 211644, 2023.
- 1240 [61] Daowei Zhang and Heng Li. Efficient surrogate modeling based on improved vision transformer neural network for history matching. *SPE Journal*, pages 1–17, 2023.
- 1241 [62] Yong Do Kim and Louis J. Durlofsky. Convolutional – recurrent neural network proxy for robust optimization and closed-loop reservoir management. *Computational Geosciences*, pages 1–24, 1 2023. ISSN 1420-0597. doi: 10.1007/S10596-022-10189-9/TABLES/1.
- 1242 [63] Meng Tang, Yimin Liu, and Louis J. Durlofsky. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. *Journal of Computational Physics*, 413, 7 2020. ISSN 10902716. doi: 10.1016/J.JCP.2020.109456.
- 1243 [64] M. Tang, Y. Liu, and L.J. Durlofsky. Deep-learning-based surrogate flow modeling and geological parameterization for data assimilation in 3d subsurface flow. *Computer Methods in Applied Mechanics and Engineering*, 376, 2021. doi: 10.1016/j.cma.2020.113636. cited By 39.
- 1244 [65] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Generating videos with scene dynamics, 2016.
- 1245 [66] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean square error, 2016.
- 1246 [67] Ruben Villegas, Jimei Yang, Seunghoon Hong, Xunyu Lin, and Honglak Lee. Decomposing motion and content for natural video sequence prediction, 2018.
- 1247 [68] Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. Mocogan: Decomposing motion and content for video generation, 2017.
- 1248 [69] Xingjian SHI, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information*

- 1284        *Processing Systems*, volume 28. Curran Asso- 1326  
 1285        ciates, Inc., 2015. URL [https://proceedings.neurips.cc/paper\\_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf).  
 1286  
 1287  
 1288  
 1289 [70] Michael Iliadis, Leonidas Spinoulas, and 1331  
 1290        Aggelos K. Katsaggelos. Deep fully- 1332  
 1291        connected networks for video compressive 1333  
 1292        sensing, 2017.  
 1293  
 1294 [71] Kai Xu and Fengbo Ren. Csvideonet: A real- 1335  
 1295        time end-to-end learning framework for high- 1336  
 1296        frame-rate video compressive sensing. In *2018 IEEE Winter Conference on Applications of 1337 Computer Vision (WACV)*, pages 1680–1688. 1338  
 1297        IEEE, 2018.  
 1298  
 1299 [72] Michael Dorkenwald, Timo Milbich, Andreas 1341  
 1300        Blattmann, Robin Rombach, Konstanti- 1342  
 1301        nos G. Derpanis, and Björn Ommer. Stochas- 1343  
 1302        tic image-to-video synthesis using cinns, 1344  
 1303        2021.  
 1304  
 1305 [73] Aleksander Holynski, Brian Curless, 1346  
 1306        Steven M. Seitz, and Richard Szeliski. Ani- 1347  
 1307        mating pictures with eulerian motion fields, 1348  
 1308        2020.  
 1309  
 1310 [74] Karsten Pruess, Curtis M Oldenburg, and 1350  
 1311        GJ Moridis. Tough2 user’s guide version 2. 1351  
 1312        Technical report, Lawrence Berkeley National 1352  
 1313        Lab.(LBNL), Berkeley, CA (United States), 1353  
 1314        1999.  
 1315  
 1316 [75] François Chollet. Xception: Deep learning 1355  
 1317        with depthwise separable convolutions. In 1356  
 1318        *Proceedings of the IEEE conference on com- 1357  
 1319        puter vision and pattern recognition*, pages 1358  
 1320        1251–1258, 2017.  
 1321  
 1322 [76] Jie Hu, Li Shen, and Gang Sun. Squeeze- 1360  
 1323        and-excitation networks. In *Proceedings of 1361  
 1324        the IEEE conference on computer vision and 1362  
 1325        pattern recognition*, pages 7132–7141, 2018.  
 1326  
 1327 [77] Dmitry Ulyanov, Andrea Vedaldi, and Vic- 1364  
 1328        tor Lempitsky. Instance normalization: The 1365  
 1329        missing ingredient for fast stylization. *arXiv 1366  
 1330        preprint arXiv:1607.08022*, 2016.  
 1331  
 1332  
 1333  
 1334 [78] Albert Tarantola. *Inverse problem theory 1326  
 1335        and methods for model parameter estimation.* 1327  
 1336        SIAM, 2005.  
 1337  
 1338 [79] D.S. Oliver, A.C. Reynolds, and N. Liu. 1329  
 1339        *Inverse theory for petroleum reservoir char- 1330  
 1340        acterization and history matching*, volume 1331  
 1341        9780521881517. 2008. doi: 10.1017/CBO9780511535642. cited By 766.  
 1342  
 1343 [80] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, 1334  
 1344        and Eero P. Simoncelli. Image quality assess- 1335  
 1345        ment: from error visibility to structural simi- 1336  
 1346        larity. *IEEE Transactions on Image Process- 1337  
 1347        ing*, 13:600–612, 4 2004. ISSN 1941-0042. doi: 1338  
 1348        doi.org/10.1109/TIP.2003.819861.  
 1349  
 1350 [81] Ilya Loshchilov and Frank Hutter. Decoupled 1334  
 1351        weight decay regularization. *arXiv preprint 1335  
 1352        arXiv:1711.05101*, 2017.  
 1353  
 1354 [82] Diederik P Kingma and Jimmy Ba. Adam: 1341  
 1355        A method for stochastic optimization. *arXiv 1342  
 1356        preprint arXiv:1412.6980*, 2014.  
 1357  
 1358 [83] Nicolas Remy, Alexandre Boucher, and Jian- 1346  
 1359        bing Wu. *Applied Geostatistics with SGEMS: 1347  
 1360        A User’s Guide*. Cambridge University Press, 1348  
 1361        2009.  
 1362  
 1363 [84] G. W. Verly. *Sequential Gaussian Cosim- 1350  
 1364        ulation: A Simulation Method Integrating 1351  
 1365        Several Types of Information*, pages 543– 1352  
 1366        554. Springer Netherlands, Dordrecht, 1993. 1353  
 1367        ISBN 978-94-011-1739-5. doi: 10.1007/ 1354  
 1368        978-94-011-1739-5\_42.  
 1369  
 1370 [85] M.J. Pyrcz, J.B. Boisvert, and C.V. Deutsch. 1355  
 1371        A library of training images for fluvial 1356  
 1372        and deepwater reservoirs and associated 1357  
 1373        code. *Computers Geosciences*, 34(5):542– 1358  
 1374        560, 2008. ISSN 0098-3004. doi: <https://doi.org/10.1016/j.cageo.2007.05.015>.  
 1375  
 1376 [86] Misael M. Morales and Michael Pyrcz. 1362  
 1377        GeostatsGuy/MLTrainingImages: Machine- 1363  
 1378        LearningTrainingImages\_v1.0.0, March 1364  
 1379        2023. URL <https://doi.org/10.5281/zenodo.7702128>.  
 1380  
 1381 [87] Knut-Andreas Lie. *An introduction to 1367  
 1382        reservoir simulation using MATLAB/GNU 1368  
 1383        MATLAB*. Springer, 2008.

1369        *Octave: User guide for the MATLAB Reser-*  
1370        *voir Simulation Toolbox (MRST)*. Cambridge  
1371        University Press, 2019.

1372 [88] Ashish Vaswani, Noam Shazeer, Niki Par-  
1373 mar, Jakob Uszkoreit, Llion Jones, Aidan N  
1374 Gomez, Łukasz Kaiser, and Illia Polosukhin.  
1375 Attention is all you need. *Advances in neural*  
1376 *information processing systems*, 30, 2017.

1377 [89] Q. Li and G. Liu. *Risk assessment of the*  
1378 *geological storage of CO<sub>2</sub>: A review*. 2016.  
1379 doi: 10.1007/978-3-319-27019-7\_13. cited By  
1380 39.

1381 [90] R.A. Chadwick, R. Arts, and O. Eiken. 4d  
1382 seismic quantification of a growing co<sub>2</sub> plume  
1383 at sleipner, north sea. *Petroleum Geol-*  
1384 *ogy Conference Proceedings*, 6(0):1385–1399,  
1385 2005. doi: 10.1144/0061385. cited By 188.

1386 [91] R.A. Chadwick and D.J. Noy. History-  
1387 matching flow simulations and timelapse  
1388 seismic data from the sleipner co<sub>2</sub> plume.  
1389 *7th Petroleum Geology Conference [FROM*  
1390 *MATURE BASINS to NEW FRONTIERS]*  
1391 *(London, 3/30/2009-4/2/2009) Proceedings*,  
1392 2:1171–1182, 2010. cited By 31.

1393 [92] Ismael Dawuda and Sanjay Srinivasan. Geo-  
1394 logic modeling and ensemble-based history  
1395 matching for evaluating co<sub>2</sub> sequestration  
1396 potential in point bar reservoirs. *Frontiers in*  
1397 *Energy Research*, 10:867083, 2022.