



Geological Facies modeling based on progressive growing of generative adversarial networks (GANs)

Suihong Song^{1,2,3} · Tapan Mukerji³ · Jiagen Hou^{1,2} 

Received: 18 June 2020 / Accepted: 25 March 2021

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

Geological facies modeling has long been studied to predict subsurface resources. In recent years, generative adversarial networks (GANs) have been used as a new method for geological facies modeling with surprisingly good results. However, in conventional GANs, all layers are trained concurrently, and the scales of the geological features are not considered. In this study, we propose to train GANs for facies modeling based on a new training process, namely progressive growing of GANs or a progressive training process. In the progressive training process, GANs are trained layer by layer, and geological features are learned from coarse scales to fine scales. We also train a GAN in the conventional training process, and compare the conventionally trained generator with the progressively trained generator based on visual inspection, multi-scale sliced Wasserstein distance (MS-SWD), multi-dimensional scaling (MDS) plot visualization, facies proportion, variogram, and channel sinuosity, width, and length metrics. The MS-SWD reveals realism and diversity of the generated facies models, and is combined with MDS to visualize the relationship between the distributions of the generated and training facies models. The conventionally and progressively trained generators both have very good performances on all metrics. The progressively trained generator behaves especially better than the conventionally trained generator on the MS-SWD, MDS plots, and the necessary training time. The training time for the progressively trained generator can be as small as 39% of that for the conventionally trained generator. This study demonstrates the superiority of the progressive training process over the conventional one in geological facies modeling, and provides a better option for future GAN-related researches.

Keywords Facies modeling · Generative adversarial networks · GANs · Progressive growing of GANs · Channel facies · Karst cave

Key Points

- Progressive growing of GANs is used for geological facies modeling, where geological features are learned from large scales to fine scales.
- Multi-scale sliced Wasserstein distance, multi-dimensional scaling plot, facies proportion, variogram, and channel sinuosity, width, and length are proposed as metrics to evaluate GANs.
- Progressive growing of GANs behaves better than conventional training processes in many metrics including computation time.

✉ Jiagen Hou
jghou63@hotmail.com

¹ State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China

² College of Geoscience, China University of Petroleum (Beijing), Beijing 102249, China

³ Stanford University, 367 Panama St, Stanford, CA 94305, USA

1 Introduction

Subsurface geological facies modeling is a very important part of the workflow for accurate assessment of subsurface resources such as groundwater, petroleum, and carbon storage. Geological facies modeling is a process of integrating various observed data (e.g., well data, seismic data, and outcrops) and geological patterns, and predicting the range of spatial geological facies distributions in the subsurface. Many geostatistical methods have been used for facies modeling, including variogram-based methods, multiple points statistics (MPS)-based methods, object-based (Boolean) methods, and process-mimicking methods (see e.g. [1, 2]). These traditional methods have various advantages and disadvantages. Some of them are still under research, such as the development of tree-based direct sampling MPS method [3] and many variations of forward stratigraphic modeling (e.g. [4]).

With the quick improvement of algorithms and enlargement of datasets, deep learning (DL) has been providing robust solutions for many complicated problems that cannot be addressed using conventional methods. These problems can be within any field including geosciences. Tahmasebiet et al. [5] reviewed commonly used machine learning (including deep learning) algorithms and their state-of-the-art applications in various problems of geosciences. These problems include prediction of rock parameters, reconstruction of porous media, modeling of ground water level, soil moisture prediction, carbon leakage management. Widely used DL techniques vary from convolutional neural networks, recursive neural networks, variational autoencoder, to generative adversarial networks, etc. DL techniques can also be combined with conventional statistical methods for better performances, such as the combination of recursive convolutional neural networks with MPS method [6].

From the perspective of DL, geological facies modeling is a typical generative problem, where a generative model is trained to reproduce a probability distribution given many observed samples [7]. Widely used deep generative models include deep sigmoid belief network [8], neural autoregressive density estimation [9], pixel recurrent neural networks (RNN) and pixel convolutional neural networks (CNN) [10], variational autoencoders (VAE) [9, 11], and especially generative adversarial networks (GAN) [12]. Major variants of GAN include deep convolutional GAN (DCGAN) [13], Wasserstein GAN [14, 15], cycle GAN [16], progressive GAN [17], style GAN [18], and bidirectional GAN [19]. These generative models have been successfully used in many areas, including digit generation [9], image generation [10, 17], audio generation [20], domain transformation [16], super-resolution image creation [21], text-to-image translation [22], and object segmentation [23].

Generative models have also been used in the geosciences, especially in the geological facies modeling problem. Laloy et al. [24] used VAE for low-dimensional representation of binary geological models. Mosser et al. [25] and Mosser et al. [26] used DCGAN for generating three-dimensional solid-void structure of porous media and micro-CT-scale oolitic Ketton limestone. Chan and Elsheikh [27] used Wasserstein DCGAN to generate geological facies models.

In the above GAN applications, all layers of GANs were trained concurrently. Karras et al. [17] proposed an alternative to the conventional training process - the progressive growing of GANs or the progressive training process, i.e. training GANs layer by layer, and proved the superiority of this progressive training process over the conventional training process in multiple applications. The focus of this paper is to investigate the progressive GAN training process for geologic facies modeling, and compare the progressive training process with the conventional training process, based on a set of evaluation metrics, including realistic reproduction of facies

distributions and computation time. In addition, we also build two large training datasets for facies modeling. These datasets can be used for many facies modeling-related future works.

This paper is organized as follows. In section 2, we introduce related background materials on GANs. Section 3 presents the workflows and settings on how to train GANs for facies modeling with both the conventional and the progressive training processes. In section 4, we build a set of evaluation metrics for the trained GANs. Section 5 presents the facies model training datasets. Then, in section 6 we compare and analyze the results of the trained conventional GANs and progressive GANs. Finally, conclusions are provided in section 7.

2 Background on GANs

2.1 GAN framework

The framework of GANs was proposed by Goodfellow, Pouget-Abadie [12] to address the generative problem. Given many observed real samples x_r 's from an unknown distribution p_{data} over a high-dimensional space \mathcal{X} , i.e., $x_r \sim p_{data}$, the goal of GANs is to train a generative model that can reproduce a distribution p_G to approximate p_{data} . Figure 1 shows the typical workflow of a GAN. In general, a GAN includes two trainable blocks, a generator (G_θ) and a discriminator (D_φ) (θ and φ are trainable parameters). We define a latent variable Z with a known distribution p_z (e.g., Gaussian) over a low-dimensional space \mathcal{Z} , and define z as a sample from Z , i.e., $z \sim p_z$. The generator G_θ maps z into x_G over the space \mathcal{X} , i.e., $x_G = G_\theta(z)$; the distribution of x_G is p_G , i.e., $x_G \sim p_G$. The discriminator D_φ maps the generated sample x_G and the given data sample x_r into two scalar values, which are called scores; the two scores are abbreviated as s_G and s_r , i.e., $s_G = D_\varphi(x_G)$, $s_r = D_\varphi(x_r)$. The scores evaluate the realism of the inputs of D_φ . The loss function of GANs is based on some type of distance between s_G and s_r (Eq. (1)); in essence, the loss represents the distance between the p_G and p_{data} . We will discuss the loss function in more detail in section 2.2. The discriminator D_φ and the generator G_θ are alternatively trained by maximizing and minimizing the loss (Eq. (1)), until a certain stopping criterion is reached. In practice, usually a batch of z 's and x_r 's are taken as inputs for the training of GANs.

$$\begin{aligned} \min_{G_\theta} \max_{D_\varphi} L(G_\theta, D_\varphi) \\ = \min_{G_\theta} \max_{D_\varphi} \mathbb{E}_{x_G \sim p_G, x_r \sim p_{data}} \text{Dist}(s_G, s_r) \end{aligned} \quad (1)$$

In the above equation, $L(G_\theta, D_\varphi)$ is the loss function, $\text{Dist}(s_G, s_r)$ is some type of distance between s_G and s_r , and \mathbb{E} is the expectation over $x_G \sim p_G$ and $x_r \sim p_{data}$.

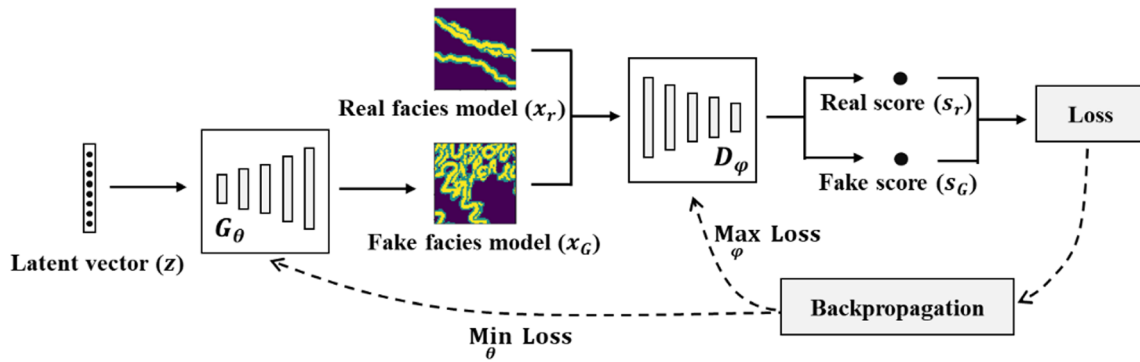


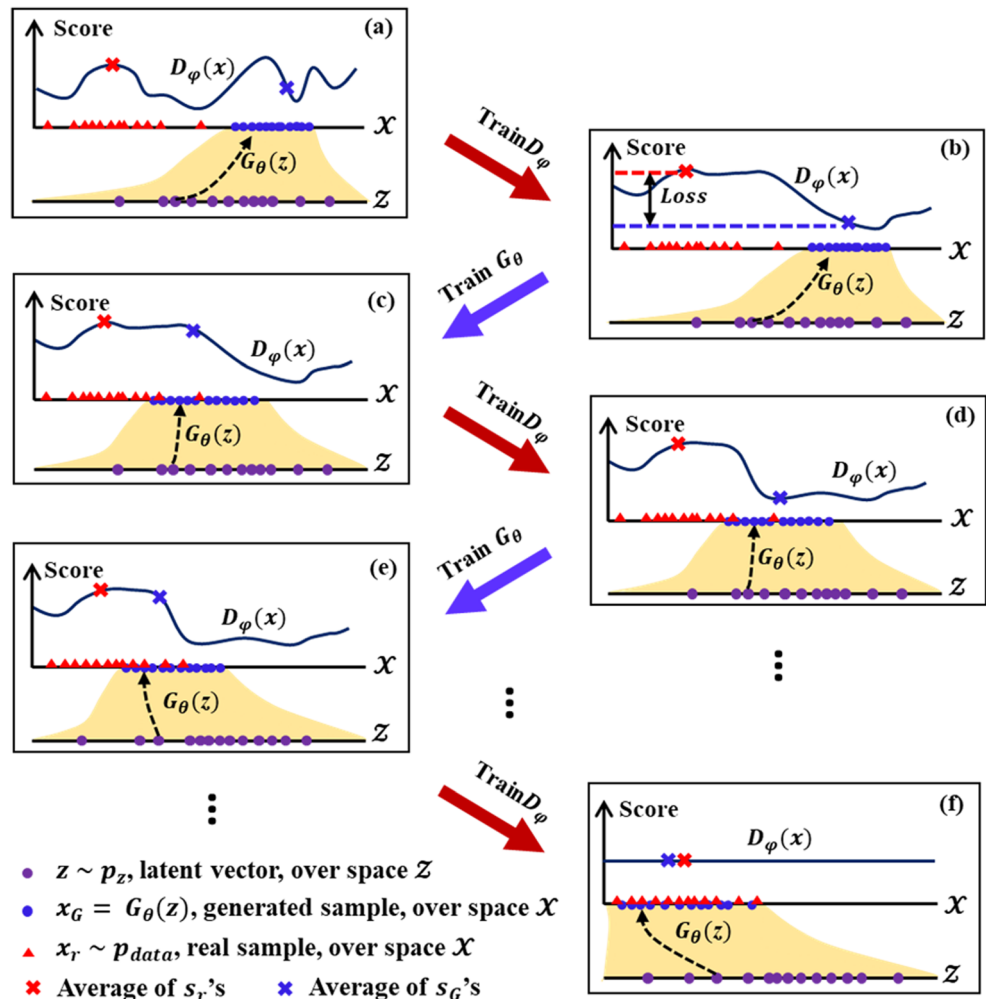
Fig. 1 The basic GAN framework in the context of the geological facies modeling. The generator G_θ maps a latent vector (z) into a “fake” generated facies model (x_G). The discriminator D_ϕ maps x_G and a “real” facies model (x_r) into two scores, one real score and one fake score. The

scores evaluate the realism of the inputs of D_ϕ . Finally, the loss is calculated based on Eq. (1); D_ϕ and G_θ are alternatively trained by maximizing and minimizing the loss

Figure 2 gives an intuitive way to understand the mechanism behind GANs. For better understanding, here we specify the loss function in Eq. (1) as $L(G_\theta, D_\phi) = \|\mathbb{E}_{x_G \sim p_G} s_G - \mathbb{E}_{x_r \sim p_{data}} s_r\|_{L1}$, which is the L1 distance between the expectation of s_G 's and the expectation of s_r 's. The loss function is affected by both $G_\theta(z)$ (the yellow mapping from z to x_G in

Fig. 2) and D_ϕ (the dark blue curve in Fig. 2). Whenever D_ϕ is trained, the dark blue curve is adjusted to be larger on the left and smaller on the right, to increase the loss function, discriminating better between s_G 's and s_r 's, e.g., from (a) to (b) or from (c) to (d) in Fig. 2. On the contrary, when $G_\theta(z)$ is trained, x_G 's are shifted a step closer towards x_r 's, to decrease the loss,

Fig. 2 A schematic illustration of the training of GANs. Training the discriminator (D_ϕ) adjusts the dark blue curve to be larger on the left and smaller on the right, to increase the loss function, to better discriminate the generated samples (x_G 's, blue) from the real samples (x_r 's, red), e.g., from (a) to (b) or from (c) to (d). On the contrary, training the generator ($G_\theta(z)$) maps x_G 's a step closer towards x_r 's, to decrease the loss function, e.g., from (b) to (c) or from (d) to (e). Every successive pair of the training of D_ϕ and G_θ actually pushes x_G 's a step closer to x_r 's, e.g., from (a) to (c) or from (c) to (e); finally x_G 's and x_r 's completely mix in (f), representing $p_G = p_{data}$



e.g., from (b) to (c) or from (d) to (e) in Fig. 2. Every successive pair of the training of D_φ and G_θ pushes x_G 's a step closer to x_r 's, e.g., from (a) to (c) or from (c) to (e) in Fig. 2; finally, x_G 's and x_r 's completely mix, representing $p_G = p_{data}$, e.g., (f) in Fig. 2.

Basically, the generator and the discriminator are functions with trainable parameters, and these functions can be of any form. When GANs were first proposed, the generator and the discriminator were multilayer perceptrons [12]. In recent years, depending on the practical needs, different types of architectures have been proposed for the generator and the discriminator. Generally, if the outputs of the generator are spatially correlated data (e.g., images), the architectures of generator and discriminator are designed to be convolutional neural networks (CNN) (e.g., [13, 28, 29]); if the outputs of the generator are time-related sequence data (e.g., natural language), the architectures of the generator and discriminator may be designed to be recurrent neural networks (RNN) (e.g., [30]). The architecture design of the generator and the discriminator can be very complicated in complicated applications (e.g., [31, 32]).

2.2 GAN loss

Since the GAN framework was proposed, many different types of loss functions have been studied [33], and each of them corresponds to a type of distance between p_G and p_{data} .

Originally, Goodfellow, Pouget et al. [12] proposed the loss function given below in Eq. (2), where the last layer of D_φ is a sigmoid function.

$$L(G_\theta, D_\varphi) = \mathbb{E}_{x_r \sim p_{data}} [\log D_\varphi(x_r)] + \mathbb{E}_{z \sim p_z} [\log(1 - D_\varphi(G_\theta(z)))] \quad (2)$$

This loss function was proved to be equivalent to having a Jensen-Shannon divergence between p_{data} and p_G [12].

Arjovsky et al. [14] showed that the Wasserstein distance is more sensitive than the Jensen-Shannon divergence, and proposed the Wasserstein loss function based on the Wasserstein distance between p_{data} and p_G shown in Eq. (3)

$$L(G_\theta, D_\varphi) = \mathbb{E}_{x_r \sim p_{data}} D_\varphi(x_r) - \mathbb{E}_{z \sim p_z} D_\varphi(G_\theta(z)) \quad (3)$$

where D_φ does not have the sigmoid function in the last layer and should be changing slowly.

Gulrajani et al. [15] improved the Wasserstein loss function by using a gradient penalty to enforce the gradient of D_φ to be small. This improved loss function is shown in Eq. (4).

$$L(G_\theta, D_\varphi) = \mathbb{E}_{x_r \sim p_{data}} D_\varphi(x_r) - \mathbb{E}_{z \sim p_z} D_\varphi(G_\theta(z)) - \lambda \mathbb{E}_{\hat{x} \sim p_{\hat{x}}} \left[\left(\|\nabla_{\hat{x}} D_\varphi(\hat{x})\|_2 - 1 \right)^2 \right] \quad (4)$$

Here, D_φ does not have the sigmoid function in the last layer; λ is a predefined weight, and \hat{x} is sampled between

$x_r \sim p_{data}$ and $x_G \sim p_G$ using $t \sim \text{uniform}(0, 1)$, i.e., $\hat{x} = tx + (1-t)x_G$.

2.3 Progressive growing of GANs

The training of GANs is actually a process of forcing the generator to learn all the features of p_{data} ; the more features are learned, the closer p_G and p_{data} would be ((a) to (c) in Fig. 2); if all features are learned, then $p_G = p_{data}$ ((f) in Fig. 2). Features of a distribution have various scales; for example, in terms of the distribution of human face images, the scale of face gesture feature is larger than the scale of mouth size feature. In most GAN related researches, all layers of GANs are trained concurrently and the scales of features are not considered, so the generator has to manage and learn different scales of features completely by itself; this may result in an inefficient way of feature learning of the generator, e.g., some fine-scale features may be learned earlier than the large-scale features.

Therefore, Karras et al. [17] proposed a new training methodology for GANs: a progressive growing of GANs or a progressive training process, in which the generator and the discriminator are trained layer by layer. The workflow for progressive growing is as follows. First, the original high-resolution training data x_r are downsampled into lower-resolution training data at different resolution levels, e.g., downsampling 1024×1024 training images into 512×512 , $256 \times 256, \dots, 4 \times 4$ training images. These different levels of low-resolution training data represent different scales of features of the original-resolution training data. The lower-resolution downsampled training data capture only the coarse scale features represented in the training data. Second, the shallowest layers of the generator and the discriminator are activated, and the GAN is trained with the lowest-resolution training data. After training, the largest-scale features represented by the lowest-resolution training data are learned by these activated shallowest layers of the generator. Third, the next shallower layer of the generator and the discriminator are further included, and the GAN is trained with the second lowest-resolution training data. After training, the features at the next level of resolution represented by the input training data are learned by the newly activated shallower layers of the generator. Then, increasingly more layers of the generator and the discriminator are progressively included, and larger-resolution training data are used to train the GAN. In this way, the shallower layers of the generator learn the larger-scale features of the training data, while the deeper layers of the generator learn the finer-scale features of the training data. Throughout the training process, all activated layers of both the generator and the discriminator remain trainable. New layers of the generator and the discriminator are included smoothly, meaning their contribution increases from 0 to 1 to avoid disrupting the already well-trained, shallower layers.

Table 1 Architectures of the generator and the discriminator used in this study

Generator							Discriminator			
Block	Layer	Activation	Output shape	No. of trainable parameters	Block	Layer	Activation	Input shape	No. of trainable parameters	
Block 1 (4×4)	Latent vector	–	128×1	–	Block 5 (64×64)	Input image	–	64×64×1	–	
	Fully connected	LReLU	2048×1	262 k		Convolutional 1×1	LReLU	64×64×1	128	
	Reshape	–	4×4×128	–		Convolutional 3×3	LReLU	64×64×64	37 k	
Block 2 (8×8)	Convolutional 3×3	LReLU	4×4×128	148 k	Block 4 (32×32)	Convolutional 3×3	LReLU	64×64×64	74 k	
	Upsampling	–	8×8×128	–		Downsampling	–	64×64×128	–	
	Convolutional 3×3	LReLU	8×8×128	148 k		Convolutional 3×3	LReLU	32×32×128	148 k	
Block 3 (16×16)	Convolutional 3×3	LReLU	8×8×128	148 k	Block 3 (16×16)	Convolutional 3×3	LReLU	32×32×128	148 k	
	Upsampling	–	16×16×128	–		Downsampling	–	32×32×128	–	
	Convolutional 3×3	LReLU	16×16×128	148 k		Convolutional 3×3	LReLU	16×16×128	148 k	
Block 4 (32×32)	Convolutional 3×3	LReLU	16×16×128	148 k	Block 2 (8×8)	Convolutional 3×3	LReLU	16×16×128	148 k	
	Upsampling	–	32×32×128	–		Downsampling	–	16×16×128	–	
	Convolutional 3×3	LReLU	32×32×128	148 k		Convolutional 3×3	LReLU	8×8×128	148 k	
Block 5 (64×64)	Convolutional 3×3	LReLU	32×32×128	148 k	Block 1 (4×4)	Convolutional 3×3	LReLU	8×8×128	148 k	
	Upsampling	–	64×64×128	–		Downsampling	–	8×8×128	–	
	Convolutional 3×3	LReLU	64×64×64	74 k		Minibatch std. dev.	–	4×4×128	–	
Total trainable parameters	Convolutional 3×3	LReLU	64×64×64	37 k	Fully connected	Convolutional 3×3	LReLU	4×4×129	149 k	
	Convolutional 1×1	Linear	64×64×1	65		Fully connected	LReLU	2048×1	262 k	
						1.41 M	Fully connected	Linear	128×1	129
				1.41 M	Total trainable parameters				1.41 M	

***Convolutional 3×3/Convolution 1×1: Convolutional layer with kernel size of 3×3/1×1

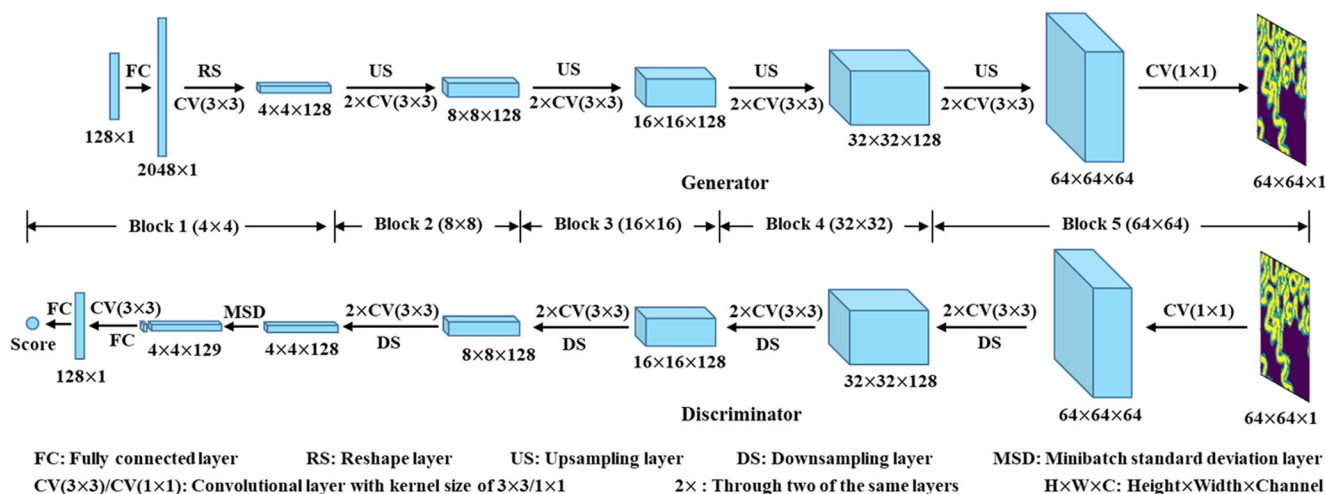


Fig. 3 The architectures of the generator and the discriminator used in this study. The input of the generator is a 128-dimensional vector sampled from the Gaussian distribution, i.e., $z \sim \text{Gaussian}(0, 1)$; the input propagates through the fully connected layer, the reshape layer,

multiple convolutional layers, and multiple upsampling layers; finally the generator outputs a 64×64 2-dimensional facies model. The discriminator takes the 64×64 facies model as input and finally outputs a score to evaluate the realism of the input

The progressive growing of GANs provides a process of learning different scales of features from large scales to fine scales. This philosophy is also well revealed in the multiple-grid geological facies modeling approach, which was proposed by Tran [34] and has been widely used especially in MPS modeling methods (e.g., [35–37]). In the multiple-grid modeling approach, the coarse grid is first simulated to capture large-scale structures from the training image with a coarse template, then the fine grid is simulated, conditioned to the coarse grid simulation, to capture fine-scale structures with a fine-scale template.

3 Geological facies modeling

We train GANs for the geological facies modeling problem, with both the conventional training process and the progressive training process, and compare their results. In this section, first, we introduce the GAN architectures we use in this study; then we describe the settings for the GAN training; finally, we specify the progressive training process, based on the specified architectures of the generator and the discriminator. The

conventional training process is standard (e.g., [14]), so we do not discuss it in detail.

3.1 GAN architectures used in this study

The facies model is spatially correlated, so we use CNN for both the generator and the discriminator. The architecture of the generator is shown in Fig. 3 and Table 1. The input of the generator is a 128×1 latent vector that is sampled from a standard Gaussian distribution, i.e., $z \sim \text{Gaussian}(0, 1)$; the output is a 64×64 facies model. The architecture of the generator includes 1 fully connected neural network layer with 128 input neurons and 2048 output neurons, 1 reshape layer, 4 upsampling layers, and 10 2-dimensional convolutional layers. The reshape layer converts the 2048×1 vector into the $4 \times 4 \times 128$ feature cube, corresponding to 128 feature maps of size 4×4 . Each upsampling layer dilates the heights and widths of the feature cubes by 2, using the nearest-neighbor upsampling method. The kernel size for the last convolution layer of the generator is 1×1 , while the kernel size for the other convolutional layers is 3×3 ; the stride size for all convolutional layers is 1×1 . The leaky rectified linear unit

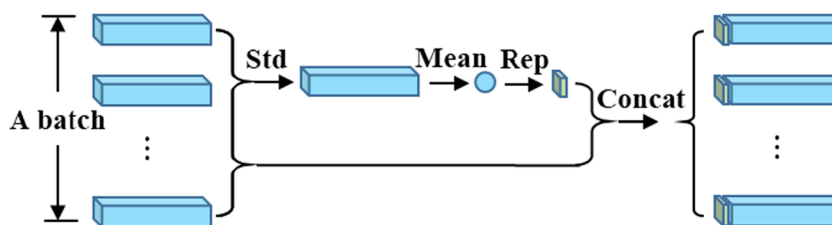
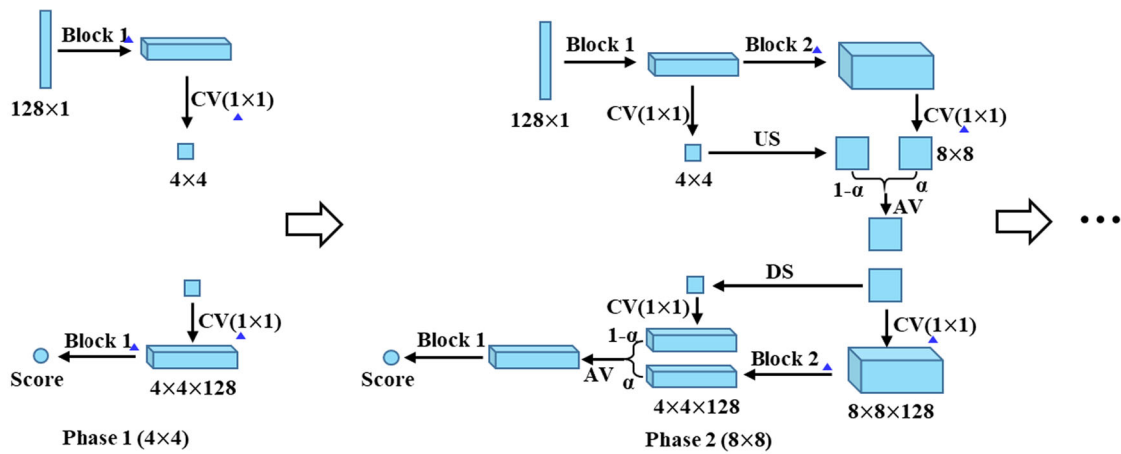


Fig. 4 Illustration of how a mini-batch standard deviation layer works. The batch of feature cubes on the left are calculated from the earlier layers of the discriminator. Each feature cube includes multiple vertical square feature maps. The mini-batch standard deviation layer first calculates the standard deviation for each feature map at each spatial location over the

mini-batch to obtain a standard deviation feature cube. Then the cube of standard deviations are averaged to obtain a single overall average value. This value is replicated into a feature map of the same resolution as other feature maps, and is concatenated to other feature maps of the mini-batch



Block 1 / Block 2: Layers of Block 1 / Block 2 in the generator or the discriminator **US:** Upsampling layer
DS: Downsampling layer **AV:** Weighted average, i.e., $(1-\alpha)x + \alpha y$ **▲:** Trainable parameters are initialized from scratch
CV(1x1): Convolutional layer with kernel size of 1x1

Fig. 5 The progressive GAN training workflow used in this study. In phase 1, the layers of block 1 in the generator and the discriminator in Fig. 3, are trained from scratch with the 4x4-size training facies models. In phase 2, the layers of block 2 in the generator and the discriminator are smoothly included. The trainable parameters of the layers marked with

blue triangle are initialized from scratch, and the remaining trainable parameters are initialized from previous phase. All activated layers in phase 2 are trained with the 8x8-size training facies models. Similar to phase 2, more and more blocks of layers in the generator and the discriminator are trained in successive phases until all layers are trained

function (LReLU) with a leaky value of 0.2 is used as the activation function in all hidden layers except the last for which only the linear function to add bias is used. As seen in Fig. 3 and Table 1, the architecture consists of five blocks of layers producing feature maps with coarser to finer resolution – 4x4, 8x8, 16x16, and 32x32, and finally 64x64. The first fully connected layer is also included in block 1 (4x4) for easier description. These five blocks are trained progressively starting from the coarse 4x4 block to finer and finer blocks ending with 64x64 output image. The progressive training process is described in more detail later.

The architecture of the discriminator is essentially symmetrical to the generator, with corresponding blocks of layers producing feature maps with coarse (4x4) to increasingly fine resolution (8x8, 16x16, 32x32, and 64x64), except that the output of the discriminator is a scalar and a mini-batch standard deviation layer (discussed below) is applied at the discriminator (Fig. 3 and Table 1). The kernel size for the first convolutional layer of the discriminator is 1x1, while the kernel size for the other convolutional layers is 3x3. The mini-batch standard deviation technique was proposed by Karras et al. [17] to increase the variability of the generated results. The calculation steps for the mini-batch standard deviation layer are as follows (Fig. 4): first, calculate the standard deviation for each feature map at each spatial location over the mini-batch; second, average these calculated standard deviation values over all feature maps and spatial locations to obtain a single value; third, replicate the value into an additional feature map with the same resolution as other feature maps; finally, concatenate the new feature map with other feature maps of the mini-batch, increasing the channels (or the

number of feature maps) by one, from 128 to 129 after concatenating the mini-batch standard deviation layer (MSD in Fig. 3). This layer could be inserted anywhere in the discriminator, but it is best to be inserted towards the end [17].

3.2 Loss function training

We use the Wasserstein loss function with gradient penalty (W-gp). During the training, the trainable parameters (i.e., θ and φ) are initialized with He initialization approach [38]. To speed up the training process, mini-batch gradient descent and the Adam optimizer [39] with the default parameters (i.e., $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$) are used. Every mini-batch is set to include 32 facies models. In many studies, the number of optimization steps for the discriminator in each loop is set to be larger than 1, i.e., the training alternates between multiple iterations of optimizing the discriminator and one step of optimizing the generator. In our study, we set the optimization steps of the discriminator to be 1 in each loop, because the Wasserstein loss can largely stabilize the training process. In this work, we call each pair of the optimizations of the discriminator and the following generator as one step of alternative training. We use Tensorflow (tensorflow.org), an open-source deep learning framework, to construct and train our GANs. 2 GPUs (NVIDIA Tesla V100-PCIE-32GB), 10 CPUs, and 80G RAM are used in parallel for the training.

The criterion for stopping the training is important during the training of GANs. Different from other deep learning approaches, the loss in GAN, by definition, does not measure the accuracy, although observations of some experiments show

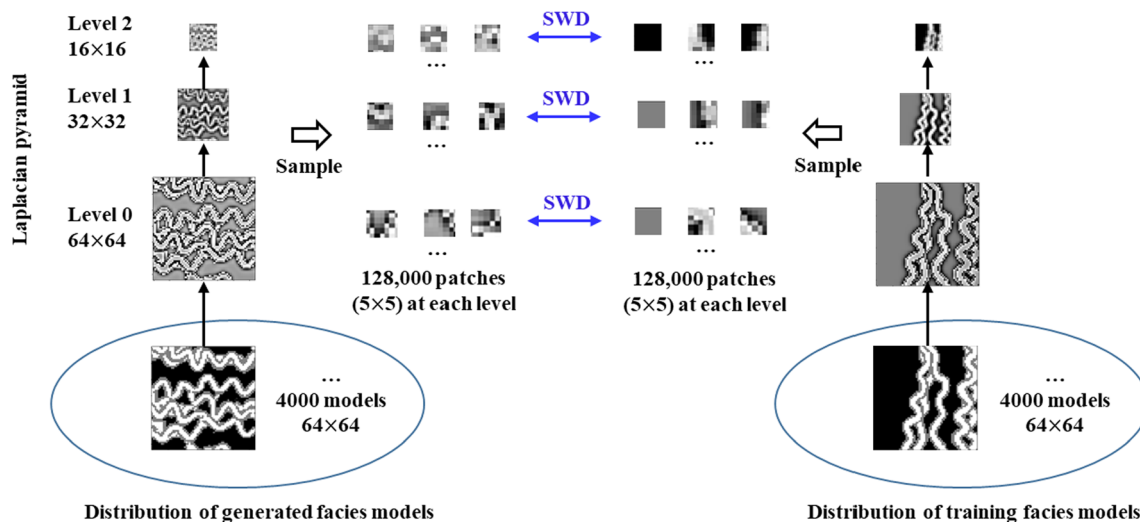


Fig. 6 We apply MS-SWD to assess the distance between the distribution of the generated facies models and the distribution of the training facies models. We calculate the Laplacian pyramid representations for 4000 generated facies models and 4000 randomly selected training models. Next, we randomly sample 32 5×5-pixel patches from the Laplacian

pyramid representation of each facies model at each level. We normalize these patches with the mean and the standard deviation of each patch. Finally, we calculate the SWD between these patches from the generated facies models and from the training facies models, at each level

that the Wasserstein loss reflects the quality of the generated images [15]. We use two evaluation metrics as the stopping criteria, i.e., the multi-scale sliced Wasserstein distance and visual inspection. The multi-scale sliced Wasserstein distance measures the distance between the distribution of the generated results and the distribution of the training data, over multiple scales; visual inspection refers to visually inspecting the generated results during the training process by humans, until the generated results are visually indistinguishable from the training data. The training stops when the multi-scale sliced Wasserstein distance converges and the generated results are visually indistinguishable from the training data. We will discuss the multi-scale sliced Wasserstein distance and the visual inspection in detail in Section 4. We evaluate the generator every 20 k steps of alternative training based on the above criteria to decide when to stop the training process.

3.3 Progressive training workflow

Figure 5 shows the progressive training workflow of this study. Different from the conventional training workflow, we downsample the original 64×64-size training facies models into 32×32, 16×16, 8×8, and 4×4-size training facies models, using averaging. These downsampled facies models and the original ones are used during the following process.

The layers of the generator and the discriminator are progressively trained block by block from coarse to fine (shallow to deep), in different phases. In phase 1, we activate the layers of block 1 (4×4) in the generator and the discriminator (Fig. 5), and add two additional convolutional layers (i.e., “CV(1×1)” in Fig. 5). The two convolutional layers convert the output (the 4×4×128 feature cube) of block 1 in the

generator into a 4×4 facies model and convert a 4×4 facies model, either a training or a generated facies model, into a 4×4×128 feature cube as the input of block 1 in the discriminator. The involved layers are initialized from scratch. We train the activated layers in this phase with 4×4-size training facies models.

In phase 2, we now activate the layers of block 2 (8×8) in the generator and the discriminator, and add two new convolutional layers (“CV(1×1)” in Fig. 5) that convert the output (8×8×128 feature cube) of block 2 in the generator into a 8×8 facies model and correspondingly convert a 8×8 facies model into a 8×8×128 feature cube. To avoid disrupting the already trained layers in the previous phase, the newly activated layers (block 2) are included smoothly in the generator and the discriminator by a weighted average, meaning that the contribution of the new layers increases gradually from 0 to 1. We upsample the 4×4 facies model generated from block 1 in the generator into a 8×8 facies model, and average that 8×8 facies model from block 1 with the 8×8 facies model converted from the output of block 2, with a weighting factor α . The generator output now is thus $(1 - \alpha) \times$ (upsampled 8×8 facies model from block 1) + $\alpha \times$ (8×8 facies model from block 2). In the discriminator, the input 8×8 facies model, either a training or a generated facies model, is both directly converted into a 8×8×128 feature cube and downsampled into a 4×4 facies model. Block 2 in the discriminator takes that 8×8×128 feature cube as input and generates a 4×4×128 feature cube. At the same time, the downsampled 4×4 facies model is converted into another 4×4×128 feature cube by a convolutional layer. These two 4×4×128 feature cubes are then averaged into a new 4×4×128 feature cube with the same weight α , i.e., $(1 - \alpha) \times$ (4×4×128 feature cube converted from 4×4 facies

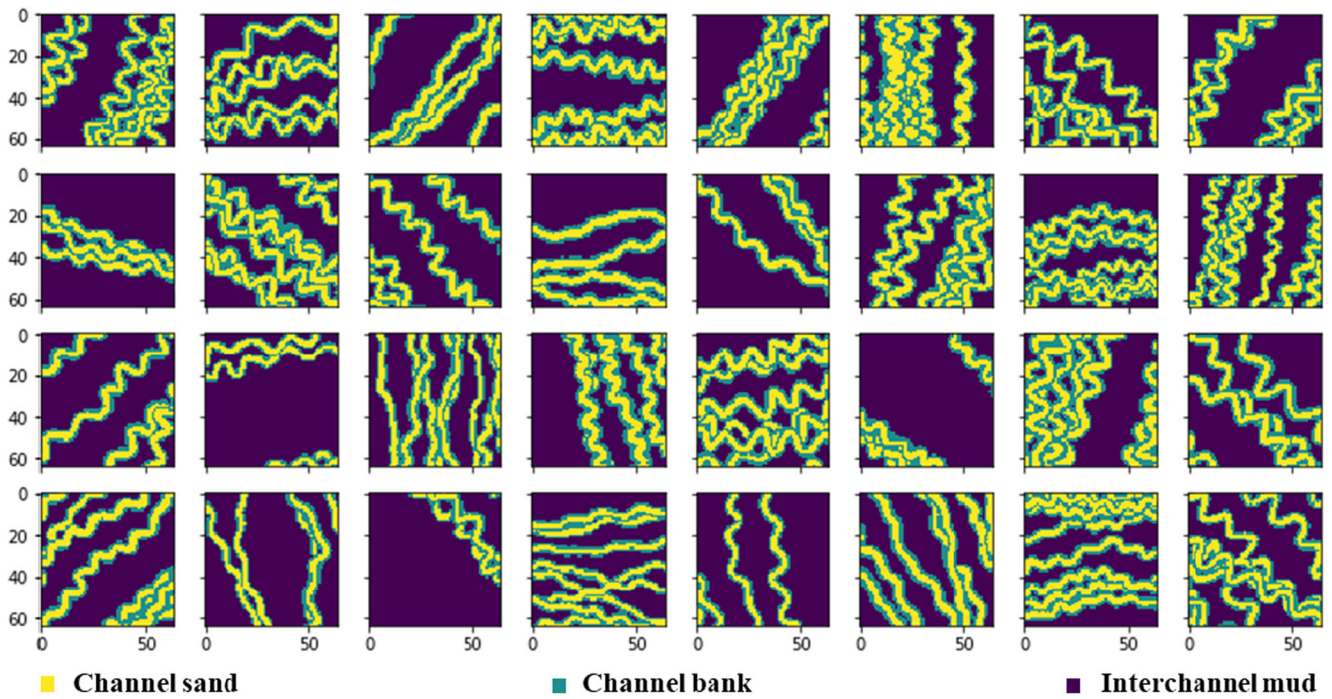


Fig. 7 Some random training facies model examples

model) $+\alpha\times(4\times4\times128$ feature cube generated from block 2). This new $4\times4\times128$ feature cube finally goes through the layers of block 1 in the discriminator to output the score. The weighting factor α represents the contribution of the newly activated layers of block 2. It first changes linearly from 0 to

1, to ensure that the newly activated layers fade in smoothly without breaking the already learned features stored in the previous block. After that, α is held constant at 1. This allows training of all parameters (in the newly activated block as well as the previous block) to the same degree, to enhance the

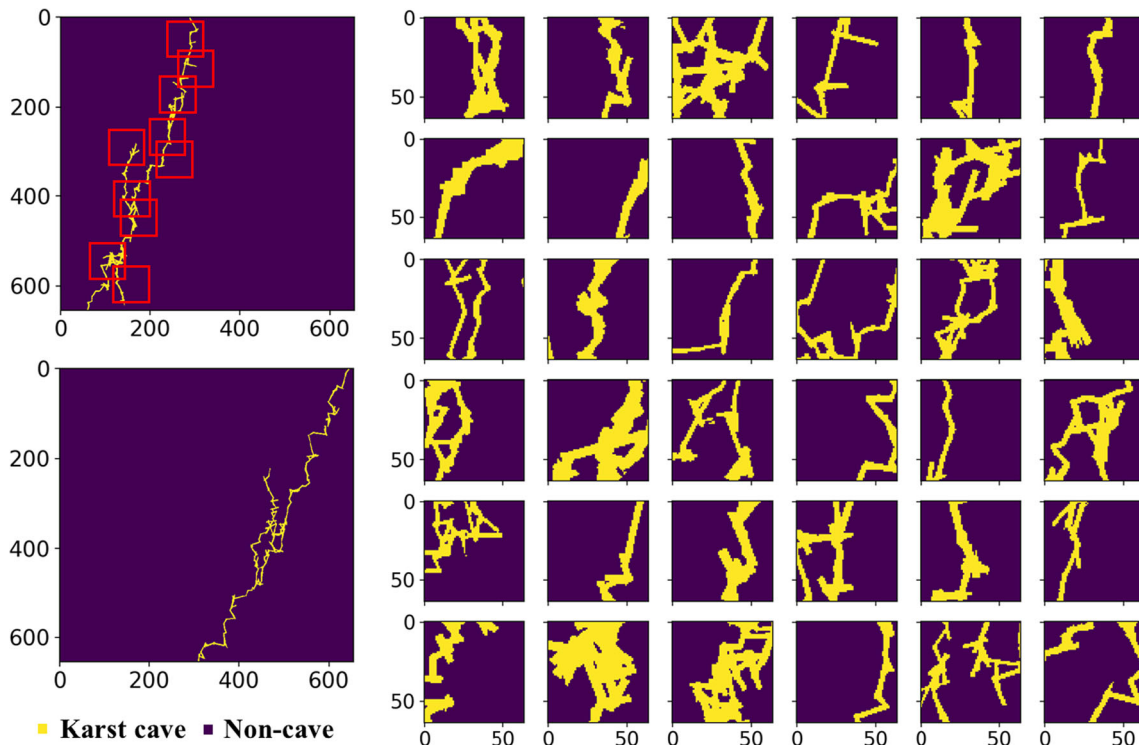


Fig. 8 Random Examples of large karst cave simulations (left) and small cropped training karst cave facies models (right). The red rectangles (64×64) illustrate how the large facies models are cropped into small facies models along karst caves

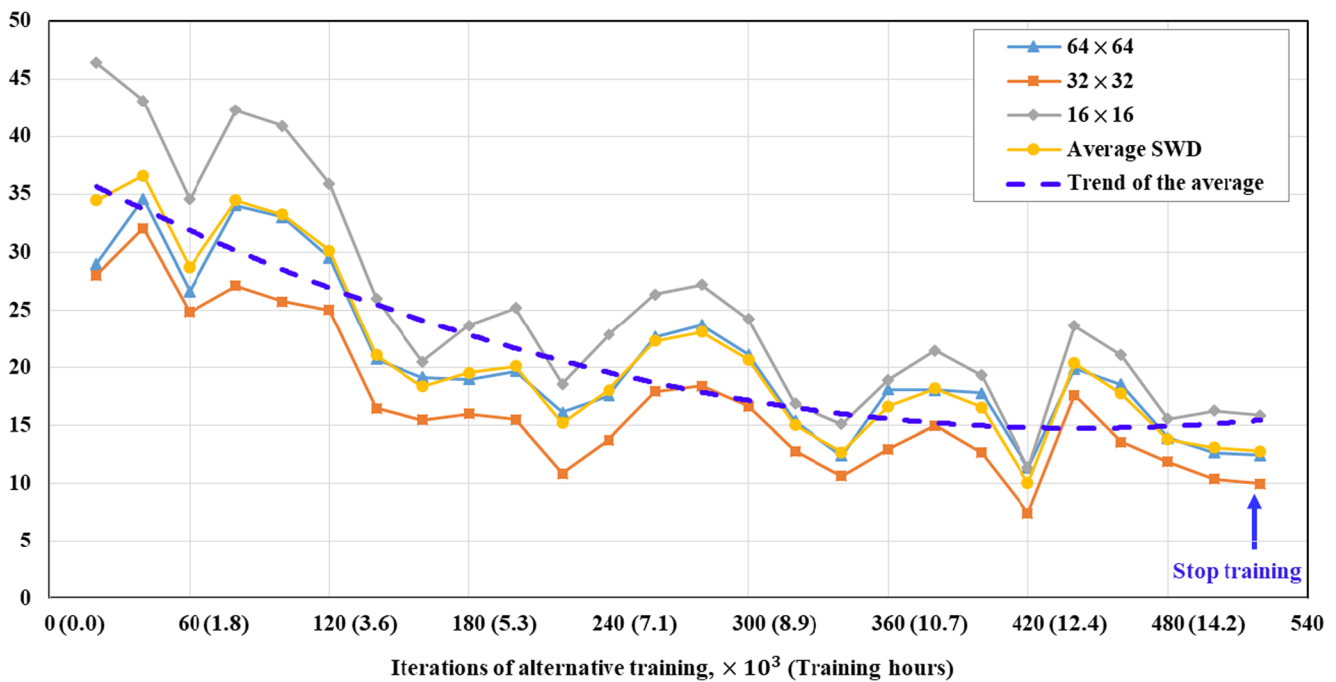


Fig. 9 The change of MS-SWD of the generator during the conventional training process. The trend of the average SWD flattens out after about 460 k alternative training

robustness of the generator and the discriminator. In this phase, the trainable parameters of the newly activated and newly added layers are initialized from scratch, while the other trainable parameters are initialized from the values they reached in the previous training phase. Finally, we train all the activated layers in this phase with 8×8 -size training facies models.

In a similar manner as in phase 2, we progressively train more and more blocks of layers in phase 3 (16×16), phase 4 (32×32), and phase 5 (64×64) until all layers in the generator and the discriminator are trained.

4 Evaluations of GANs

We evaluate the generator by assessing its generated facies models, based on the following metrics described below: visual inspection, multi-scale sliced Wasserstein distance, multi-dimensional scaling plots, facies proportion, variogram, and channel sinuosity, width, and length.

4.1 Visual inspection

Visual inspection of the generated samples is one of the most common and intuitive ways to evaluate GANs [40]. We generate large numbers of facies models, and assess their quality (i.e., realism and diversity) by comparing them with the training facies models. Visual inspection is used both in the training process as a stopping criterion and after the training process to inspect the performance of the trained generator.

Although visual inspection is the simplest approach for GAN evaluation, it is expensive, biased, requires background knowledge, and cannot fully reflect the capacity of the generator [40]. We further use several other quantitative evaluation metrics.

4.2 Multi-scale sliced Wasserstein distance (MS-SWD) and multi-dimensional scaling

Many researchers use the multi-scale structural similarity index (MS-SSIM) [41] or directly the structural similarity index (SSIM) to evaluate the generator's output (e.g., [42, 43]). Karras et al. [17] found that MS-SSIM can reveal large-scale mode collapse reliably but fails to react to smaller effects such as loss of variation in colors or textures, and MS-SSIM does not assess the realism of the generated samples. Therefore, Karras et al. [17] proposed multi-scale sliced Wasserstein distance (MS-SWD) to evaluate the generator, based on the intuition that a successful generator should produce samples whose structure is similar to the structure of the training data over all scales. We apply MS-SWD for our study (Fig. 6) as follows. First, we randomly generate 4000 facies models from the trained generator and randomly select 4000 training facies models. Second, we obtain the Laplacian pyramid representations [44] of both the generated facies models and the training facies models from resolution of 64×64 to 16×16 . The Laplacian pyramid representations reveal the structures of the original facies models at different scales. Third, we randomly extract $32 \ 5 \times 5$ -pixel patches from the Laplacian pyramid representation of each facies model at each level, to obtain

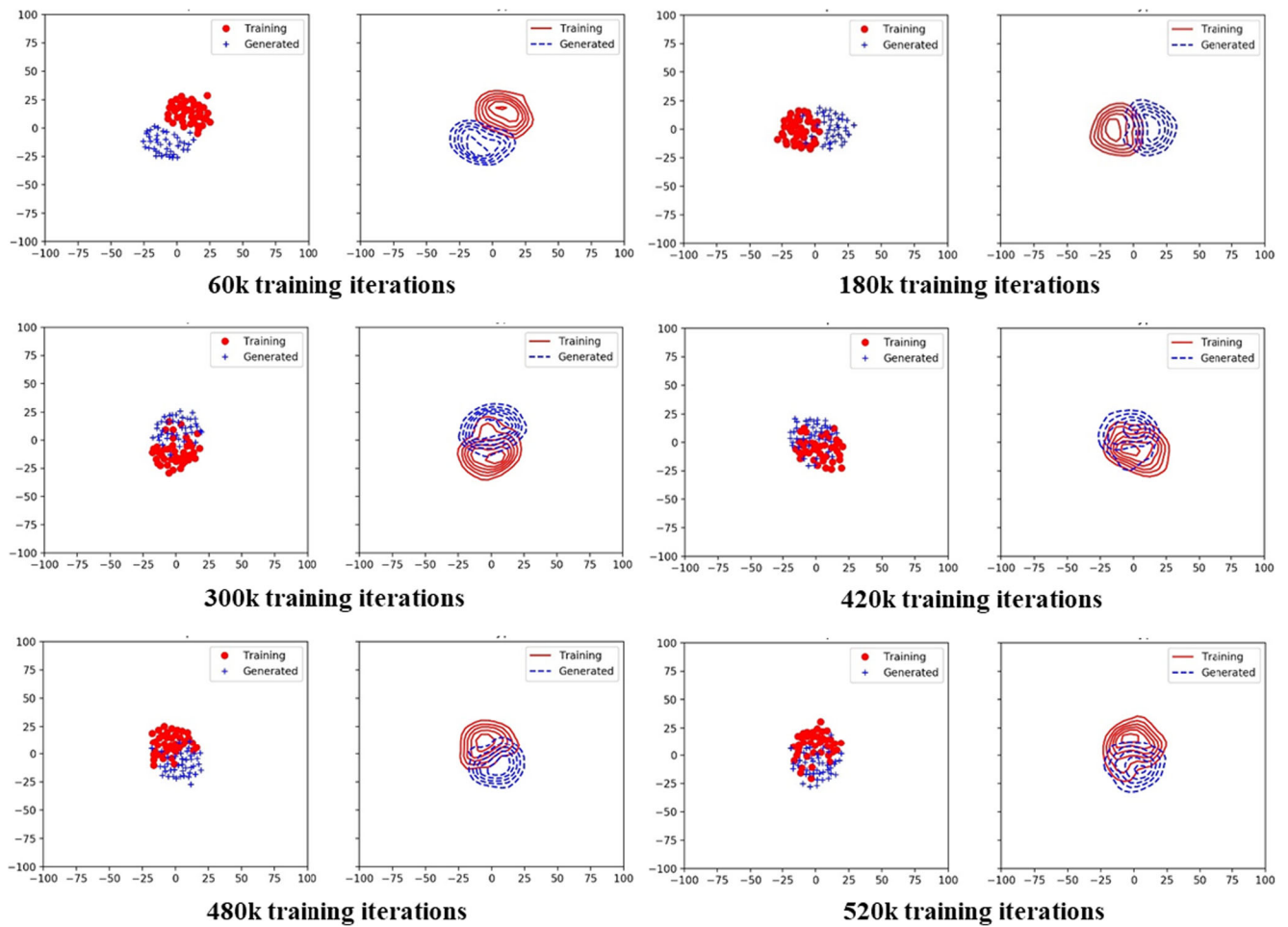


Fig. 10 MDS plots of the relationship between the generated facies models (blue) and the training facies models (red) in 2D space at different iterations of alternative training, during the conventional training process. Each point in this figure represents 40 facies models.

As the training progresses, the generated facies model distribution gets closer to the training facies model distribution, but after 420 k training iterations there is minor or no improvements of the relationship between these two distributions

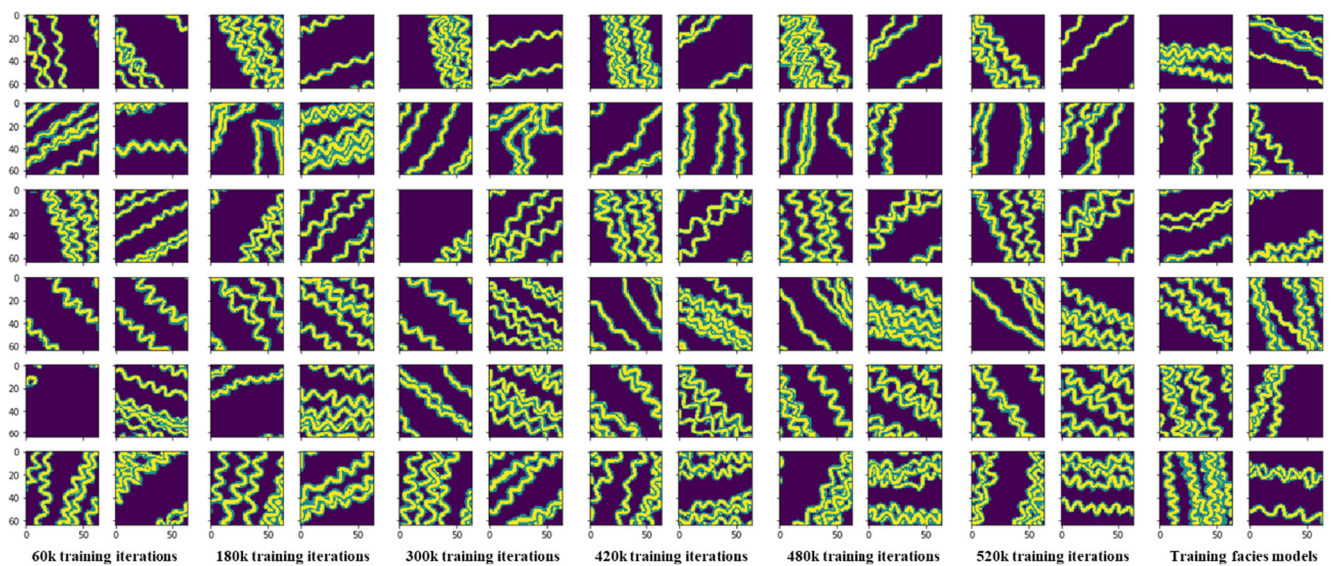
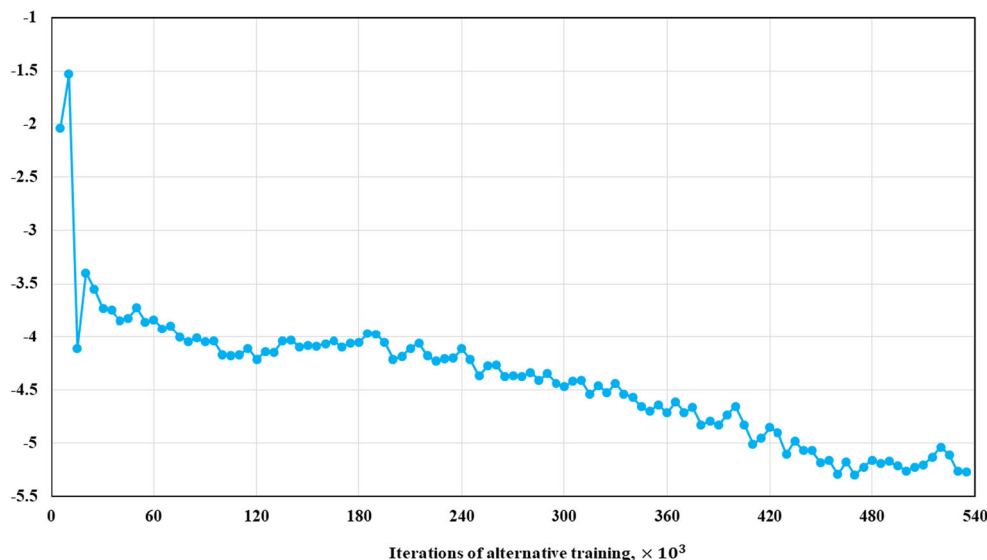


Fig. 11 Transition of the generated facies models with the same input latent vectors after different iterations of alternative training, during the conventional training process

Fig. 12 The W-gp loss versus iterations of alternative training, during the conventional training process

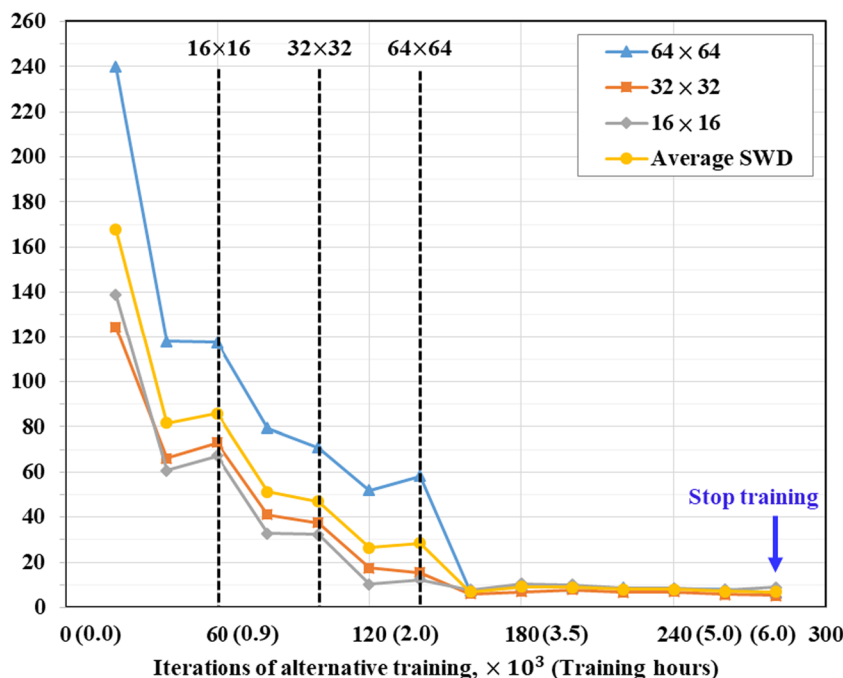


128,000 patches respectively from the generated facies models and the training facies models at each level. Fourth, we normalize these patches with respect to the mean and the standard deviation of each patch. Finally, we calculate the sliced Wasserstein distance (SWD), an efficient approximation to the Wasserstein distance [45], between the patches from the generated facies models and the patches from the training facies models at each level. The smaller the MS-SWD, the closer is the distribution of the generated facies models to the distribution of the training facies models, and better is the performance of the generator with respect to both realism and variation of the generated facies models. We average MS-SWD over different levels to obtain one value to

assess the distance between two distributions; this value is defined as the average SWD in this study. To evaluate the generator more objectively, we also randomly select two groups of the training facies models (4000 in each group) and calculate their MS-SWD, and their average SWD is regarded as the ground truth baseline. In the progressive training, the resolution of the generated facies models varies from 4×4 to 64×64; for the low-resolution (<64×64) facies models, we first upsample them into 64×64 with the nearest-neighbor method, and then calculate the MS-SWD.

Although the calculated MS-SWD can be indicative of the closeness between the distributions of the generated and the training facies models, we may still have no clear sense of

Fig. 13 The change of MS-SWD of the generator during the progressive training process. The MS-SWD converges after 160 k iterations of alternative training. The dashed black lines represent the points when the deeper blocks of 16×16, 32×32, and 64×64-resolution were added to the generator



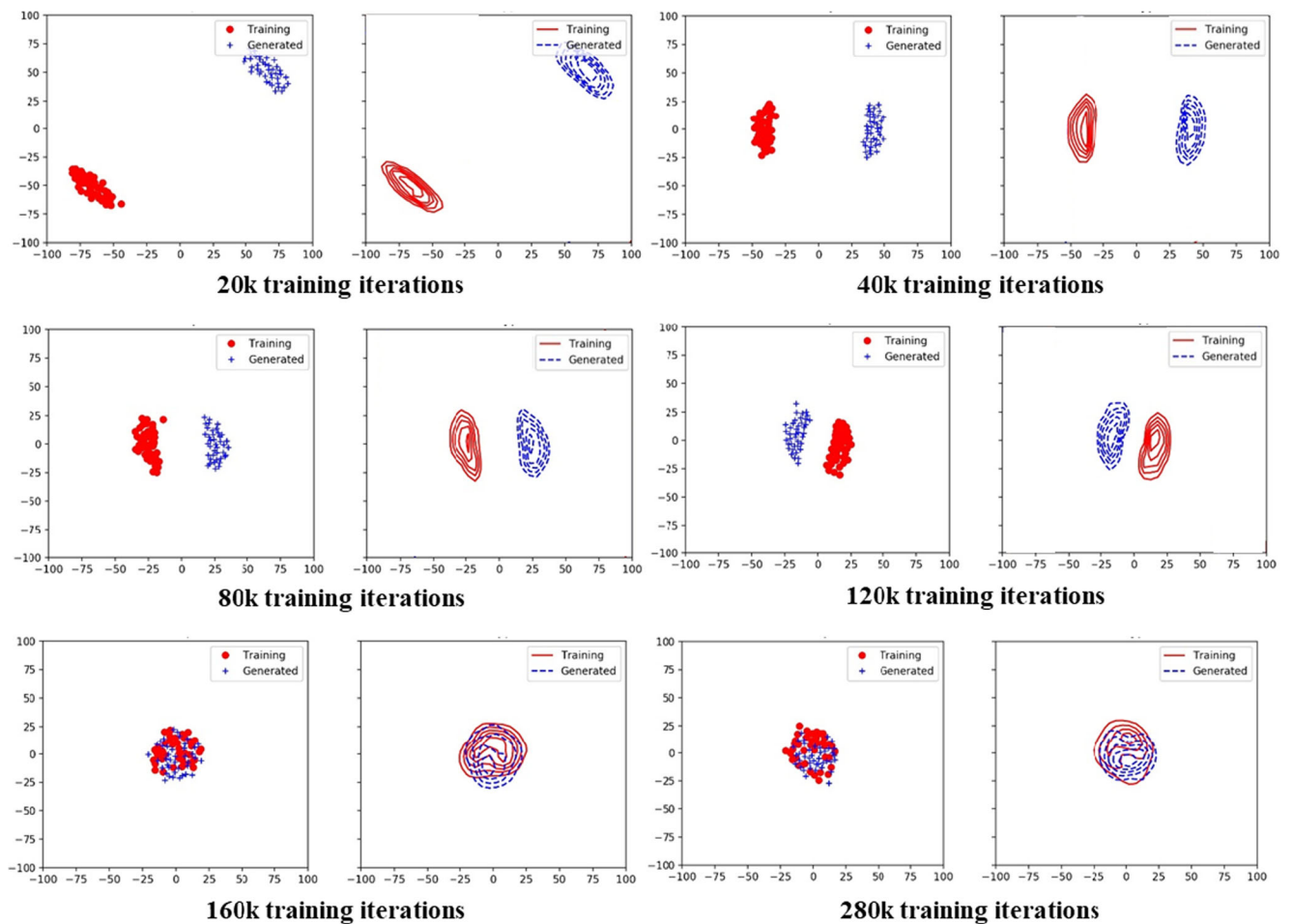


Fig. 14 MDS plots of the relationship between the generated facies models (blue) and the training facies models (red) in 2D space at different iterations of alternative training in the progressive process. As

their spatial relationship. Therefore, we propose a method to visualize the distribution of the generated and the training facies models in 2D space, based on MS-SWD and multi-dimensional scaling (MDS) approach [46]. MDS was also used to evaluate the GAN-generated facies models in [47] but based on the modified Hausdorff distance. We randomly generate 12,000 facies models using the generator and randomly select 12,000 training facies models. We randomly divide the generated and the training facies models into 300 sub-groups of size 40 each, respectively. The average SWD between every two of the 600 groups, is taken as the pairwise distance matrix used for multi-dimensional scaling. We use MDS to approximate the distribution of the 600 groups of the facies models in 2D space. This distribution can visually show the relationship between the generated facies models and the training facies models. In addition, in the same way, we also visualize the distribution of the training facies models, the distribution of the facies models generated from the generator trained in the conventional process, and the distribution of the facies models generated from the generator trained in the progressive process. This MDS plots visualization approach can

the training progresses, the generated facies model distribution gets closer to the training facies model distribution. After 160 k iterations there is no improvement of the relationship between these two distributions

help to compare the performances of GANs trained in different processes.

4.3 Facies proportion

For each facies model, we calculate a proportion of each facies type. The facies proportion distribution of the generated facies models should be close to that of the training facies models, for each facies type. To quantitatively measure this closeness, we define an error as the area between the facies proportion cdf of the generated facies models and that of the training facies models. This error is between 0 and 1. In this study, we sample 3000 generated facies models and 3000 training facies models, to compare their facies proportion distributions over different facies types.

4.4 Variogram

Variogram describes two-point spatial continuity of properties, including the geological facies. We sample 3000 generated and 3000 training facies models, calculate their

variograms along four directions (NS, NE-SW, EW, and SE-NW), and assess the generator by comparing their variograms in different directions.

4.5 Channel sinuosity, channel width, and channel length

We use image processing algorithms to calculate channel sinuosity, channel width, and channel length (arc length) for each facies model. The channel sinuosity is defined as the ratio of channel length to the straight-line length between two end points. We used geodesic distance and Euclidean distance transforms algorithms to calculate the width, arc length, and straight-line length of channels in facies models. We randomly sample 200 generated facies models and 200 training facies models to compare their distributions of channel sinuosity, channel width, and channel length. The channel length reveals the connectivity of channels, to some extent.

5 Training datasets

In terms of the facies model training dataset used for GANs, researchers have used training images (TIs) [48], as well as object-based facies modeling methods (e.g., [29, 42]) to construct their own training datasets. In addition, Nesvold and Mukerji [49] used satellite images as the training dataset. In our study, we synthesized two systematic facies training datasets in terms of different reservoir types: sinuous channel reservoirs and underground river-related karst caves.

The training sinuous channel facies models were synthesized in the commercial Petrel platform using mature object-based modeling method. The facies types include channel complex facies (channel sand and channel bank) and inter-channel mud facies. The resolution of each of the 35,640 2-dimensional facies models is 64×64 . Each facies model includes multiple channel complexes, and these channel complexes have similar global features (e.g., orientation, sinuosity, width, amplitude, etc.). During the synthesizing process, we tune the number, orientation, width, wavelength, and amplitude of the channel complexes, to create a variety of synthesized facies models. Figure 7 shows some facies model examples.

For the karst cave training models, we initially designed an object-based simulation method to produce 642 large karst cave realizations with resolution of 655×655 cells. These realizations vary in cave width, cave direction, and the number of cave branches (Fig. 8). There are totally two facies types: cave and non-cave facies. We randomly cropped these large realizations along karst caves to form 22,695 small 64×64 -resolution karst cave facies models, which will be used as the training facies models in this study. Figure 8 shows the cropping process and some large and small karst cave facies model examples.

6 Results and analyses

6.1 GANs used for sinuous channel facies modeling

The sinuous channel facies model training dataset was used for training the GANs in this case. First, we trained the GAN

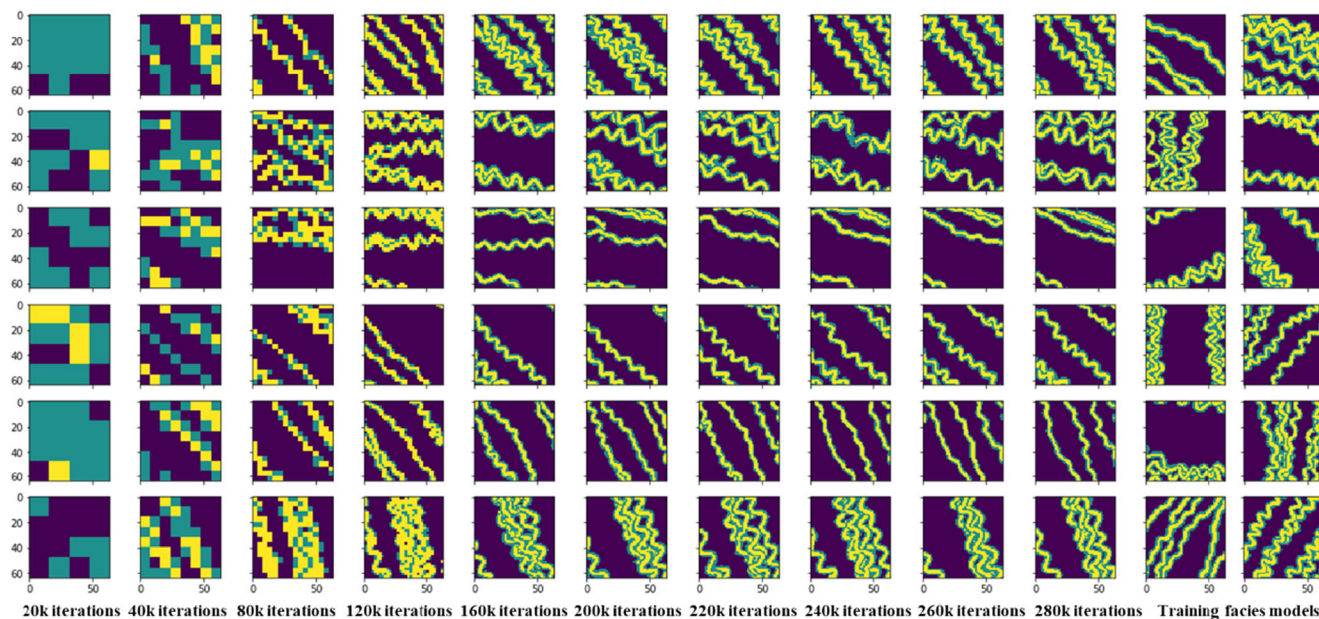
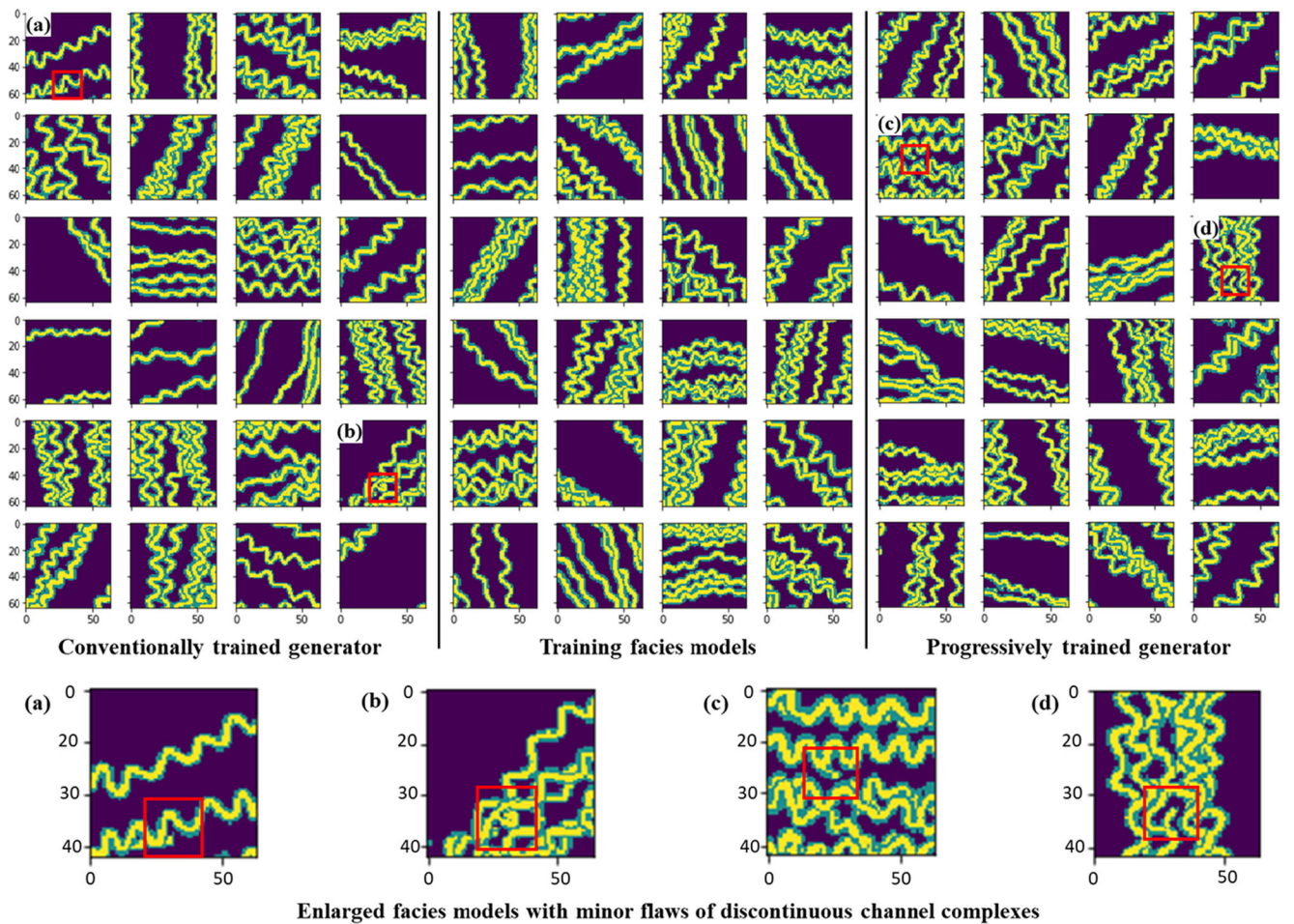
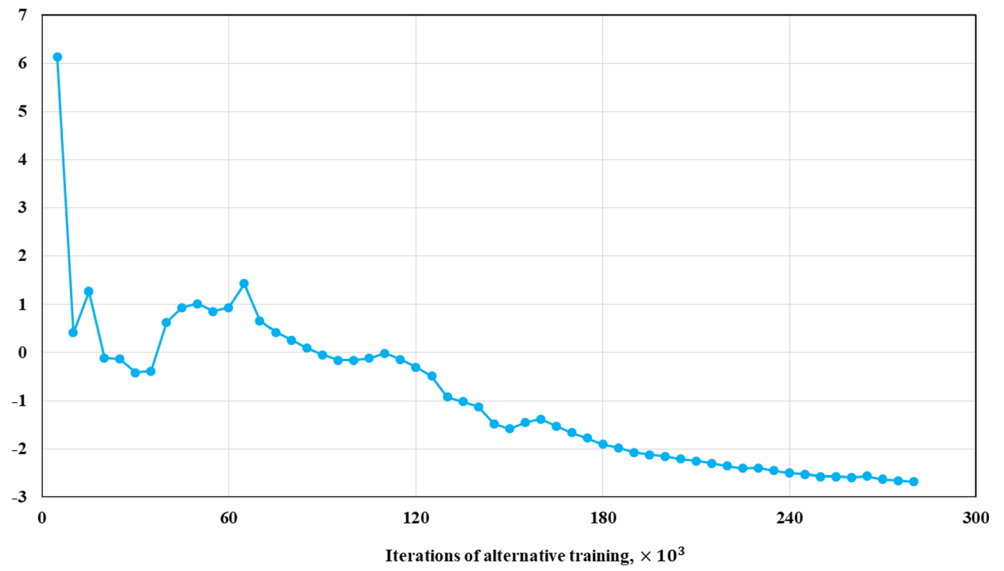


Fig. 15 Transition of the generated facies models with the same input latent vectors at different iterations of alternative training in the progressive process. Facies models are upsampled to 64×64 , if they are smaller than 64×64 . The first 4 columns are upsampled from 4×4 , 8×8 ,

16×16 , and 32×32 into 64×64 . The quality of the generated facies models increases stably, especially at the transitions when the resolution is doubled. After 280 k iterations, the generated facies models are almost indistinguishable from the training data

Fig. 16 The W-gp loss after different times of the alternative training, in the progressive manner



Enlarged facies models with minor flaws of discontinuous channel complexes

Fig. 17 Comparison among the facies models generated from the conventionally trained generator (upper left), the facies models generated from the progressively trained generator (upper right), and the training facies models (upper center). All the generated facies

models are all very similar to the training facies models in both realism and diversity, and only a small number of generated facies models ((a), (b), (c), and (d)) exhibit minor flaws of discontinuous channel complexes

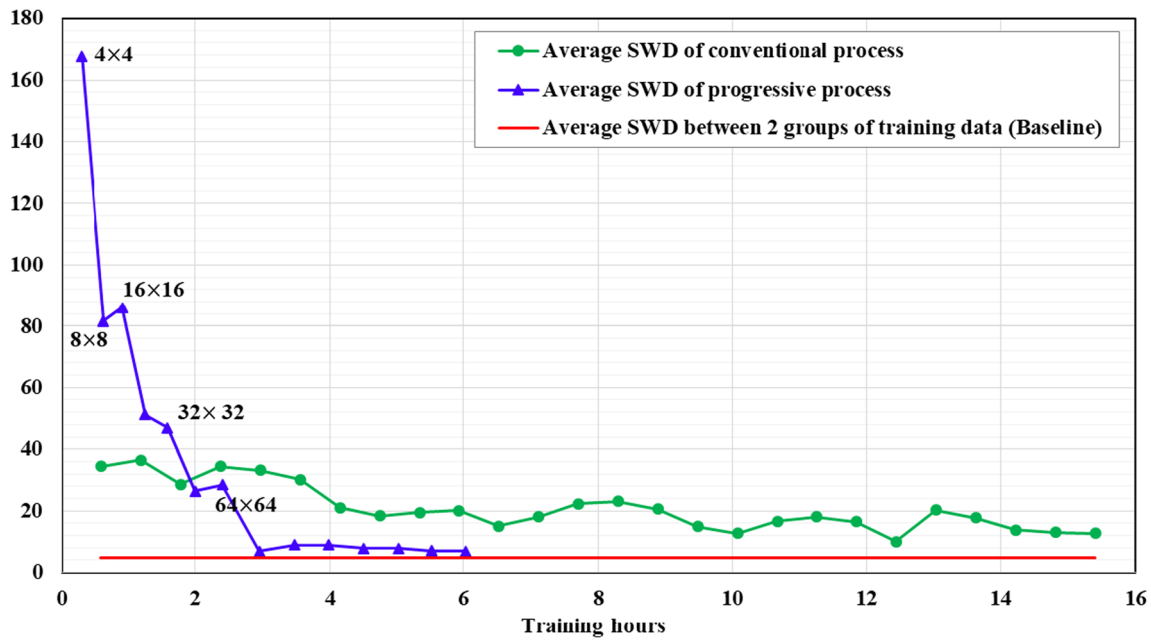


Fig. 18 The change of the average SWDs of the conventional and progressive training, during the training process

in the conventional process. The trend of the average SWD becomes smaller and finally flattens out after about 460 k iterations of alternative training (Fig. 9). Figure 10 shows the MDS plots of the distributions of the generated facies models and the training facies models in 2D space during the training process; after 420 k alternative training iterations, there is minor or no improvement in the relationship between the generated and the training facies models. By further visual inspection, we found that, after 520 k alternative training iterations, the generator can generate very realistic facies models, almost indistinguishable from the training data by human eyes. Figure 11 shows some random generated facies model examples after different iterations of the alternative training. We stopped training after 520 k iterations of alternative training, and kept that final generator for further evaluation and applications. This generator is called the conventionally trained generator hereafter. The total training time is 15.4 h using 2 GPUs (NVIDIA Tesla V100-PCIE-32GB), 10 CPUs,

and 80G RAM in parallel. Figure 12 shows the W-gp loss (see Eq. (4)) versus training iterations; this loss is also called the critic loss in GAN research community.

Next, we trained another GAN with the same architecture but now using the progressive process described above. The MS-SWD converges after about 160 k iterations of alternative training (Fig. 13). Figure 14 further shows in the MDS plots that the generated facies model distribution gets closer to the training facies model distribution as the training progresses, and these two distributions almost totally mix after 160 k iterations. Further visual inspection indicates that, after 280 k iterations the generated facies models are almost indistinguishable from the training facies models (Fig. 15). Therefore, we stopped training after 280 k iterations of alternative training, and kept that generator for further assessment and applications. This generator is called the progressively trained generator hereafter. The total training time is 6.03 h using the same computational capabilities as for the

Fig. 19 The comparison of the training facies models, the facies models generated from the conventionally trained generator, and facies models generated from the progressively trained generator, in 2D space, using MDS plot. The facies models generated from the progressively trained generator is closer in distribution to the training facies models than the facies models generated from the conventionally trained generator

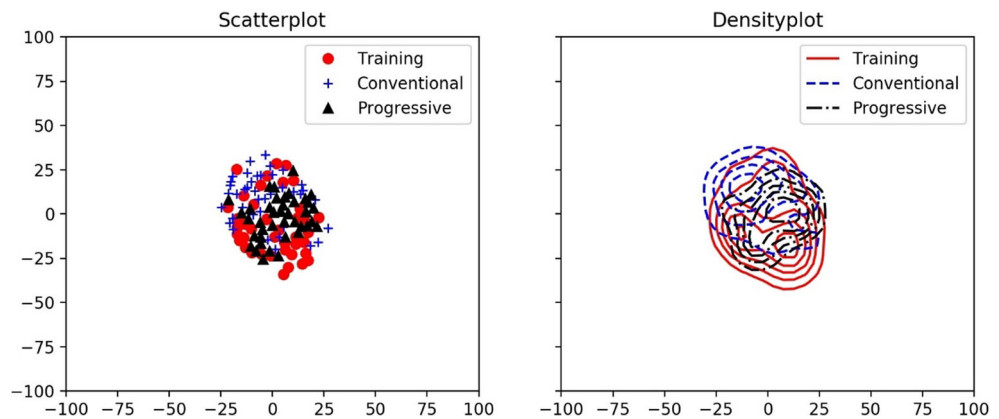


Table 2 Errors in cdf of facies proportion for the conventionally trained generator and the progressively trained generator over all facies types

Generator	Mud facies	Channel bank facies	Channel and facies
Conventionally trained	9.8%	5.6%	6.0%
Progressively trained	6.5%	5.1%	4.6%

conventionally trained GAN. Figure 15 also indicates that the quality of the generated facies models increases stability during the training, especially at the time points when the output resolution is doubled. At these time points, more layers are involved in the generator and the discriminator, so the generator and the discriminator become more robust. The generated facies models after these times points (e.g., the fourth column in Fig. 15) are based on the generated facies models before these times points (e.g., the third column in Fig. 15). The newly added deeper layers in the generator learns more detailed features of the facies models based on the features learned by the shallower layers. Figure 16 shows the W-gp loss (see (4)) versus alternative training iterations.

We compared the conventionally trained generator and the progressively trained generator using visual inspection, MS-SWD, MDS plots, facies proportion, the variogram, and the channel sinuosity, width, and length metrics.

The conventionally and progressively trained generators were both applied to generate 20,000 facies models with one GPU (NVIDIA Tesla V100-PCIE-32GB), respectively, and it takes around 3.2 s for each of the two trained generators. We manually compared these generated facies models with the training facies models, with respect to realism and diversity. The facies models generated from both generators are all very realistic and diversified, and almost indistinguishable from the training facies models, in spite of some minor flaws of discontinuous channel complexes in a small number of generated facies models (Fig. 17). It is difficult to decide which generator is superior based on visual inspection alone.

Figure 18 shows the change of the average SWDs during the conventional and progressive training against the baseline. The average SWD of the conventional training process flattens out slowly, and is farther away from the baseline. Figure 10 indicates that, in the conventional training process,

although the initial generated facies model distribution is not far from the training facies model distribution, the two distributions cannot get very close even after many iterations of alternative training. On the other hand, the average SWD of the progressive training converges very quickly, and the convergence value is very close to the baseline (Fig. 18). The progressive training process starts from low-resolution output and progressively increases the resolution. Hence during the initial 2 h of training, the average SWD of the progressive training is much larger than that of the conventional training process (Fig. 18), and the generated facies model distribution is very far from the training facies model distribution (Fig. 14). During the progressive training, whenever the resolution of the output is doubled, there is a large improvement in the average SWD and in the relationship between the generated and the training facies model distributions. The final generated facies model distribution of the progressive training almost totally mixes with the training facies model distribution (last plot in Fig. 14).

The average SWD of the progressively trained generator (6.81) is much closer to the baseline (4.69) than the average SWD of the conventionally trained generator (12.74) (Fig. 18). Figure 19 uses MDS plot to compare the distribution of the training facies models, and the distributions of the facies models generated from both the conventionally trained and the progressively trained generators, in 2D space. It is clear that the facies models generated from the progressively trained generator is closer to the training facies models than the facies models generated from the conventionally trained generator (Fig. 19, last plot in Fig. 10, and last plot in Fig. 14). In addition, the progressively trained generator is trained in only 39.2% (6.03 h) of the time required to train the conventionally trained generator (15.4 h).

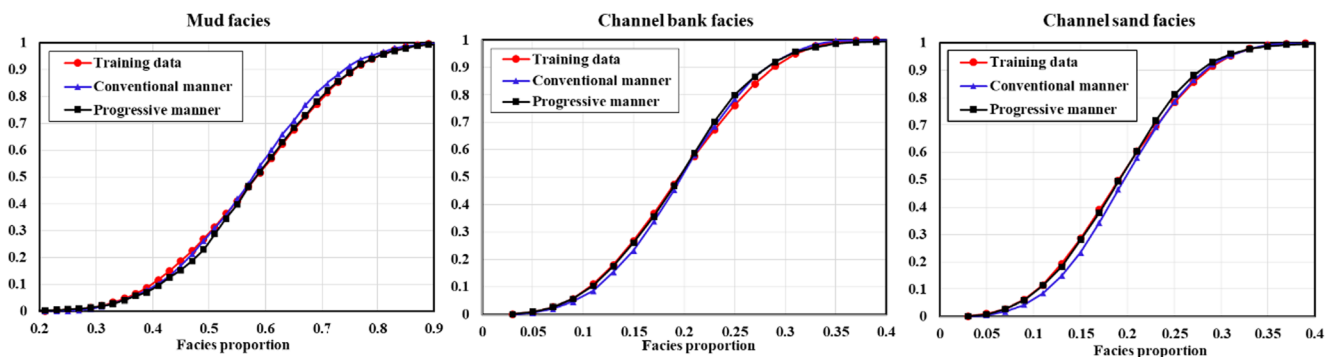


Fig. 20 Facies proportion distributions (cdf's) of the conventionally trained generator, the progressively trained generator, and the training data, over all facies types

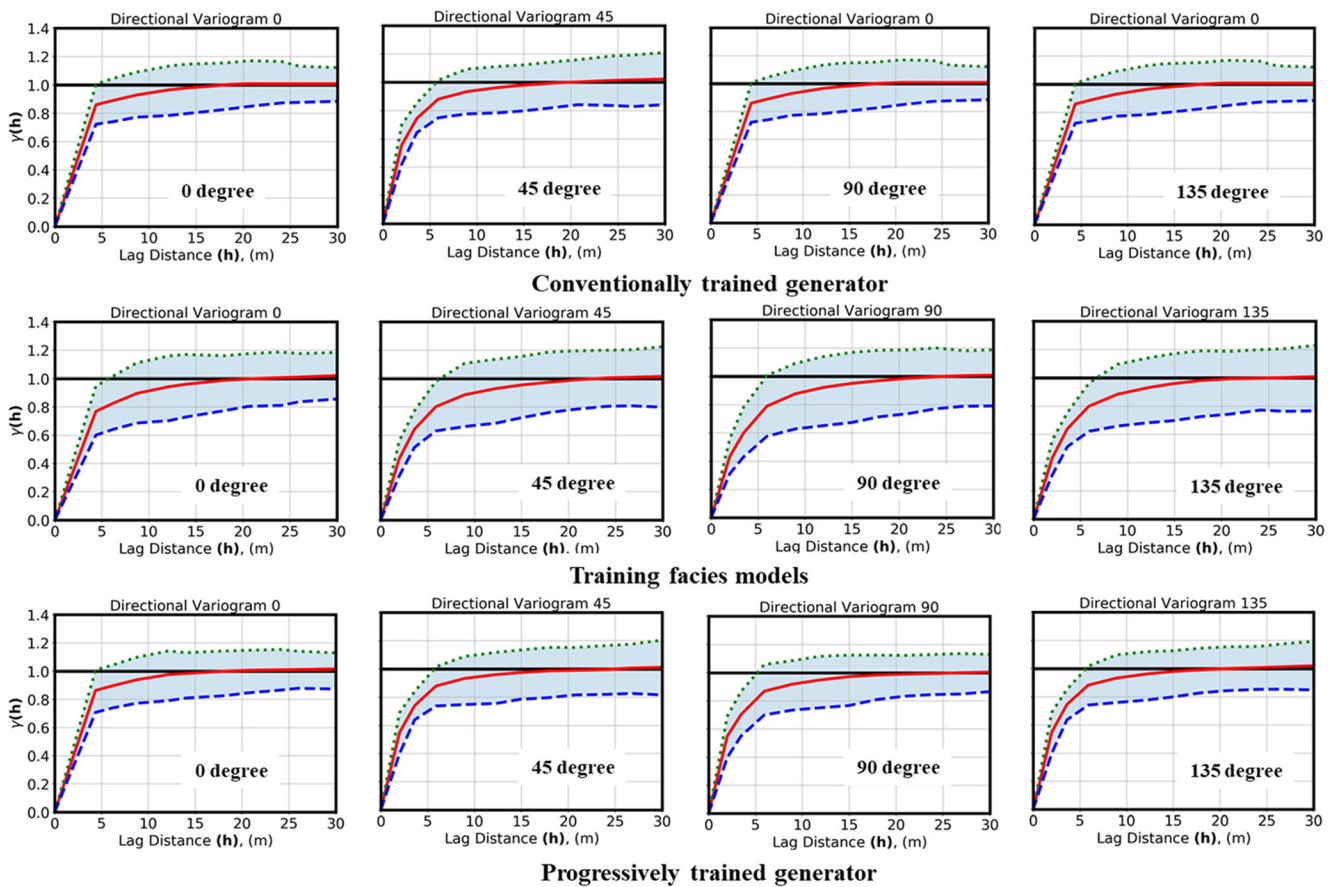


Fig. 21 The variogram distributions of the facies models generated from the conventionally trained and the progressively trained generators along different directions (i.e., 0, 45, 90, and 135 degrees), compared to the variogram distribution calculated from the training facies models. The

red curve is the average, while the dashed blue and dotted green curves are P10 and P90 (i.e. 10th and 90th percentiles) of the variogram distributions

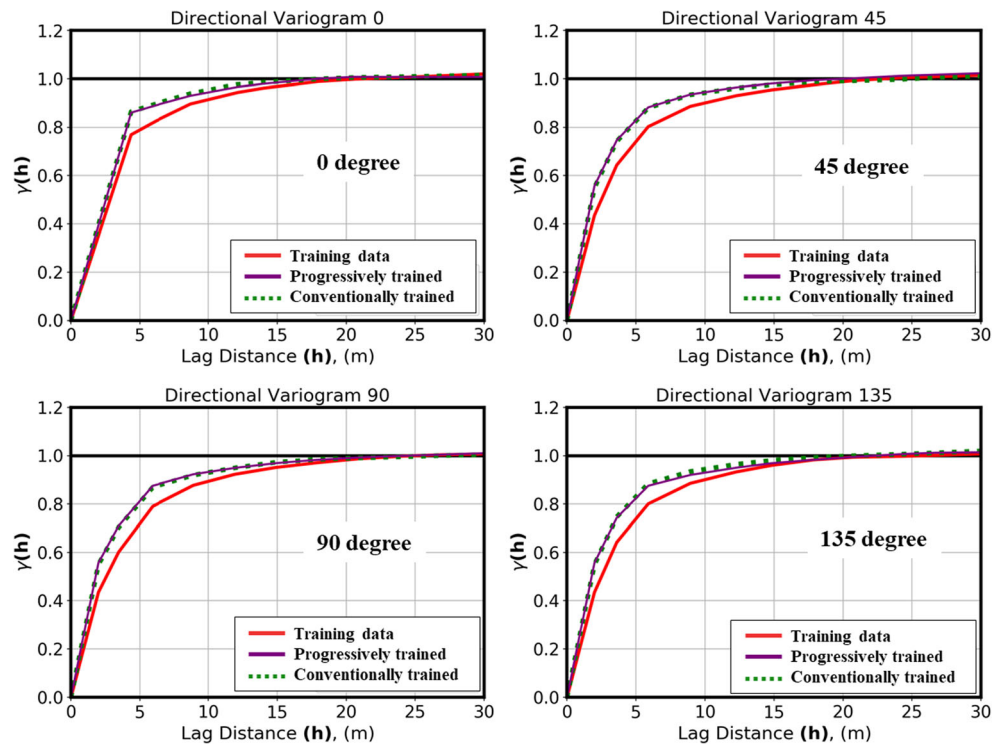
We also evaluated the facies proportion distributions of both the conventionally trained generator and the progressively trained generator, by comparing to the facies proportion distributions of the training data (Fig. 20). The cumulative distribution functions (cdfs) of both the trained generators are very close to the cdfs of the training data, over all facies types. We calculated the facies proportion errors in cdf for the two trained generators, over different facies types (Table 2). The progressively trained generator has slightly lower errors than the conventionally trained generator, over all facies types.

We calculated the variograms of the facies models generated from the conventionally trained generator, the progressively trained generator, and the training facies models, along different directions (Fig. 21). We further compared the average variograms of the generators with that of the training facies models (Fig. 22). The average variograms of both generators are very close to that of the training data, over all directions. The average variograms of the two generators almost overlap; this means that the two generators have similar performance (and flaws) in capturing two-point spatial continuity, to some extent.

We calculated the cdf distributions of channel sinuosity, channel width, and channel length for the randomly sampled training facies models, facies models produced from conventionally trained generator, and facies models from progressively trained generator (Fig. 23). The cdfs of both trained generators are almost equally close to the cdf of the training data, for channel sinuosity, channel width, and channel length. Channel length reveals the connectivity of channels, thus the close channel length distributions in Fig. 23 indicates that, both trained generators produce similar channel connectivity as in training facies models. Looking at the details of the channel length distribution we see that the conventionally trained generator (blue triangles) produces models with a slightly larger proportion of shorter channel lengths as compared to the distribution of channel length from the training set (red circles). This implies that the channel connectivity is slightly poorer for the models from the conventionally trained generator. The models from the progressively trained generator (black squares) show a better match to the training set distribution.

To sum up, the progressively trained generator and the conventionally trained generator both have very good

Fig. 22 The comparison of the average variograms for the conventionally trained generator, the progressively trained generator, and the training facies models, along different directions (i.e., 0, 45, 90, and 135 degrees). The average variograms of two generators almost overlap, and they are both very close to that of the training facies models



performances on all metrics. Once they are trained, it takes 0.16 milliseconds for each generator to simulate a 64×64-resolution facies model. The progressively trained generator behaves especially better than the conventionally trained generator on the MS-SWD metric, MDS plots, and the reduced training time.

We further compare the facies models generated from the progressively trained generator with their closest counterparts in the training dataset (Fig. 24). The closest training facies models are found based on the pixel-to-pixel Euclidean distance. We can see from Fig. 24 that, although the generated facies models are very similar to their counterparts in the training dataset, there are still large differences between them. Thus, we can conclude that, the generated facies models are

not just a naïve copy of the training dataset; instead, they are produced based on the geological pattern knowledge learned from the training dataset.

6.2 GANs used for karst cave facies modeling

Similar to the previous case, we trained two generators with conventional and progressive processes, respectively, using the karst cave training dataset. Based on visual inspection and change of MS-SWD metrics, the conventionally trained generator is obtained after 190 k iterations of alternative training. The training time is 4.7 h using 2 GPUs (NVIDIA Tesla V100-PCIE-32GB), 10 CPUs, and 80G RAM in parallel. Similarly, the progressively trained generator is obtained after

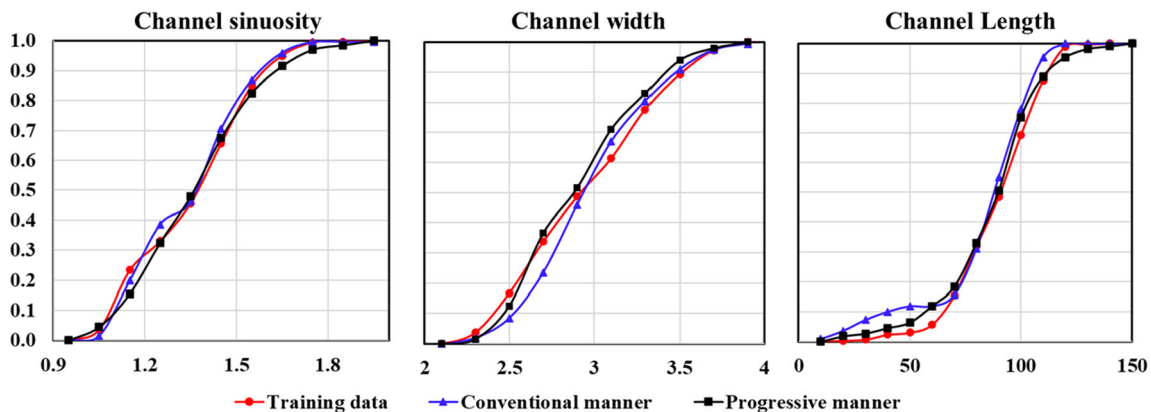


Fig. 23 Channel sinuosity (left), channel width (middle), and channel length (right) distributions (cdf's) of the randomly sampled training facies models, facies models produced from conventionally trained generator, and facies models from progressively trained generator

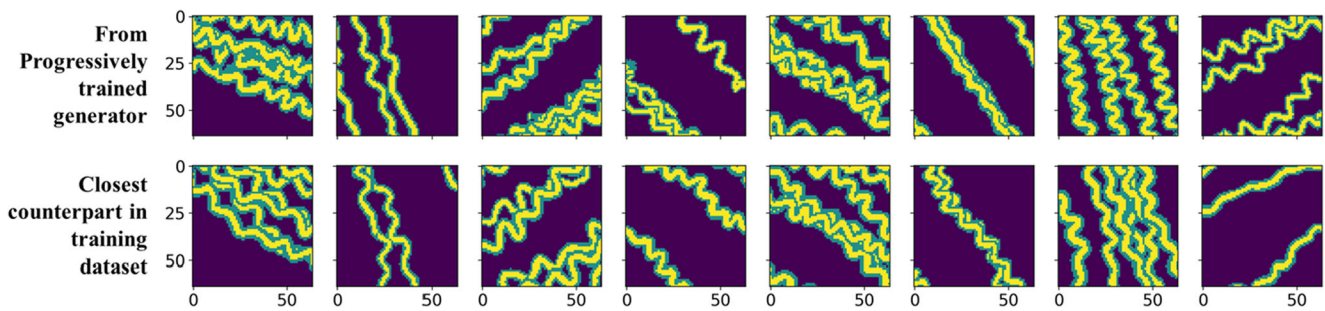


Fig. 24 Comparison between generated facies models with their closest counterparts in the training dataset

240 k iterations of alternative training, and takes 4.4 h with the same computational resources as the conventional process.

We manually compared the karst cave facies models produced by the two trained generators with the training karst cave facies models in realism and diversity (Fig. 25). All the generated facies models are very realistic and diversified, and almost indistinguishable from the training facies models. The average SWD of the progressively trained generator, conventionally trained generator, and the baseline computed from two groups of training facies models are 6.93, 11.75, and 4.69, respectively. That the average SWD of progressively trained generator is much closer to the baseline than the conventionally trained generator, indicates the superiority of the progressive training to some extent. Figure 26 shows the MDS plot to compare the distributions of the training facies models with the distributions of the facies models produced from the two generators, in 2D space. The facies models generated

from the progressively trained generator is closer to the training facies models than those generated from the conventionally trained generator. We further compared the cdf distributions of the karst cave proportion for randomly sampled training facies models and facies models produced from the two trained generators (Fig. 27). The cdf of the progressively trained generator is slightly closer to the cdf of the training data than the conventionally trained generator.

In this case, compared to the conventional training process, the progressive training process only saves about 6% of the training time, while the saved training time in previous case is about 60%. The reason might be that, the karst cave distribution patterns to be learned in this case is easier than the sinuous channel distribution patterns in the previous case, given that generally only one or two karst caves exist in one facies model and that the caves are more aligned along the north-south direction (Fig. 8). With easier feature knowledge to be

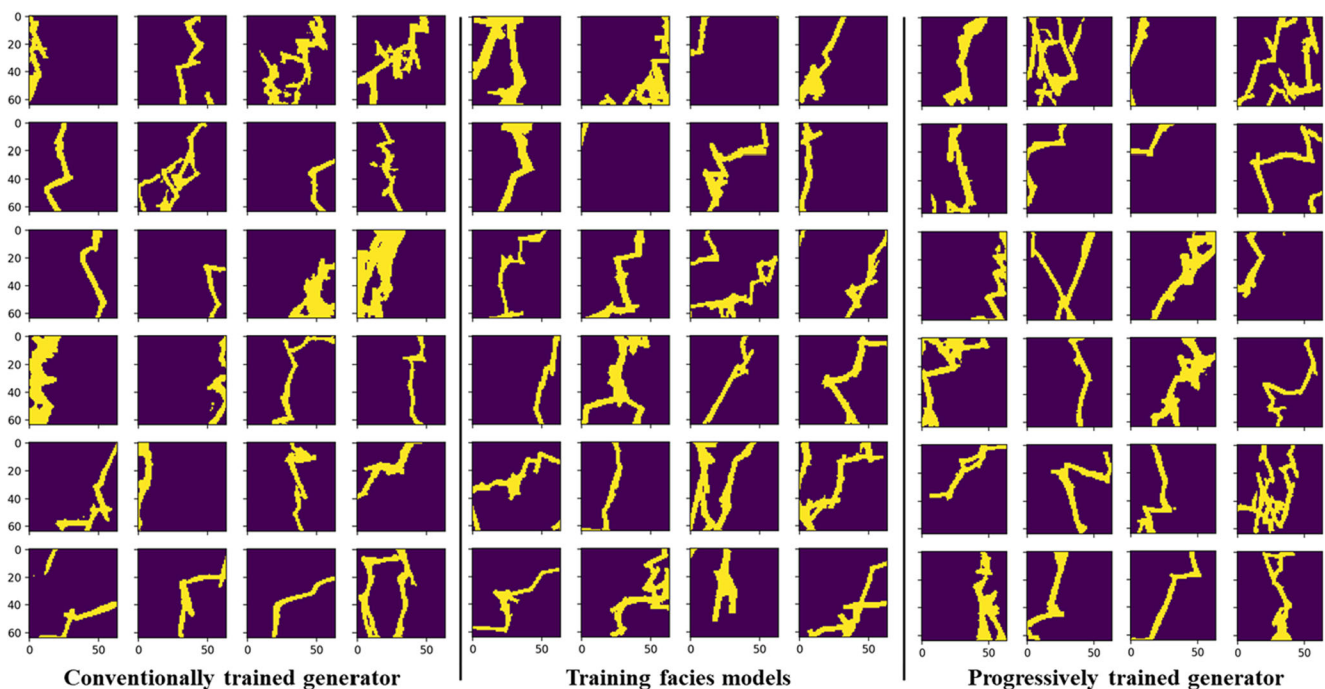
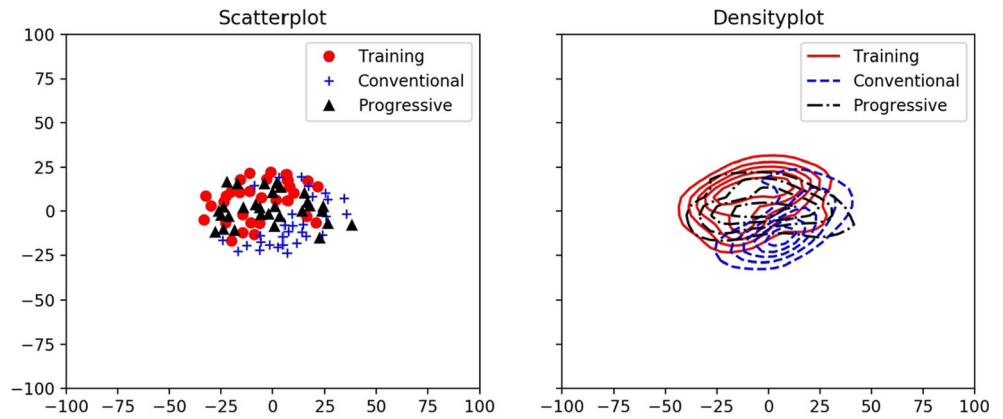


Fig. 25 Comparison among the facies models generated from the conventionally trained generator (left), the facies models generated from the progressively trained generator (right), and the training facies models

(middle). All the generated facies models are very similar to the training facies models in both realism and diversity

Fig. 26 The comparison of the training facies models, the facies models generated from the conventionally trained generator, and facies models generated from the progressively trained generator, in 2D space, using MDS plot. The distribution of facies models generated from the progressively trained generator is much closer to the training facies models than the facies models generated from the conventionally trained generator



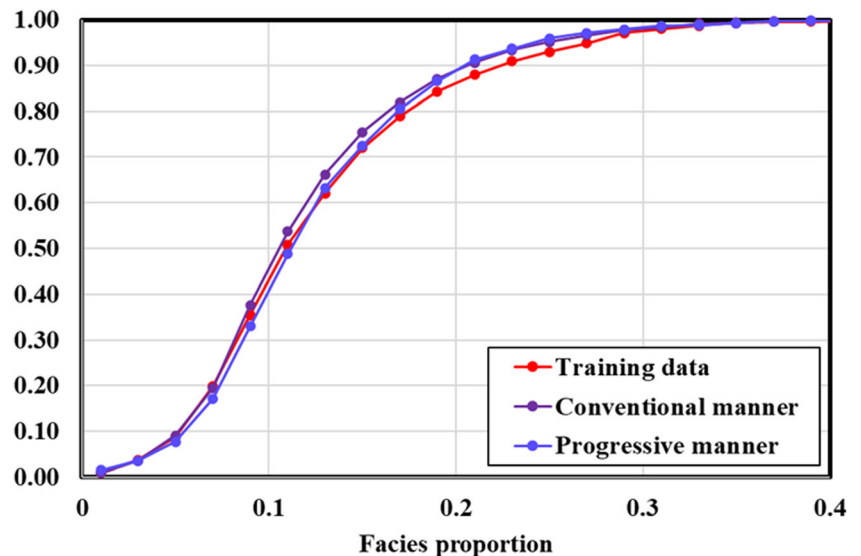
learned, even the direct conventional training process can learn good features quite efficiently, thus the progressive training process cannot show much advantage in at least the training time.

7 Conclusions

Different from the conventional Generative Adversarial Network (GAN) training process, in the progressive GAN training process, or the progressive growing of GANs, GANs are trained layer by layer, allowing features to be learned by the generator from large coarse scales to fine scales. We trained GANs for geological facies modeling using both the progressive training workflow and the conventional training workflow, and assessed the two trained generators. We used visual inspection, multi-scale sliced Wasserstein distance (MS-SWD), multi-dimensional scaling (MDS) plots, facies proportion, variogram, channel sinuosity, channel width, and channel length as the evaluation metrics for the

two trained generators. MS-SWD reveals the realism and variety of the generated facies models over different scales, compared to the training facies models; MS-SWD was also combined with MDS to visualize the distributions of the training and the generated facies models in 2D space. The two trained generators both have very good performances on all metrics. The progressively trained generator behaves especially better than the conventionally trained generator on the MS-SWD metric, MDS plots, and the GAN training time. The training time for the progressively trained generator can be only 39.2% of that for the conventionally trained generator. However, if the features to be learned are simple, then this advantage of the progressive process on training time might be smoothed. Both trained generators can reproduce very good facies proportion, variograms, and channel sinuosity, width, and length that are similar to the training data. Once the generators are trained, it takes 0.16 milliseconds for each generator to simulate a new 64×64-resolution facies model. We also built two large facies model training datasets for this study: one is sinuous channel reservoir type, and the other is karst cave reservoir type. These

Fig. 27 Karst cave proportion distributions (cdf's) of the facies models produced from the conventionally trained generator, the progressively trained generator, and the training data



datasets can also be used for future facies modeling-related works. Future work involves extending the progressive growing of GANs into conditional facies modeling and 3D applications.

Acknowledgments This work was supported by the National Natural Science Foundation of China (42072146). We acknowledge the sponsors of the Stanford Center for Earth Resources Forecasting (SCERF) and support from Prof. Steve Graham, the Dean of the Stanford School of Earth, Energy and Environmental Sciences. Some of the computing for this project was performed on the Sherlock cluster at Stanford University. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. Codes, data, and some results of this work are available at the GitHub site (https://github.com/SuihongSong/GeoModeling_Unconditional_ProGAN).

References

- Pyrcz, M.J., Deutsch, C. V.: *Geostatistical Reservoir Modeling*. (2014)
- Mariethoz, G., Caers, J.: *Multiple-Point Geostatistics: Stochastic Modeling with Training Images*. (2014)
- Zuo, C., Yin, Z., Pan, Z., MacKie, E.J., Caers, J.: A tree-based direct sampling method for stochastic surface and subsurface hydrological modeling. *Water Resour. Res.* **56**, (2020). <https://doi.org/10.1029/2019WR026130>
- Yan, N., Mountney, N.P., Colombera, L., Dorrell, R.M.: A 3D forward stratigraphic model of fluvial meander-bend evolution for prediction of point-bar lithofacies architecture. *Comput. Geosci.* **105**, 65–80 (2017). <https://doi.org/10.1016/j.cageo.2017.04.012>
- Tahmasebi, P., Kamrava, S., Bai, T., Sahimi, M.: *Machine Learning in Geo- and Environmental Sciences: from Small to Large Scale*, (2020)
- Avalos, S., Ortiz, J.M.: Recursive convolutional neural networks in a multiple-point statistics framework. *Comput. Geosci.* **141**, 104522 (2020). <https://doi.org/10.1016/j.cageo.2020.104522>
- Ian Goodfellow, Yoshua Bengio, A.C.: *Deep Learning*. (2015)
- Gan, Z., Henaou, R., Carlson, D., Carin, L.: Learning Deep Sigmoid Belief Networks with Data Augmentation. In: *Journal of Machine Learning Research* (2015)
- Larochelle, H., Murray, I.: The Neural Autoregressive Distribution Estimator. In: *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)* (2011)
- Van Den Oord, A., Kalchbrenner, N., Vinyals, O., Espoholt, L., Graves, A., Kavukcuoglu, K.: Conditional Image Generation with PixelCNN Decoders. In: *Advances in Neural Information Processing Systems* (2016)
- Rezende, D.J., Mohamed, S., Wierstra, D.: Stochastic backpropagation and approximate inference in deep generative models. In: *31st International Conference on Machine Learning, ICML 2014* (2014)
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. In: *Advances in Neural Information Processing Systems 27* (2014)
- Radford, A., Metz, L., Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv e-prints. arXiv:1511.06434 (2015)
- Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. arXiv e-prints. arXiv:1701.07875 (2017)
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.: Improved Training of Wasserstein GANs. arXiv e-prints. arXiv:1704.00028 (2017)
- Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: *Proceedings of the IEEE International Conference on Computer Vision* (2017)
- Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv e-prints. arXiv:1710.10196 (2017)
- Karras, T., Laine, S., Aila, T.: A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv e-prints. arXiv:1812.04948 (2018)
- Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M., Courville, A.: Adversarially Learned Inference. arXiv e-prints. arXiv:1606.00704 (2016)
- Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K.: WaveNet: A Generative Model for Raw Audio. arXiv e-prints. arXiv:1609.03499 (2016)
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv e-prints. arXiv:1609.04802 (2016)
- Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative Adversarial Text to Image Synthesis. arXiv e-prints. arXiv:1605.05396 (2016)
- Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-Image Translation with Conditional Adversarial Networks. arXiv e-prints. arXiv:1611.07004 (2016)
- Laloy, E., Hérault, R., Lee, J., Jacques, D., Linde, N.: Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Adv. Water Resour.* **110**, 387–405 (2017). <https://doi.org/10.1016/j.advwatres.2017.09.029>
- Mosser, L., Dubrulle, O., Blunt, M.J.: Reconstruction of three-dimensional porous media using generative adversarial neural networks. *Phys. Rev. E.* **96**, (2017). <https://doi.org/10.1103/PhysRevE.96.043309>
- Mosser, L., Dubrulle, O., Blunt, M.J.: Stochastic reconstruction of an Oolitic limestone by generative adversarial networks. *Transp. Porous Media.* **125**, 81–103 (2018). <https://doi.org/10.1007/s11242-018-1039-9>
- Chan, S., Elsheikh, A.H.: Parametrization and generation of geological models with generative adversarial networks, (2017)
- Wu, J., Zhang, C., Xue, T., Freeman, W.T., Tenenbaum, J.B.: Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. arXiv e-prints. arXiv:1610.07584 (2016)
- Zhang, T.F., Tilke, P., Dupont, E., Zhu, L.C., Liang, L., Bailey, W.: Generating geologically realistic 3D reservoir facies models using deep learning of sedimentary architecture with generative adversarial networks. *Pet. Sci.* **16**, 541–549 (2019). <https://doi.org/10.1007/s12182-019-0328-4>
- Mogren, O.: C-RNN-GAN: Continuous recurrent neural networks with adversarial training. arXiv e-prints. arXiv:1611.09904 (2016)
- Ma, L., Jia, X., Sun, Q., Schiele, B., Tuytelaars, T., Van Gool, L.: Pose Guided Person Image Generation. arXiv e-prints. arXiv:1705.09368 (2017)
- Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., Metaxas, D.: StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. arXiv e-prints. arXiv:1612.03242 (2016)

33. Lucic, M., Kurach, K., Michalski, M., Gelly, S., Bousquet, O.: Are GANs Created Equal? A Large-Scale Study. arXiv e-prints. arXiv: 1711.10337 (2017)
34. Tran, T.T.: Improving variogram reproduction on dense simulation grids. *Comput. Geosci.* **20**, 1161–1168 (1994). [https://doi.org/10.1016/0098-3004\(94\)90069-8](https://doi.org/10.1016/0098-3004(94)90069-8)
35. Strebelle, S.: Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* **34**, 1–21 (2002). <https://doi.org/10.1023/A:1014009426274>
36. Liu, Y.: Using the Snesim program for multiple-point statistical simulation. *Comput. Geosci.* **32**, 1544–1563 (2006). <https://doi.org/10.1016/j.cageo.2006.02.008>
37. Arpat, G.B., Caers, J.: Conditional simulation with patterns. *Math. Geol.* **39**, 177–203 (2007). <https://doi.org/10.1007/s11004-006-9075-3>
38. He, K., Zhang, X., Ren, S., Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: IEEE International Conference on Computer Vision. pp. 1026–1034 (2015)
39. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv Prepr. arXiv1412.6980. (2014). <https://doi.org/10.1109/ICCE.2017.7889386>
40. Borji, A.: Pros and Cons of GAN Evaluation Measures. arXiv e-prints. arXiv:1802.03446 (2018)
41. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. pp. 1398–1402 (2003)
42. Mosser, L., Dubrule, O., Blunt, M.J.: Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. *Math. Geosci.* **52**, 53–79 (2020). <https://doi.org/10.1007/s11004-019-09832-6>
43. Zhong, Z., Sun, A.Y., Jeong, H.: Predicting CO₂ plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. *Water Resour. Res.* **55**, 5830–5851 (2019). <https://doi.org/10.1029/2018wr024592>
44. Burt, P.J., Adelson, E.H.: Readings in Computer Vision: Issues, Problems, Principles, and Paradigms. Presented at the (1987)
45. Rabin, J., Peyré, G., Delon, J., Bernot, M.: Wasserstein barycenter and its application to texture mixing. In: Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., and Bronstein, M.M. (eds.) *Scale Space and Variational Methods in Computer Vision*. pp. 435–446. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
46. Scheidt, C., Caers, J.: Representing spatial uncertainty using distances and kernels. *Math. Geosci.* **41**, 397–419 (2009). <https://doi.org/10.1007/s11004-008-9186-0>
47. Azevedo, L., Paneiro, G., Santos, A., Soares, A.: Generative adversarial network as a stochastic subsurface model reconstruction. *Comput. Geosci.* **24**, 1673–1692 (2020). <https://doi.org/10.1007/s10596-020-09978-x>
48. Laloy, E., Hérault, R., Jacques, D., Linde, N.: Training-image based Geostatistical inversion using a spatial generative adversarial neural network. *Water Resour. Res.* **54**, 381–406 (2018). <https://doi.org/10.1002/2017WR022148>
49. Nesvold, E., Mukerji, T.: Geomodeling Using Generative Adversarial Networks and a Database of Satellite Imagery of Modern River Deltas. In: *Petroleum Geostatistics* (2019)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.