

Supplementary information for ‘Generative inverse modeling for improved geological CO₂ storage prediction via conditional diffusion models’

We here provide further details regarding the network architecture (Section S1) and the training protocol, including data preprocessing and hyperparameter settings (Section S2). We further present additional evaluation and comparison of the CGAN and CDM on 2D case (Section S3) and 3D case (Section S4).

S1. Detailed information on network architecture

Fig. S1 shows the detailed architecture of each block. The conv block consists of two convolutional layers, group normalization, and the GELU activation function. This block is utilized in both downsample and upsample blocks. The downsample block employs a max pooling layer and two conv blocks to process data from the previous layer, using the SILU activation function and a linear layer to extract features from the embeddings. Two tensors are then added together and passed to the next block. The self-attention block uses two skip connections to handle data from either the downsample or upsample blocks. The first connection adds the output from the attention layer to the reshaped input and then passes it through the forward layer, which includes layer normalization, a linear layer, and the GELU activation function. The second connection adds the output from the forward layer to the attention layer’s output, followed by a reshaping operation to produce the final tensor. The upsample block is a little more complicated. Initially, it uses an upsample layer to process data from the previous block, concatenating the output with data from the skip connection. This combined tensor is then sent through two conv blocks. Additionally, the upsample block uses the SILU activation function and a linear layer to extract features from the embeddings. Finally, the tensors from the conv block and the linear layer are added together to form the output.

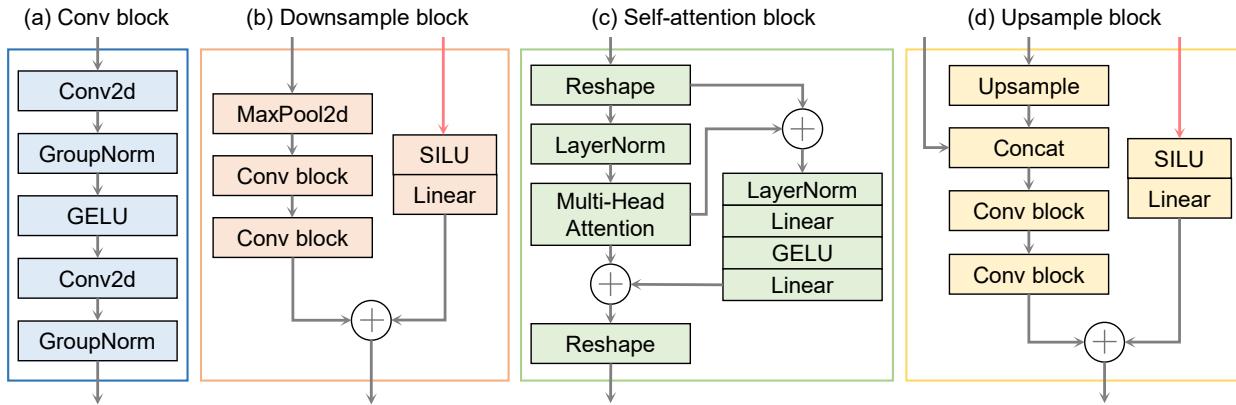


Figure S1: Detailed architecture of (a) Conv block, (b) Downsample block, (c) Self-attention block, and (d) Upsample block.

S2. Training protocol

Before training the neural networks, we first pre-process the training data as follows. We apply a min-max normalization to transform all data \mathbf{x} to the range $[-1, 1]$, that is

$$x_i \leftarrow \frac{2 \times [x_i - \min(\mathbf{x})]}{\max(\mathbf{x}) - \min(\mathbf{x})} - 1 \quad (\text{S1})$$

where the min and max operators are applied across all corresponding data points. For the log-permeability fields, we consider all corresponding pixel values in the entire training dataset. For each type of observed data, we consider the minimum and maximum recorded flow responses for all simulation time steps in the entire training dataset.

We provide the key training hyperparameters in Table S1.

Table S1: Training hyperparameters.

Hyperparameter	Value
Training epoch	600
Batch size	16
β_1	0.0001
β_T	0.02
Learning rate	0.0002
Optimizer	Adam
Diffusion steps	1000
Guidance weight	3

S3. 2D case

In this section, the realization 2 and realization 3 shown in Fig. 4 of the main article are used as reference fields for evaluation and comparison of CGAN and CDM. Fig. S2 and Fig. S3 show the mean and standard deviation (std) of the log-permeability fields estimated using both methods. We can see that the posterior mean fields obtained from the CGAN and CDM overall capture the high-permeability and low-permeability regions of the reference fields, with the CDM mean field being more consistent with the reference field. The statistical results of the *RMSE* values for 100 realizations further indicate the superiority of the CDM over the CGAN. In addition, it can be observed that the CGAN leads to almost a collapse of the ensemble standard deviation, that is, the realizations in posterior ensemble are very similar to each other. The CDM results in larger values of standard deviation, which means that there are more distinct log-permeability realizations in the posterior ensemble.

S4. 3D case

In this section, two realizations randomly selected from the test set are used as reference fields for comparison of the quality of the 3D log-permeability fields estimated by the CGAN and CDM. Fig. S4 and Fig. S5 show the mean of the estimated log-permeability fields and the absolute error between the mean field and the reference field, respectively. We can see that compared to the CGAN, the mean field obtained from the CDM aligns more closely with the reference field, with the major features of the reference field have been captured. This further corroborates our conclusion.

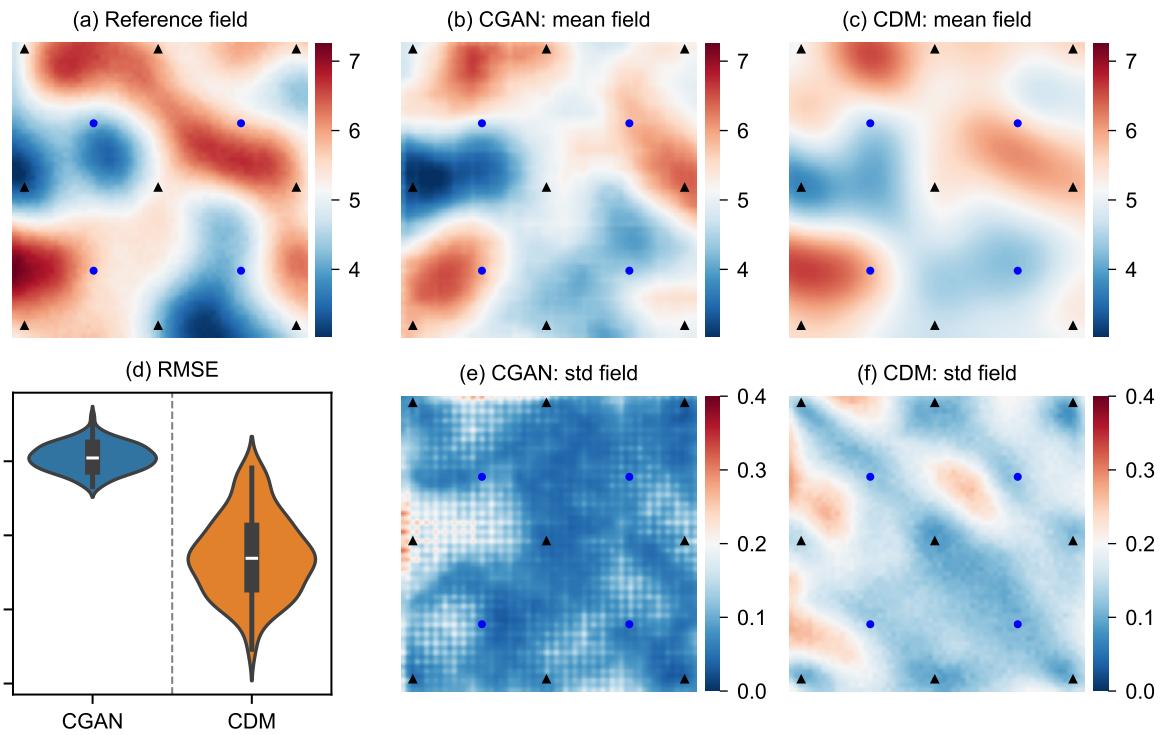


Figure S2: (a) Reference log-permeability field. (b-c) Mean of the log-permeability fields estimated by the CGAN and CDM, respectively. (d) Comparison of the RMSE values for 100 realizations obtained from the CGAN and CDM. (e-f) Standard deviation (std) of the log-permeability fields estimated by the CGAN and CDM, respectively.

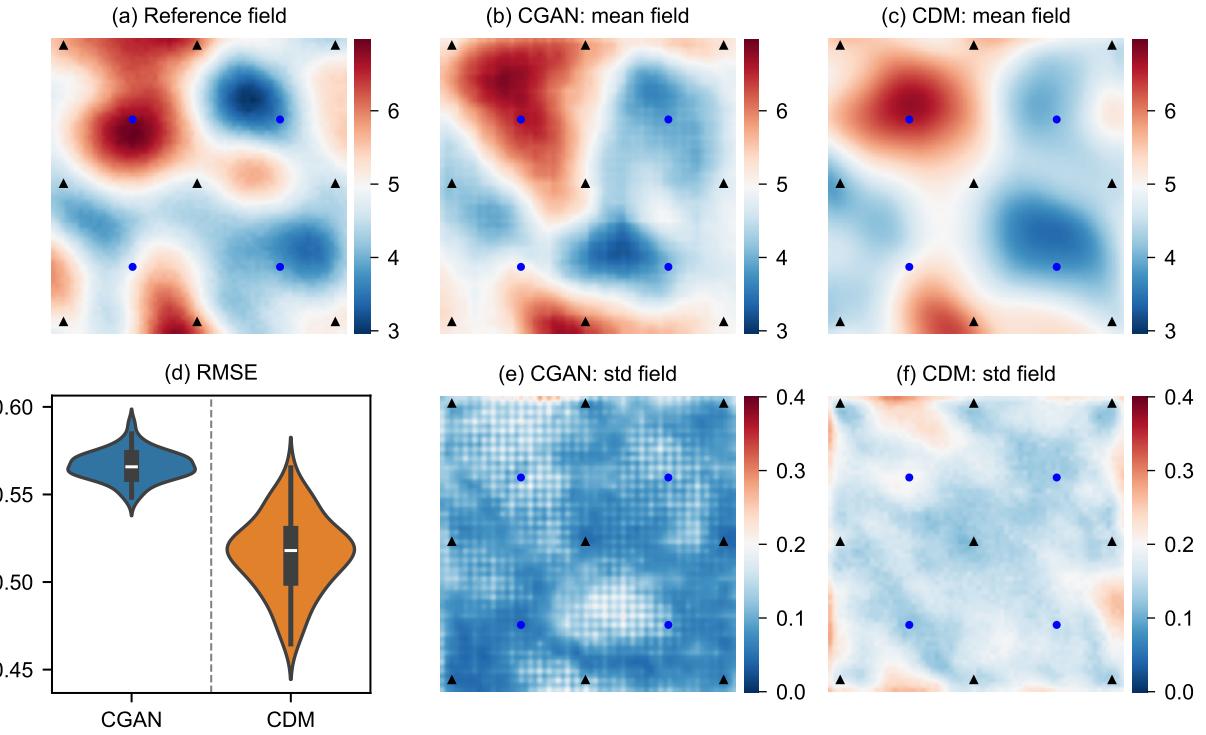


Figure S3: (a) Reference log-permeability field. (b-c) Mean of the log-permeability fields estimated by the CGAN and CDM, respectively. (d) Comparison of the RMSE values for 100 realizations obtained from the CGAN and CDM. (e-f) Standard deviation (std) of the log-permeability fields estimated by the CGAN and CDM, respectively.

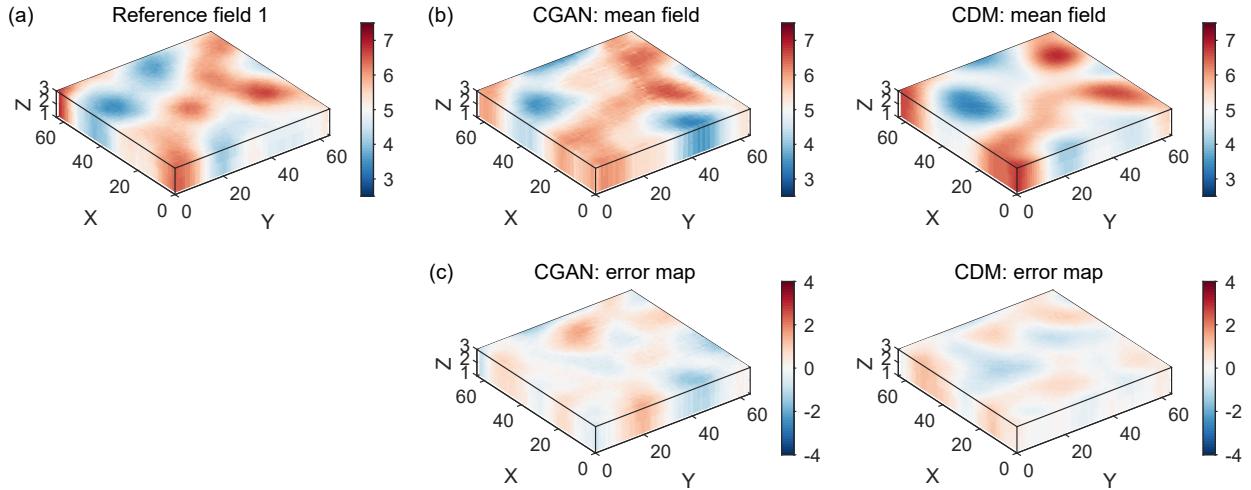


Figure S4: (a) Reference log-permeability field. (b) Mean of the log-permeability fields estimated by the CGAN and CDM, respectively. (c) Absolute error between the mean field and the reference field, respectively.

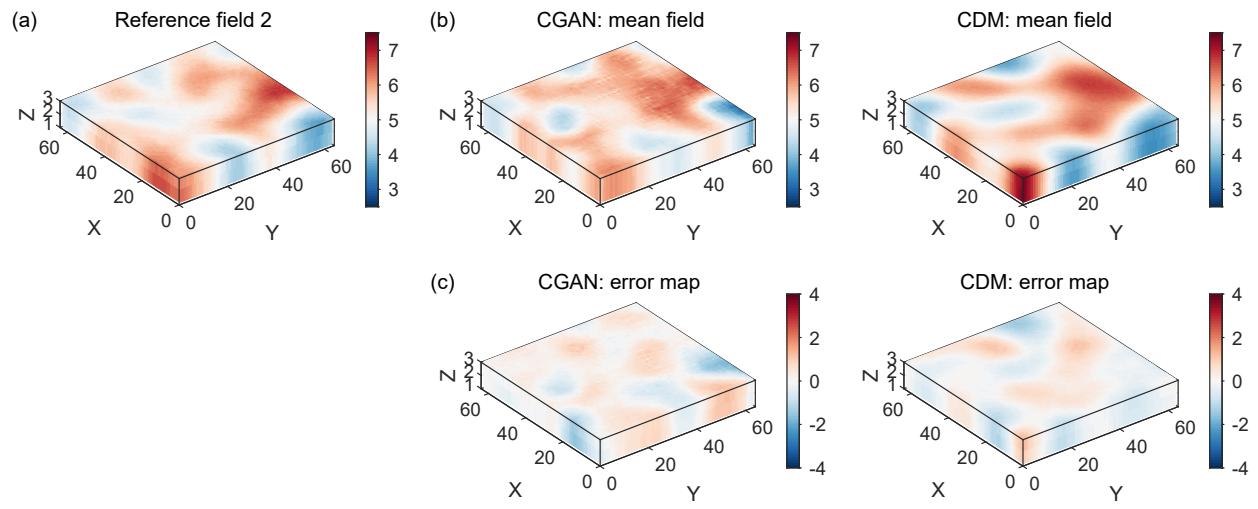


Figure S5: (a) Reference log-permeability field. (b) Mean of the log-permeability fields estimated by the CGAN and CDM, respectively. (c) Absolute error between the mean field and the reference field, respectively.