



OPEN Physics-embedded inverse analysis with algorithmic differentiation for the earth's subsurface

Hao Wu¹✉, Sarah Y. Greer^{1,2} & Daniel O'Malley¹

Inverse analysis has been utilized to understand unknown underground geological properties by matching the observational data with simulators. To overcome the underconstrained nature of inverse problems and achieve good performance, an approach is presented with embedded physics and a technique known as algorithmic differentiation. We use a physics-embedded generative model, which takes statistically simple parameters as input and outputs subsurface properties (e.g., permeability or P-wave velocity), that embeds physical knowledge of the subsurface properties into inverse analysis and improves its performance. We tested the application of this approach on four geologic problems: two heterogeneous hydraulic conductivity fields, a hydraulic fracture network, and a seismic inversion for P-wave velocity. This physics-embedded inverse analysis approach consistently characterizes these geological problems accurately. Furthermore, the excellent performance in matching the observational data demonstrates the reliability of the proposed method. Moreover, the application of algorithmic differentiation makes this an easy and fast approach to inverse analysis when dealing with complicated geological structures.

Inverse analysis gives the solution of inverse problems aiming to find unknown properties of an object, or a medium, from observing a response of this object or medium¹. The inverse analysis process describes finding the matched predictions through a forward model calculation, which takes the parameters describing unknown properties as input, to the observational data^{2,3}. A representative example is seismic inversion, which often involves triggering a source wavefield at the earth's surface and collecting the scattered data at receivers from various positions along the surface. Accounting for the received data, it is possible to find the heterogeneous subsurface structures, such as the existence of an oil deposit, a cave, or a mine¹. In earth science, geological reservoir characterization is of essential value for maximizing oil production from mature hydrocarbon provinces, detecting fluid distributions (groundwater, oil, gas, etc.)^{4,5}, and many other important issues affecting our daily lives⁶⁻⁸. In addition, reservoir properties show spatial heterogeneities from pore to reservoir scale, and it is critical to properly resolve the heterogeneity effects on the underground fluid flow system^{9,10}. However, since physical properties can not be observed directly in the field, inverse analysis techniques have to be applied in order to understand the heterogeneous reservoir properties depending on observational data, such as pressure for hydraulic conductivity fields.

In this study, we propose a novel method for inverse analysis, which generalizes different inverse analysis approaches and can include embedded physics understanding. In addition, we test this novel inverse analysis method for different earth science problems: heterogeneous hydraulic conductivity of a groundwater flow system, hydraulic fracture distribution in a gas-producing reservoir, and seismic inverse for subsurface aquifer determination. Porous media has been the source of valuable fluids such as groundwater and petroleum, as well as both liquid and natural gas¹¹. In a groundwater flow system, fully understanding the heterogeneous subsurface hydraulic conductivity distribution is of importance for estimation of drinking groundwater utilization and contamination mitigation⁶⁻⁸. In addition, the earth's subsurface has also been used for the injection of slurried wastes, like hazardous chemicals or radioactive byproducts^{12,13}, and certain geological reservoirs have been used for CO₂ storage and recovery^{10,14-17}. Characterizing the underground structures, which enables the prediction of the fluid flow system behavior, is essential for successfully using geological sources and avoiding environmental contamination for the projects mentioned above. Notably, we focus on two different scales of heterogeneity in this study. Furthermore, among the producing wells drilled in North America since the 1950s, around 70% of gas wells and 50% of oil wells have been hydraulically fractured. Once a hydraulic fracture is generated, fluid in the reservoir will flow out or into the fracture face and then, along the fracture path, flow out

¹Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ²Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ✉email: wu_hao@lanl.gov

or into the injection or production well^{18–20}. Over the past decades, hydraulic fracture simulation has become a significant part of the design and analysis of oil fields through reservoir characterization and simulation²¹. This study depicts the hydraulic conductivity distribution of a hydraulic fracture network by applying inverse analysis for future oil production estimation and optimization. Last but not least, we exploit a physics-embedded generative model for seismic inversion problems for predictions about underground lithology²². Seismic inversion aims to reconstruct the subsurface structure based on seismic measurements, like trapping mechanisms for hydrocarbon reservoirs and fracture distribution for groundwater storage²³. During the seismic inversion, reservoir properties of interest, such as lithologies, can be transformed from elastic properties (e.g., velocities), which are inverted from seismic data²⁴. Here, resulting from inverse analysis, seismic inversion of underground geological properties is easy and fast.

The inverse analysis provides support for underground feature characterization in earth sciences. Despite its effectiveness, inverse analysis is challenging to conduct, and is computationally expensive⁸. In addition, another concern of inverse analysis is that it may lead to many viable solutions resulting in a second-round calibration or investigation due to the underconstrained or ill-posed features²⁵. There are many feasible ways to conduct inverse analysis, for example, the geostatistical approach^{26,27}, physics-based imaging methods²⁸, and machine learning (ML)^{29,30}. The stochastic geostatistical inversion approach was recognized at unknown parameters estimation³¹, which describes the unknown underground geological properties, like permeability. However, the randomness of the variables reflects the lack of certainty about their values, which are coded as the probability distributions of the quantities. As a result, the solution to an uncertainty quantification turns out to be the maximum likelihood probability distribution of the target variables, based on all the information completed to be interpreted³². Given this feature of the traditional geostatistical approach, the high computational cost may be one concern when dealing with large-scale systems. Even if some recent developments overcome this challenge, like the principal component geostatistical approach (PCGA)^{26,27}, reducing the dimension parameter space by only focusing on principal components of the covariance matrix, the struggle still cannot be avoided when dealing with highly complicated surface structures that are not amenable to a two-point correlation structure. In the use of reflection seismic data, various migration methods are often used to map recorded surface data to their corresponding subsurface reflection points. More robust methods, such as reverse-time migration and full-waveform inversion, can work on models with complex geologic structures but require significant computational cost^{33,34}. An additional representative method is ML, which recently improved inverse analysis in the geological area^{35,36}. These ML studies did use a physics model during the initial training, which led to a scenario with a steep up-front cost to generate the training data, even if they can produce excellent results³⁷. In some situations, the computational cost of generating training data is higher than the pure inverse analysis step, which can be self-defeating. It results from the limitation of these ML models that each training data point needs to run through the forward model for generation, which is a high computational cost. This process means explicitly developing a label for each training data point and going through the supervised ML training afterward. However, the VAE training does not request the generation of data set with a label. In addition, the success of the ML models is subject to the appropriate selection of ML structures, while the uncertainty of results cannot be predicted from different structures (e.g., sometimes shallow networks can reach better results than deep networks)^{37,38}. In the work of³⁹, they applied deep learning for inverse problems about image reconstruction, which illustrates the pressing need for model structure selection, even if there is a considerable improvement in the demand for the training data.

To overcome the aforementioned challenges, including physical calibration during inverse analysis has been successful in enhancing the final results and the model performance^{40–43}. Take the work of⁴⁴, for instance; they addressed the physics-informed diagnostics by testing various ML algorithms' capability for turbulence flow. Interestingly, this physics-informed ML contributes to taking advantage of the mathematical properties of the underlying physics foundation, yielding interpretable strategies from numerical methods and computational fluid dynamics. As a result, it increased the reliability of ML schemes by its high efficiency and accuracy⁴⁵. Other similar examples, like the work of^{46,47}, present similar approaches to incorporating knowledge of physics as a soft physics constraint for the loss function penalty in the area of quantum mechanics. In earth sciences, physics-embedded inverse analysis has been widely applied. In our previous work⁴⁸, we proposed an approach (RegAE) to achieve an easy and fast inverse analysis to interpret complex heterogeneous hydrogeologic reservoir properties by applying the variational autoencoder (VAE), which combines the strengths of the traditional geostatistical approach and recent ML techniques. In addition, we tested the different neural network architectures based on result stability and reliability. Similarly, the physics-informed autoencoders have been investigated for underground fluid flow prediction, and it provides a comprehensive understanding of model stability and prediction certainty improvement⁴⁹. As a result, involving the understanding of physics foundation by applying the mathematical properties of physics laws when generating the target fields is a powerful tool for inverse analysis problems, which provides fundamental support for accuracy and efficiency enhancement. Triggered by the findings in our previous work, here, we address the importance of including physics understanding during inverse analysis and provide a more generalized idea to systematically create an inverse analysis for different applications. Therefore, this study describes our novel approach to physics-embedded inverse analysis and demonstrates its efficacy on multiple subsurface problems, including subsurface flow and seismic wave propagation.

Underground reservoir property characterization is complicated because the underlying system is unknown⁵⁰. For inverse analysis, it is essential to rely on observational data to discover the underground structures and features since it is impossible to directly observe all the detailed information about the whole system in the field. Given this reason, especially for large-scale inverse models, the use of many observations is essential⁵¹. Thanks to the rapid development of sensor networks, we can collect a wealth of variable fidelity observations and monitor the evolution of complex phenomena at large spatial and temporal scales³⁷. Consequently, it leads to a scenario where the inverse analysis based on observational data can be performed. Beyond this, another key factor to achieving successful inverse analysis is result calibration by matching observational data. Even though enough

observational data has been utilized during calibration to reach good performance of the inverse model, regularization, a numerical technique involving adding a term to the objective function, is highly valued for improving results. Adding a regularization term to the objective function seeks to develop additional desired features to the inverse solution, such as smoothness, convexity, or respecting prior knowledge of geologic features. In addition, optimization is the most time-consuming step during inverse analysis, but we apply algorithmic differentiation to increase the computational efficiency⁵². Algorithmic differentiation can compute gradients with a low computational cost for complicated computer programs by applying the chain rule repeatedly. Specifically, reverse model algorithmic differentiation is good at calculating high-dimensional derivatives (in this study, we map a high-dimensional input to a low-dimensional output), which is often useful for inverse analysis problems with substantial computational savings. Because of regularization and algorithmic differentiation techniques, inverse analysis becomes more feasible to estimate interest quantities reasonably based on available data³².

The remainder of this manuscript describes the workflow of the general physics-embedded inverse analysis, the inverse analysis results for different problems, and the benefits and improvement of this approach in “Methods”, “Examples and results”, and “Discussion” sections. Finally, we present our conclusion about applying this approach in “Conclusion” section.

Methods

The physics-embedded inverse analysis starts with the physics-embedded generative model generation. Specifically, in this study, the physics-embedded generative models describing the quantities of interest are the heterogeneous hydraulic conductivity distribution, hydraulic fracture distribution, and seismic P-wave velocity. These are the targets of what inverse analysis is trying to predict through observational data matching. Several key factors are picked to represent the variability of the targets through the physics-embedded generative model for the stability test. For example, in the hydraulic fracture problem, five key factors (which can be understood as latent variables) are utilized to represent the lengths of the hydraulic fractures in a cluster, which is correlated with permeability. The physics-embedded generative model describes the relationship between the latent variables to the target properties we are interested in. In addition, the physics-embedded generative model embeds physical knowledge of the system. Continuing with the hydraulic fracture example, once the lengths of the hydraulic fractures are figured out, the permeability of the fractures could be calculated based on the mathematical models, e.g., we apply the fracture size-transmissivity relationship. Finally, we build a model that characterizes the hydraulic fracture permeability distribution based on the representative latent variables. Broadly speaking, the physics-embedded generative model links the small number of latent variables to many target properties of interest, encoding the relationship between them. The physical knowledge embedded in the generative model increases the reliability and accuracy of the inverse analysis.

Once the physics-embedded generative model has been constructed, the second step is the objective function set up, where a forward physical model takes the output from the physics-embedded generative model to simulate the mechanisms of the study system, like fluid flow for hydraulic conductivity fields. Using the hydraulic fracture problem as an example again, a gas production situation has been simulated. The fluid flow from fractures to the pumping well, specifically the pressure drop, is calculated through the forward model. The objective function characterizes the difference between the observational data and the predicted output from the forward model. The final step is performing the inverse analysis using gradient-based optimization with the gradients being computed by algorithmic differentiation. During this step, the output from the forward model is compared with the observational data through the loss function to achieve final optimized results. The detailed workflow is illustrated in Fig. 1.

During this study, the notation \mathbf{p} and $\hat{\mathbf{p}}$ are used to represent a physical reservoir property field in vector form, \mathbf{z} and $\hat{\mathbf{z}}$ represent the latent variables, and \mathbf{h} and $\hat{\mathbf{h}}$ represent a vector of observations for inverse analysis and the calculation from the forward physical model, respectively. This study obtained observations \mathbf{h} from different problems directly from the related reference fields \mathbf{p} through the forward model. The forward physical model predictions $\hat{\mathbf{h}}$ are obtained based on the guess of target properties $\hat{\mathbf{p}}$ through the iterations. The error used to measure the difference between the true and predicted values should be differentiable for the inverse analysis. This loss takes the simplified form of the sum of squared residuals in the examples studied here.

The optimization problem of inverse analysis is formulated in terms of the key factors or latent variables, $\hat{\mathbf{z}}$ and includes regularization in the objective function. The optimization problem of inverse analysis is formulated in terms of the latent variables, \mathbf{z} :

$$f(\mathbf{z}) = [\Sigma(\mathbf{h} - \hat{\mathbf{h}})^2] + [\text{regularization}] \quad (1)$$

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} f(\mathbf{z}) \quad (2)$$

where the regularization term varies depending on application, and we call $f(\mathbf{z})$ the objective function. The specific objective function is discussed in the subsections of Examples and Results. During the optimization process, through gradient calculation, seeking the minimum value of the objective function represents the finding of the optimized final results. Regularization adds additional benefits to the inverse analysis, avoiding side effects like overfitting. In addition, algorithmic differentiation is used to compute the gradients for the optimization. Our method not only includes the physics understanding in the inverse analysis, but also finds the potential to conduct optimization calculations for complicated geological problems easily and efficiently, which is algorithmic differentiation. Specifically, the algorithmic differentiation library we applied is Zygote.jl⁵², and we use the differentiable physics simulator, DPFEHM⁵³. As for optimizing the objective function, a gradient-based optimization method is utilized, which is the limited-memory Broyden–Fletcher–Goldfarb–Shanno⁵⁴ (L-BFGS) method

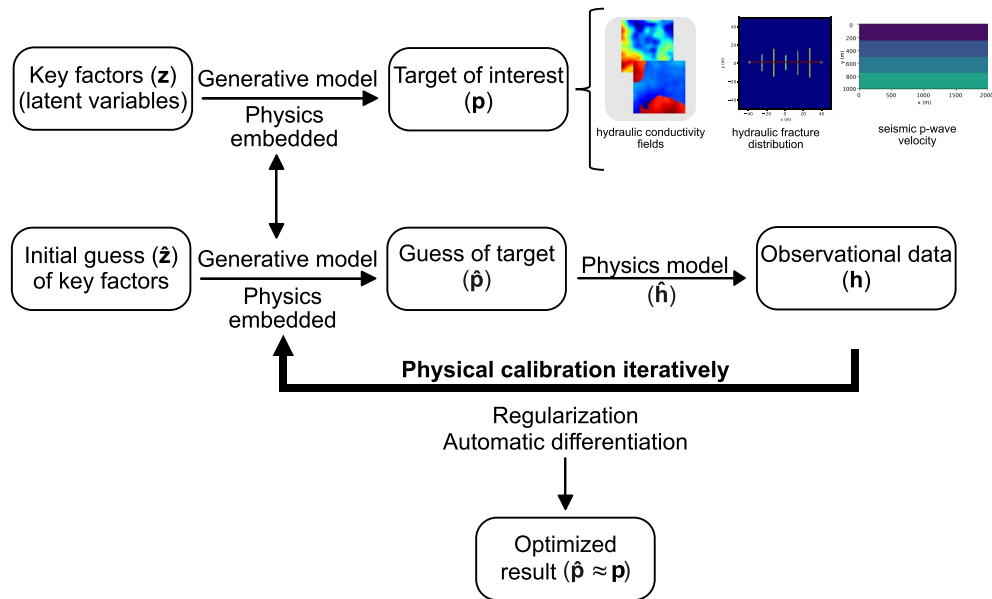


Figure 1. Physics-embedded inverse analysis leverages the underground geological property characterization with physics law understanding and algorithmic differentiation. Inverse properties are calculated based on the physics-embedded generative model, and calibration is performed on the output of the forward model. The first row characterizes the “true” situation in the field, and the generative model describes the physics understanding of the problem. This generative model is applied in the inverse analysis (second row) to represent the physics-embedded feature. The application of regularization and algorithmic differentiation makes the optimization process simple.

with a Hager-Zhang line search⁵⁵. The Optim.jl⁵⁶ software package is specified for this process. Of course, other gradient-based optimization routines could also be used. For all the problems, to start the inverse analysis, the initial guess of key factors is set to be 0.

Examples and results

This study provides a generative approach to physics-embedded inverse analysis. We focus on three problems: heterogeneous hydraulic conductivity field, hydraulic fracture distribution, and seismic inversion of P-wave velocity property. Two types of heterogeneity have been considered for heterogeneous hydraulic conductivity fields. For larger-scale heterogeneous fields, to improve the inverse analysis performance, the ML method was applied. To estimate the performance of inverse analysis, the comparison of reference fields and the final inverse results are conducted for different problems and are presented in Figs. 2, 4, 5 and 6. The convergence of different inverse analyses showing the optimization process is described in the supplementary information, Figs. S1–S4. The comparison of observational data and the prediction of the forward physical model after inverse analysis are shown in the supplementary information, Figs. S5–S7.

In Figs. 4, 5 and 6, the comparison results, the first rows are the three “true” reference fields. The following rows demonstrate the inverse results, while the last rows are the difference calculated between the “true” and results estimated by the inverse analysis. On top of the inverse results (second row) the relative error is displayed, which measures how close the inverse result is to the reference field and is defined as

$$\frac{\|\mathbf{p} - \hat{\mathbf{p}}\|^2}{\|\mathbf{p} - \bar{\mathbf{p}}\|^2}, \tag{3}$$

where $\bar{\mathbf{p}}$ is the mean of the reference field. Especially for the Gaussian hydraulic conductivity field in Fig. 2, only one reference field (as the Gaussian field is easier to be characterized than the bimodal field, which demonstrates three examples in Fig. 4) has been represented to investigate the performance of the proposed approach.

Principal component geostatistical approach for Gaussian hydraulic conductivity. One of this work’s focuses is a heterogeneous hydraulic conductivity field. First, we discuss a multivariate Gaussian field of small-scale heterogeneity in the hydraulic conductivity field. A 200 m × 200 m subsurface aquifer is simulated with a unit thickness. Two hundred eigenvalues \mathbf{z} , as latent variables (the principal components), have been introduced to go through a Gaussian distribution, with mean 0, variance 1, and correlation length 50 m, to create the heterogeneity of the research area. We use the GaussianRandomFields.jl package to generate the multivariate Gaussian field for the Julia programming language⁵⁷. The heterogeneous hydraulic conductivity field \mathbf{p} (200 × 200) is shown in Fig. 2a. For the physics-embedded generative model, we utilize the principle components of the covariance matrix to represent the Gaussian distribution. The principle components of the covariance matrix are

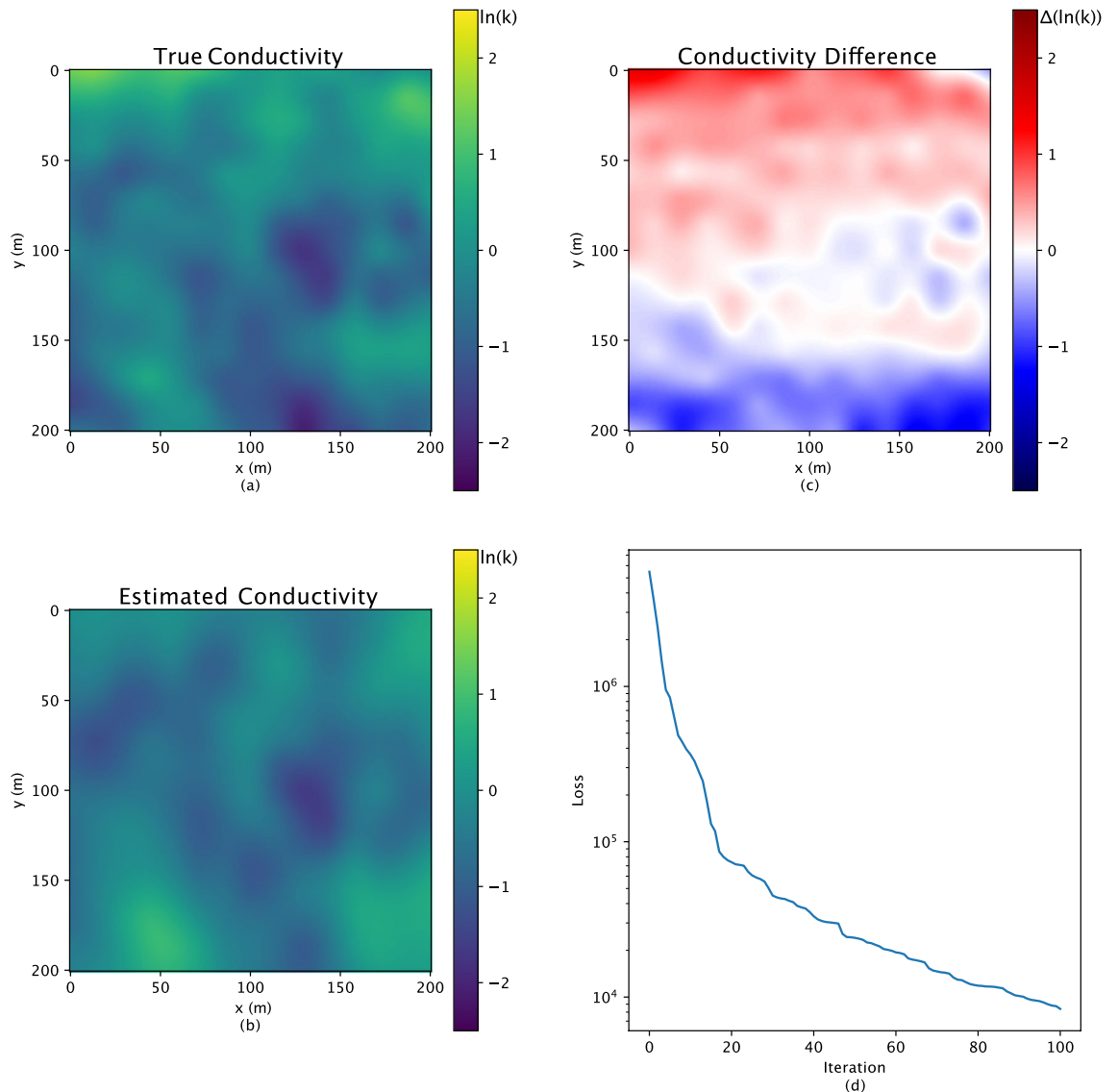


Figure 2. The reference conductivity field is shown in subfigure (a), and the corresponding inverse results are shown in subfigure (b). The hydraulic conductivity shows on the \ln scale. Subfigure (c) shows the difference between the related reference fields and the inverse results. The convergence of the inverse analysis is shown in subfigure (d).

calculated by Karhunen-Loève theorem. This embeds knowledge of the statistical structure of the permeability fields. Our approach generalizes existing PCGA and accelerates it through algorithmic differentiation (PCGA was designed for black-box models where algorithmic differentiation is not possible). We intend to establish that our approach can replicate a method familiar to geoscientists (especially hydrologists) and provide better performance.

More importantly, the background groundwater flow is simulated through a forward physical model for inverse analysis calibration. The boundary condition yields a constant 5 m head drop from left to right. In addition, in the center of the research area, water is injected at a rate of $1.0 \text{ m}^3/\text{s}$. The observation used to inform the inverse analysis is the hydraulic head, from a static forward Darcy's law and considering mass conservation, on a 16×16 regular grid within the domain. Figure 3a shows the reference head distribution, and the positions for all the observations are shown in Fig. 3d. The objective function is specified as the following equation, which considers latent variables as well:

$$f(\mathbf{z}) = [\Sigma(\|\mathbf{h} - \hat{\mathbf{h}}\|^2)] + [\Sigma(\mathbf{z}^2)]. \quad (4)$$

For the Gaussian hydraulic conductivity field in Fig. 2, the “true” reference field and the estimated result are similar, particularly for the center area, which indicates the good performance of the inverse analysis. However, the error reaches around one order of magnitude, only existing at the top and bottom edges. We hypothesize

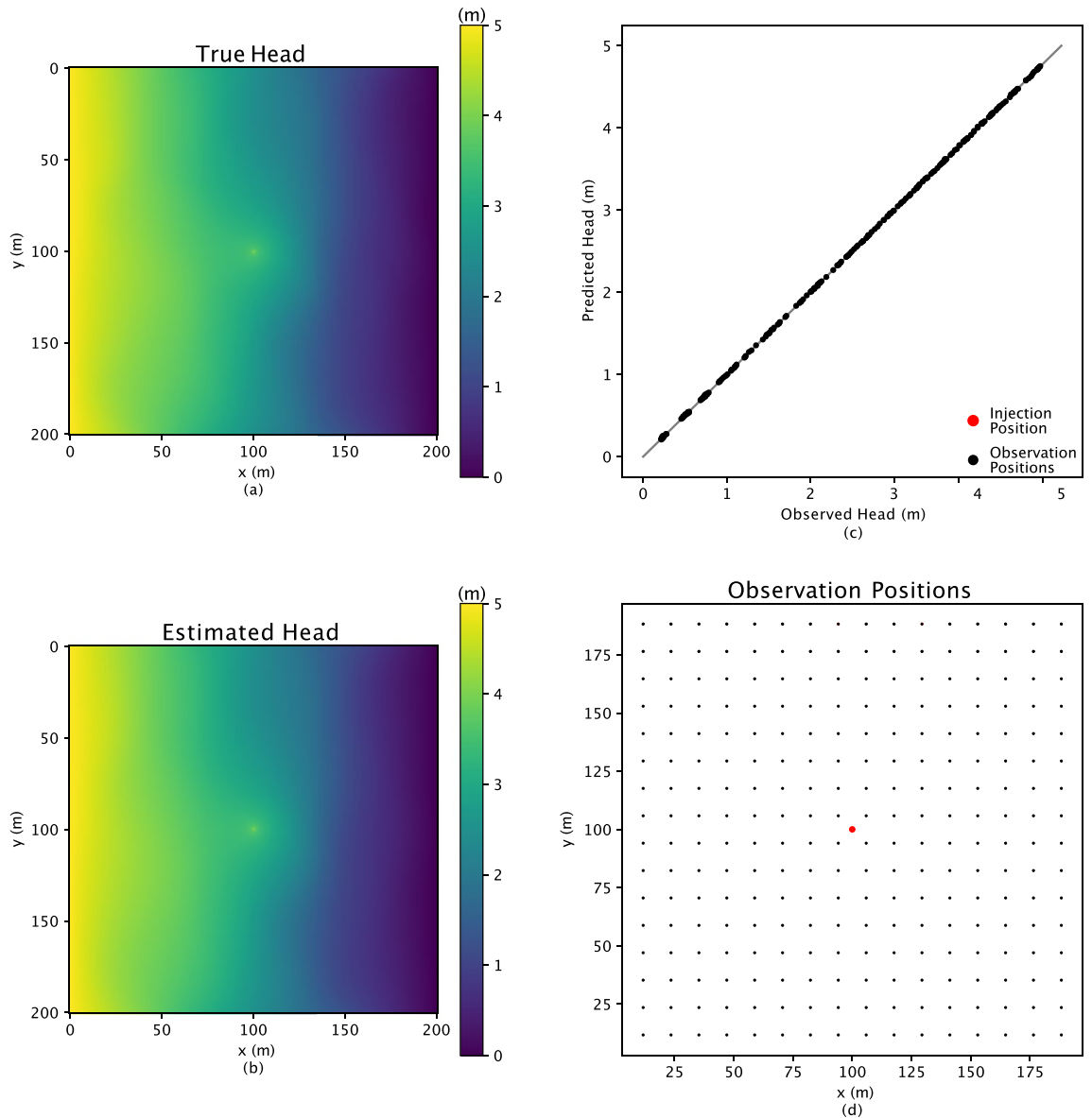


Figure 3. The head distribution of the reference conductivity field is shown in subfigure (a), and subfigure (b) shows the corresponding inverse estimated head distribution. The black dots in subfigure (d) represent the 16×16 regular grids for observation points, and the one red dot illustrates the injection position, with a rate of $1.0 \text{ m}^3/\text{s}$. Finally, the comparison of the observed head and predicted head for all the observational points are shown in subfigure (c).

this is partly due to the fixed pressure boundary conditions, which make the observations less sensitive to the hydraulic conductivity near these boundaries.

ML approach for bimodal hydraulic conductivity. Beyond the Gaussian field, to show our approach generalizes additional methods, we show how it generalizes RegAE⁴⁸. RegAE is a method that can solve more challenging permeability fields than the principal component geostatistical approach. The domain is a $100 \text{ m} \times 100 \text{ m}$ subsurface aquifer with a unit thickness. In this type of field, the higher heterogeneity is applied and is represented by two hydrogeologic facies with distinct properties, each of which is a multivariate Gaussian distribution. The two multivariate Gaussian structures are shown as conductivity 1 and 2 in Table 1. More importantly, the “Split” model has a different multivariate Gaussian structure that has been utilized to indicate which of the facies is present at a given location. As a result, the new type of field shows a bimodal hydraulic conductivity distribution, and the reference fields are represented in Fig. 4a, d, and g.

In terms of the higher heterogeneity of the bimodal fields, a generative machine learning model VAE is included to capture the hydrogeological properties distribution in this study. VAE⁵⁸ is a generative ML model with neural network architecture and has widespread application for image data⁵⁹. VAE consists of two parts: an encoder and a decoder. The encoder step maps a high-dimensional space \mathbf{p} (such as pixels in an image) into a

Models	Mean (m/s)	Variance (m ² /s ²)	Correlation length (m)
Conductivity 1	10 ⁻⁵	1.0	50.0
Conductivity 2	10 ⁻⁸	1.0	50.0
Split	10 ⁻⁸	1.0	200.0

Table 1. The statistical parameters used to generate the hydraulic conductivity data.

smaller parameter space. Specifically, the smaller parameter space is the key factor containing the features of the image, which is the hydraulic conductivity distribution in this study, and used to be called the latent variables \mathbf{z} . In reverse, the decoder maps the latent variables back to their original high-dimensional space. In the bimodal case, the reference fields have a resolution of 100×100 pixels, which is a high dimensional space. The training of VAE goes through these two encoder and decoder steps. The dimensions of the latent variable, \mathbf{z} , are 50, 100, and 200, and 100 epochs are performed during the VAE training for characterizing the relationship between the latent variables and the property distribution. For more details on how the VAE was trained, refer to⁴⁸. After the training of the VAE, \mathbf{z} is applied to represent the physics meaning of the hydraulic conductivity distribution feature. At the same time, the decoder step accounts for the physics-embedded generative model. Here, the physics is embedded through the process of training the VAE on images that contain the physical understanding of the subsurface—in this case, the two facies. In this study, three \mathbf{z} values, 50, 100, and 200, are tested for the inverse analysis performance in terms of key factors.

Like the Gaussian case, for bimodal fields, a constant head drop of 1 m from the left boundary to the right has been set up for the fluid flow system. The head distribution is calculated through the forward physical model (Darcy's law) of groundwater dynamics at a static state and considering the mass conservation. Oppositely, a coarse 5×5 regular grid within the domain is demonstrated for observation during inverse analysis calibration, shown in Fig. 4a, since VAE is a powerful tool to capture the property distribution with less input information. The regularization term of optimization for bimodal fields takes latent variables covariance into account, and the objective function is defined as:

$$f(\mathbf{z}) = [\Sigma(\|\mathbf{h} - \hat{\mathbf{h}}\|^2)] + [\mathbf{z} - \bar{\mathbf{z}}]^T \Sigma_z^{-1} [\mathbf{z} - \bar{\mathbf{z}}], \quad (5)$$

where Σ_z is the covariance matrix for the latent variables, and $\bar{\mathbf{z}}$ is the mean of the latent variables. A more detailed description of inverse analysis for Gaussian and bimodal fields is specified in⁴⁸.

More heterogeneous, bimodal hydraulic conductivity fields are simulated through VAE as the physics-embedded generative model, which is good at spacial feature characterization, to test our idea of the physics-embedded inverse analysis; the results are shown in Fig. 4. For bimodal fields in Fig. 4, the broad similarity in each facies between the reference fields and simulated results implies that the inverse analysis approach captures the salient aspects of the hydraulic conductivity distribution features. Even if the relative error is comparably higher than those from Gaussian fields, the phenomena lie in the higher complexity of bimodal fields. Meanwhile, for the difference in Fig. 4c, f, and i, the major error only occurs at the edge of the two facies; oppositely, in each face, the difference is small and close to zero. In conclusion, including the physics-embedded generative model during inverse analysis, even for the ML-specific approach, turns out to be a good application for different types of heterogeneous hydraulic conductivity fields based on having consistently good results. However, even if it approves the application of the physics-embedded inverse analysis, specifically for more complicated fields, like the edges in bimodal fields, it needs more calibration or on-site investigation for future field applications. On the other hand, it also implies that more detailed physics understanding or background should be included when dealing with complicated field situations.

Hydraulic fracture network. Most drilled wells have been fractured in the oil and gas production field, resulting from fluid pressure differences^{11,18,19}. Fully understanding the distribution and properties of the hydraulic fractures is of essential importance for production estimation and reservoir protection. In the cases of drilled wells, they are now turned fully horizontally into the target geologic formations. At the same time, for almost all depths of interest, the hydraulic fracture will be normal to the direction of the horizontal well. In this study, a cluster with five hydraulic fractures has been selected to present the process of inverse analysis during gas production⁶⁰. A medium-scale matrix of size $100 \text{ m} \times 100 \text{ m}$, with a 78 m drilled well in the center position, was used to represent the research domain. Hydraulic fractures are in the normal direction to the drilled well and distributed in a constant interval between them. The length of the hydraulic fractures follows a power law, given by⁶¹

$$f(r) = \frac{(1-p)}{(\mathbf{R}_1)^{(1-p)} - (\mathbf{R}_0)^{(1-p)}} * r^{(-p)}, \quad (6)$$

where r is the length of hydraulic fracture, p is the power, \mathbf{R}_1 and \mathbf{R}_0 are the maximum and minimum of the hydraulic fracture length range, and $f(r)$ is the possibility of a certain length r . Based on a literature review of fracture length distributions⁶², the power p is set up to be 1.8, while the length range of fractures spans from 10 m to 90 m in this problem. After the fracture length has been determined, a size-transmissivity relationship, which describes the transmissivity of fractures and shows a positively correlated power law with the length of fractures, is introduced. The size-transmissivity relationship is defined in^{63,64} as

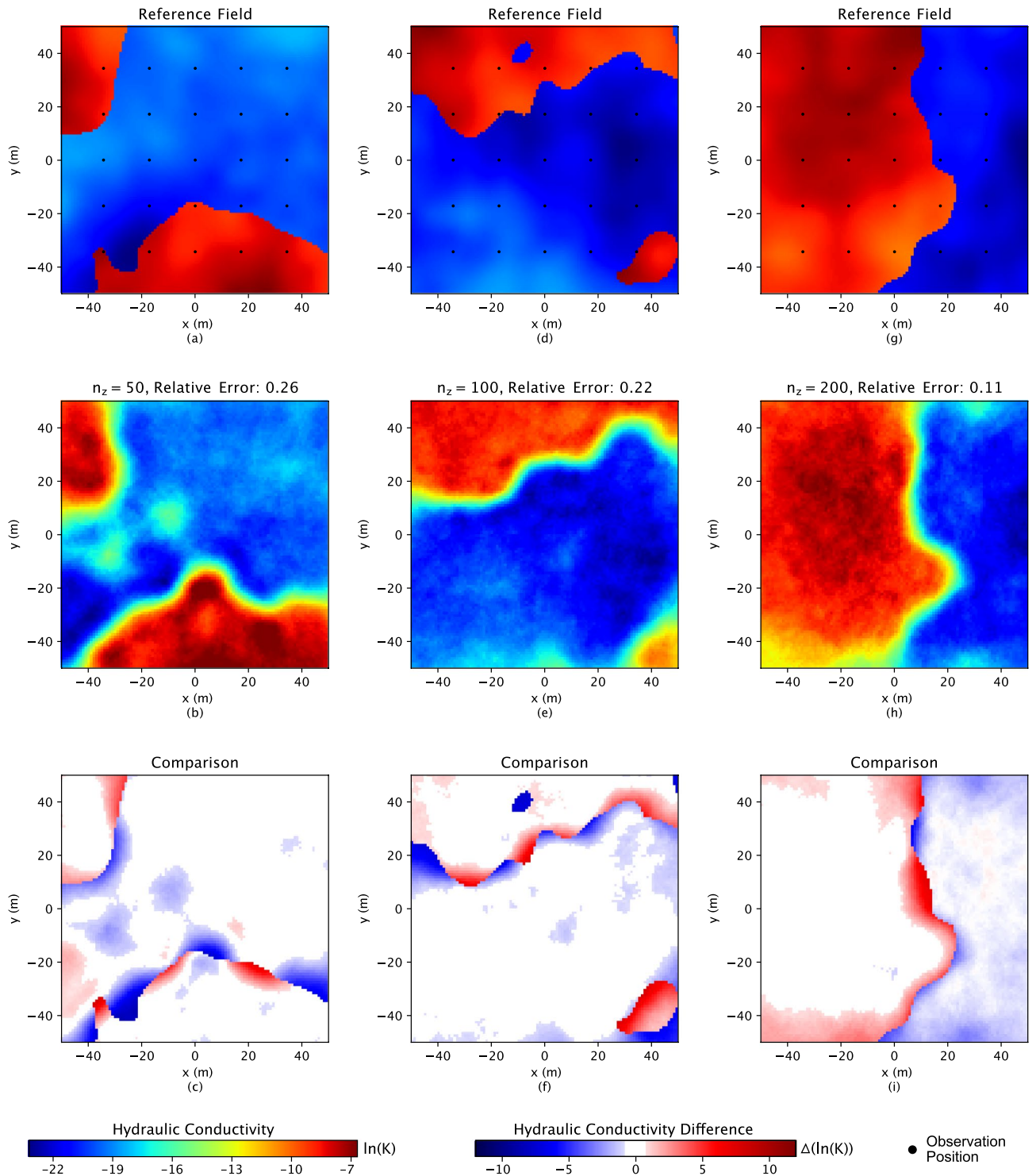


Figure 4. Three bimodal reference conductivity fields are shown in subfigures (a), (d), and (g). The black dots in the first row represent the 5 × 5 regular grids for observation points. The corresponding inverse results are shown in the second row, subfigures (b), (e), and (h). The hydraulic conductivity shows on the \ln scale. The relative errors above inverse figures represent the difference between related reference and simulation figures. The last row shows the difference between the related reference fields and the inverse results.

$$\log(T) = \log(\alpha * r^\beta), \tag{7}$$

where T is the fracture transmissivity, and α and β are related parameters with values $1.3 * 10^{-9}$, and 0.5, respectively. In this study, the reservoir has a 10 m thickness. In addition, the permeability of the matrix and the drilled

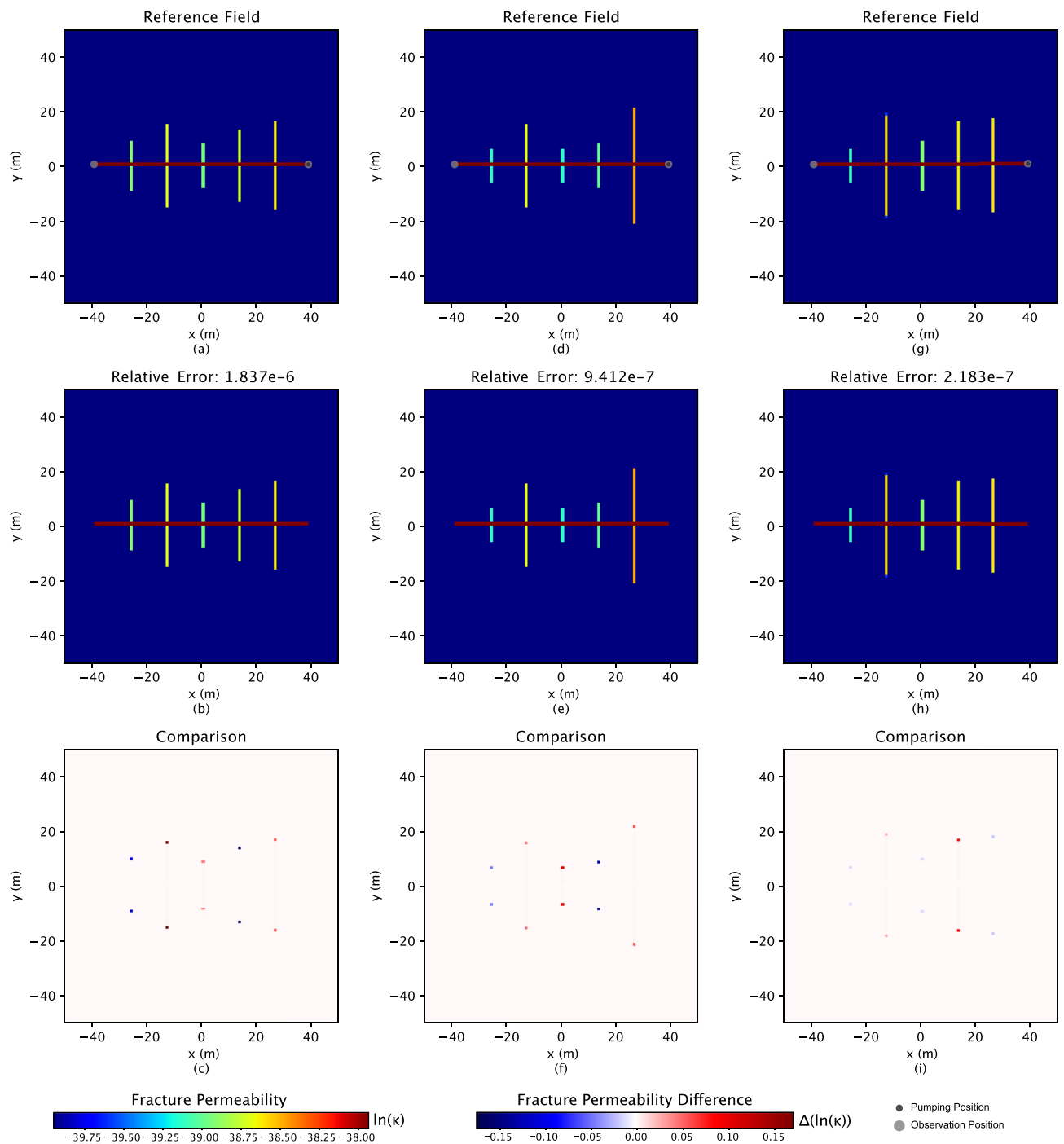


Figure 5. Three hydraulic fracture reference conductivity fields are shown in subfigures (a), (d), and (g). The two grey dots in the first row represent the positions for observation, and the black dot illustrates the gas pumping position. The corresponding inverse results are shown in the second row, subfigures (b), (e), and (h). The permeability shows on the \ln scale. The relative errors above inverse figures represent the difference between related reference and simulation figures. The last row shows the difference between the related reference fields and the inverse results. The color bars only apply for the hydraulic fractures, and the permeability for the matrix and well are -52.16 and -24.53 (\ln scale), respectively.

well are set to be at the scale of $10^{-22} m^2$ and $10^{-10} m^2$, respectively. Specifically, at the two tips of the fractures, a harmonic mean is introduced to represent the change from fracture to the matrix. Hence, five random key factors (latent variables) were selected to represent the possibility of five fracture lengths. Then the permeability of the fractures in the cluster (p) is calculated through the physics-embedded generative model mentioned above for the following inverse analysis.

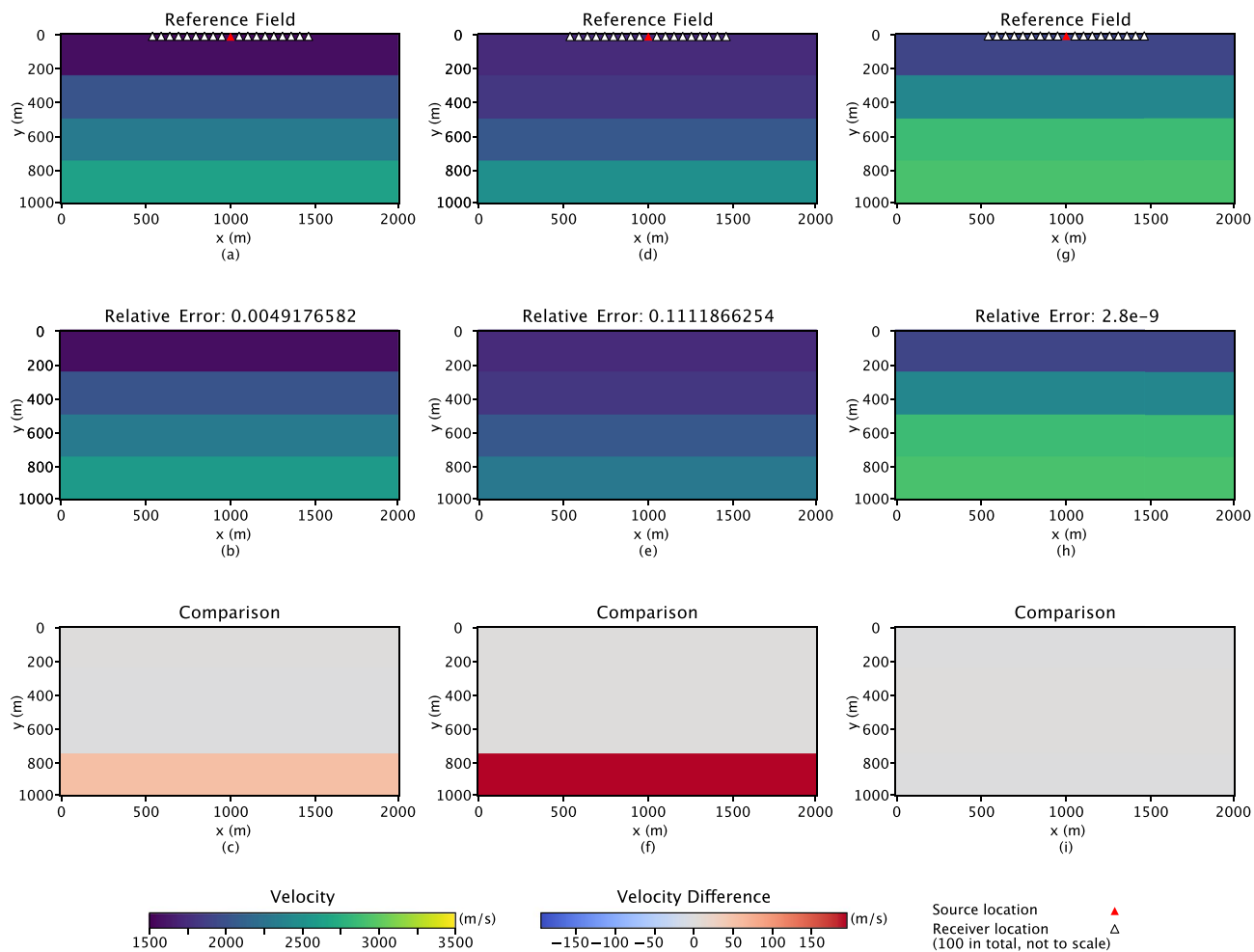


Figure 6. Three reference seismic P-wave velocity fields are shown in subfigures (a), (d), and (g). The red triangle in the first row represents the positions for source location, and the white triangles illustrate the 100 receiver locations, which are not to the real scale. The corresponding inverse results are shown in the second row, subfigures (b), (e), and (h). The relative errors above inverse figures represent the difference between related reference and simulation figures. The last row shows the difference between the related reference fields and the inverse results.

In this study, gas production has been simulated. Especially the gas pumping position is set to be at the right end of the drilled well with a rate of $0.82 \text{ m}^3/\text{s}$ based on data from the Marcellus Shale Energy and Environment Laboratory (MSEEL)⁶⁵. The shale layer of gas production is located at a depth of 2300 m, and the temperature and pressure of the subsurface are 75°C and 15 MPa, respectively. A transient flow model based on mass conservation and Darcy's law is built for the pressure drop calculation, assuming single-phase gas flow based on the experience at MSEEL, which is very dry. Two observation points are located at the two ends of the drilled well, as shown in Fig. 5. As the pumping goes on, the observation lasts for two weeks, with a 30-min frequency of data collection. The objective function for this problem utilizes the same equation as Eq. (4) in “Principal component geostatistical approach” section.

Similar results in the Gaussian fields for the hydraulic fracture problem are shown in Fig. 5. Surprisingly, the high similarity between the reference fields and the simulated results illustrates the characterizing ability of inverse analysis when including the physics understanding of the relationship between fracture distribution and permeability. In addition, the extremely low relative error additionally supports the conclusion. The error in the permeability of the hydraulic fractures mainly exists at the two ends of the fractures and is small. The more we understand the physics background in the hydraulic fracture problem, the higher possibility we can predict the underground fluid flow system and make a more reliable estimation of oil and gas production.

Seismic inversion. Seismic inversion estimates subsurface properties by matching predicted data generated on a proposed model to observed data collected at receiver locations. This study only tests the inverse analysis from the seismic records \mathbf{h} observed at the surface to elastic properties \mathbf{p} , which is the velocity at which the P-wave passes through subsurface layers. The research area is an underground reservoir of size $2 \text{ km} \times 1 \text{ km}$, consisting of four horizontal layers. In addition, our physics-embedded generative model embeds the domain knowledge that the velocity tends to increase with depth. Four key factors (latent variables) are developed to

compute the increasing velocity trend with depth. The reference fields of the geological layer properties (ρ with size 2000×1000) are shown in Fig. 6a, d, and g.

To generate the observed data, a seismic wave has to be triggered; in this application, the location of the source point is fixed at the center of the domain on the surface. Since we constrain our models to only vary with depth and not horizontally, we only use one source per model in our experiments. In this study, the wave has been computed using a finite difference model. In addition, 100 seismic receivers are located symmetrically beside the source position along the surface, as shown in Fig. 6. Data sets are recorded at the receivers for a record length of 0.8 s. The data from the 100 receivers for the whole simulation time has been implemented for calibration during inverse analysis. The objective function for this problem utilizes the same equation as Eq. (4) in “Principal component geostatistical approach”.

Finally, we also investigate seismic inversion by applying the physics-embedded inverse analysis. Not surprisingly, the simulated data from the estimated model closely matches the reference field data, indicating a good match between the reference and estimated models. The maximum error is around 11%, which is acceptable for the deep layers. It is more convincing when considering the difference subfigures; the error is only notable for the deepest layer, while all the shallow layers illustrate outstanding results. Hence, we can conclude that the generative approach of physics-embedded inverse analysis is successful for the seismic inversion problem. Meanwhile, one conclusion from these results is that shallow layers have higher reliability during the inverse analysis, and more investigation or calibration is needed when facing deep layers.

Comparison of observational results. Beyond only the comparison between the “true” reference fields and the simulated inverse results, the comparison between the observational data and outputs of the forward model is represented in Figs. 3, S5–S7 to further demonstrate the performance of the physics-embedded inverse analysis. After reviewing all the observation data comparison figures, an inevitable conclusion is that the inverse analysis successfully captures all the observational information features. Therefore, it further approves the application of the physics-embedded inverse analysis for underground reservoir property characterization based on observational information on the earth’s surface.

Convergence. The convergence results are depicted in the supplementary information in Figs. 3 and S1–S4. The convergence is generally obtained from 20 to 120 iterations for all the problems. Generally, the higher accuracy (hydraulic fracture problem) needs more iterations to reach good results; however, this relationship also is affected by the complication of the reference fields and the quality of the physics-embedded generative model. More discussion about the computational time and cost follows in the next section.

Discussion

Inverse analysis is fundamental to help us find the underground structure and the geological properties distribution. However, since the underground situation is complicated and we only have some observational data at specific locations, the inverse analysis is sometimes hard to conduct, and the results accuracy is unreliable sometimes. It is essential to improve the accuracy and reliability of inverse analysis in geosciences. Since our model incorporates the physics of the underlying problems, it reaches accurate final results and decreased model time. This demonstrates the advantages in considering the physics background which is the importance of our study^{40,41}. The physics-embedded inverse analysis provides an approach including the physics background understanding to perform inverse analysis effectively. At the same time, the application of algorithmic differentiation shows fast and efficient gradient calculation during optimization, as is discussed below. Our goal here is to demonstrate an inverse analysis approach that uses a physics-embedded generative model by showing how it generalizes some existing methods and can be used more broadly in both subsurface flow problems and seismic inverse problems. Three inverse problems are completed to investigate the accuracy of the proposed approach; mainly, for the heterogeneous hydraulic conductivity fields, we discussed the two types of heterogeneity and the various physics-embedded generative models during the inverse analysis. Generally, the comparison results in Fig. 2, 4, 5 and 6, which show the high similarity between the reference fields and the simulated results, support the discovery of the underground properties using these inverse methods. However, for Gaussian hydraulic conductivity problem, which is based on the statistical characterization of the system as the physics understanding, brings some concerns about accuracy only at the edges of the research area. At the same time, for the more heterogeneous problem, the bimodal fields, including the VAE method, which can thoroughly characterize the property distribution in each face from image data, illustrate its strong capability for the complicated scenario. However, the VAE leading inverse analysis struggles at the boundary of the two facies. Therefore, it shows that more complicated problems need a complete understanding of the physics background to reach the perfect performance of the inverse analysis.

Specifically, the hydraulic fracture problem provides the best inversion results of all the research problems. Understanding the physics mechanism of the fractures generation and distribution convinced us of its immense potential for highly successful inversion of the hydraulic fracture network, which provides the following estimation or protection plans for oil and gas production. However, in this study, the two ends of the fracture also draw attention to more calibration. It indicates the difficulty of inversion when considering the connection of the hydraulic fracture network with the existing natural fracture system. In addition, we only pick one cluster of hydraulic fractures to conduct the inverse analysis; in an actual situation, the production well is several km long, where there are many hydraulic fracture clusters along with it. Similarly, in the scenario about seismic inversion, the low relative error and high similarity illustrate the success of the physics-embedded inverse analysis approach. The error mainly focuses on the deep layers, even if the error is relatively small, which indicates that more consideration may be needed for deeper layers. Again, there are large fractures, caves, and mines in the subsurface

environment, which creates the discontinuity of the properties of interest and brings difficulty to the inversion. However, our generative physics-embedded inverse analysis provides an approach to easily and rapidly conduct underground property characterization. Our approach generalizes several different inverse analysis approaches (e.g., PCGA and RegAE), and the accuracy of the final results depends on the choice of the generative model. Even if we need to discuss the more complicated problems in the future, the solution would be to improve the physical background understanding, which leads to applying an appropriate generative model to the related simulations. As a result, we only provide some fundamental insight into how to invert the underground geological structures by applying the inverse analysis method.

Our analysis was performed on a machine with an Intel(R) Core(TM) i9-9960X CPU @ 3.10GHz with 32 threads and an NVIDIA RTX 2080 Ti GPU only for the VAE training. Except for the VAE training for bimodal fields, all the other problems only need to prepare the three reference fields through the physics-embedded generative model, which does not require much generation time. However, the optimization is most time-consuming, and the time to perform the inverse analysis varies somewhat depending on the reference fields. For example, for the Gaussian hydraulic conductivity problem, each epoch needs around 10 s to finish, while for the hydraulic fracture problem, the average time for each epoch is 10 m. As a result, the inverse analysis process time for all the problems mentioned in this study spans from 5 minutes to 1 day. Furthermore, the gradient calculation dominates the total computational cost of the inverse analysis. However, algorithmic differentiation efficiently improves the computation rate for the gradient calculation. Even if the reduction to \mathbf{z} parameters from \mathbf{p} interest properties makes the inverse analysis easy and fast, the application of algorithmic differentiation shows its extra benefit. The cost of computing a gradient with finite difference methods is ~ 200 model runs (proportional to the number of components of \mathbf{z}), while the cost of computing a gradient including algorithmic differentiation is ~ 2 model runs on average. Therefore, applying algorithmic differentiation helps speed up these computations by an additional factor of up to ~ 100 . In our research, there is only one objective function without any optimization constraints, which makes the optimization easier to achieve. In addition, the application of regularization allows for easy optimization. Combining these two features illustrates their vast potential for computational cost-saving for easy and efficient inverse analysis.

This study proposes the generative approach to include the physics-embedded generative model during inverse analysis. The framework can efficiently characterize various underground properties inversion and demonstrate accurate and trustworthy prediction results. Understanding the subsurface geological structure and properties helps in groundwater management and protection, oil and gas production estimation and optimization, and heterogeneous underground structure detection. Our approach provides new avenues of support for achieving good performance for inverse analysis by including the physics-embedded generative model. In addition, with algorithmic differentiation, the optimization can be completed fast and efficiently. Finally, we will explore more complicated and realistic geologic research problems by applying our proposed approach to expand its application in geologic properties inversion.

Conclusion

We have presented the application of an inverse analysis approach with a physics-embedded generative model for underground geological properties characterization, which provides an efficient method of regularization and algorithmic differentiation. In this study, a novel method for inverse analysis is proposed, which generalizes different inverse analysis approaches, and we have tested the application of this approach for various problems. We used four physics-embedded generative models: one based on the principal components arising from the geostatistical structure of the parameter fields, another using a variational autoencoder that was trained on images of the parameter maps, a third that embeds the structure of a hydraulic fracturing well (including a relationship between fracture length and permeability), and a fourth that includes geologic layers with distinct P-wave velocities. As a result, the physics-embedded inverse analysis provides accurate and consistent performance for various inverse problems. Using the physics-embedded generative model in combination with observational data enables to construction of a loss function that can be automatically differentiated. Our approach is computationally efficient and obtains an excellent solution to the inverse problem by easing the regularization process and applying algorithmic differentiation. In the future, different observational strategies need to be discussed to enhance the accuracy for more significantly complicated problems and deliver a high level of reliable inverse results based on an efficient observational plan.

Data availability

A computer program automatically generated all the data used in this manuscript. The code for generating the data, training the data, and performing the inverse analysis is available at <https://github.com/OrchardLANL/Regularization-DP-paper>. PCGA-ex.jl, RegAE-ex.jl, Fract-ex.jl, and Wave-ex.jl are the running files for the problems presented in “Principal component geostatistical approach” – “Seismic inversion” sections. Especially for the bimodal hydraulic conductivity problem, ex_bimodal.jl generates the training data set for VAE training. So these files, ex_bimodal.jl and RegAE-ex.jl need to run in order. In addition, if first time running, the related packages are available at <https://github.com/OrchardLANL>.

Received: 2 August 2022; Accepted: 21 December 2022

Published online: 13 January 2023

References

1. Ramm, A. G. *Inverse Problems: Mathematical and Analytical Techniques with Applications to Engineering* (Springer, 2006).

2. Linde, N., Renard, P., Mukerji, T. & Caers, J. Geological realism in hydrogeological and geophysical inverse modeling: A review. *Adv. Water Resour.* **86**, 86–101 (2015).
3. Zhou, H., Gómez-Hernández, J. J. & Li, L. Inverse methods in hydrogeology: Evolution and recent trends. *Adv. Water Resour.* **63**, 22–37 (2014).
4. Lake, L. *Reservoir Characterization* (Elsevier, 2012).
5. Lake, L. & Jensen, J. A review of heterogeneity measures used in reservoir characterization. *In Situ* **15**, 409–439 (1991).
6. Carrera, J. & Neuman, S. P. Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resour. Res.* **22**(2), 199–210 (1986).
7. Sun, N.-Z. *Inverse Problems in Groundwater Modeling* Vol. 6 (Springer, 2013).
8. Carrera, J., Alcolea, A., Medina, A., Hidalgo, J. & Slooten, L. J. Inverse problem in hydrogeology. *Hydrogeol. J.* **13**(1), 206–222 (2005).
9. Doughty, C. & Pruess, K. Modeling supercritical carbon dioxide injection in heterogeneous porous media. *Vadose Zone J.* **3**(3), 837–847 (2004).
10. Jayne, R. S., Wu, H. & Pollyea, R. M. Geologic CO₂ sequestration and permeability uncertainty in a highly heterogeneous reservoir. *Int. J. Greenh. Gas Control* **83**, 128–139 (2019).
11. Valkó, P. & Economides, M. J. *Hydraulic Fracture Mechanics* Vol. 28 (Wiley, 1995).
12. Pollyea, R. M., Chapman, M. C., Jayne, R. S. & Wu, H. High density oilfield wastewater disposal causes deeper, stronger, and more persistent earthquakes. *Nat. Commun.* **10**(1), 1–10 (2019).
13. Pollyea, R. M. *et al.* A new perspective on the hydraulics of oilfield wastewater disposal: How ptx conditions affect fluid pressure transients that cause earthquakes. *Energy Environ. Sci.* **13**(9), 3014–3031 (2020).
14. Wu, H., Jayne, R. S. & Pollyea, R. M. A parametric analysis of capillary pressure effects during geologic carbon sequestration in a sandstone reservoir. *Greenh. Gases Sci. Technol.* **8**(6), 1039–1052 (2018).
15. Jayne, R. S., Wu, H. & Pollyea, R. M. A probabilistic assessment of geomechanical reservoir integrity during CO₂ sequestration in flood basalt formations. *Greenh. Gases Sci. Technol.* **9**(5), 979–998 (2019).
16. Wu, H., Lubbers, N., Viswanathan, H. S. & Pollyea, R. M. A multi-dimensional parametric study of variability in multi-phase flow dynamics during geologic CO₂ sequestration accelerated with machine learning. *Appl. Energy* **287**, 116580 (2021).
17. Wu, H., Jayne, R. S., Bodnar, R. J. & Pollyea, R. M. Simulation of CO₂ mineral trapping and permeability alteration in fractured basalt: Implications for geologic carbon sequestration in mafic reservoirs. *Int. J. Greenh. Gas Control* **109**, 103383 (2021).
18. Montgomery, C. T. & Smith, M. B. Hydraulic fracturing: History of an enduring technology. *J. Pet. Technol.* **62**(12), 26–40 (2010).
19. Economides, M. J. *et al.* *Reservoir Stimulation* Vol. 2 (Prentice Hall Englewood Cliffs, 1989).
20. Greer, S. Y., Hyman, J. D. & O'Malley, D. A comparison of linear solvers for resolving flow in three-dimensional discrete fracture networks. *Water Resour. Res.* **58**(4), e2021WR031188 (2022).
21. Lecampion, B., Bungler, A. & Zhang, X. Numerical methods for hydraulic fracture propagation: A review of recent trends. *J. Nat. Gas Sci. Eng.* **49**, 66–83 (2018).
22. Russell, B. & Hampson, D. The old and the new in seismic inversion. *CSEG Rec.* **31**(10), 5–11 (2006).
23. Wang, Y. *Seismic Inversion: Theory and Applications* (Wiley, 2016).
24. Bosch, M., Mukerji, T. & Gonzalez, E. F. Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. *Geophysics* **75**(5), 75A165–75A176 (2010).
25. Zagst, R. & Pöschik, M. Inverse portfolio optimisation under constraints. *J. Asset Manag.* **9**(3), 239–253 (2008).
26. Lee, J. & Kitanidis, P. K. Large-scale hydraulic tomography and joint inversion of head and tracer data using the principal component geostatistical approach (pcga). *Water Resour. Res.* **50**, 5410–5427 (2014).
27. Kitanidis, P. K. & Lee, J. Principal component geostatistical approach for large-dimensional inverse problem. *Water Resour. Res.* **50**, 5428–5443 (2014).
28. Tarantola, A. Linearized inversion of seismic reflection data. *Geophys. Prospect.* **32**(6), 998–1015 (1984).
29. Sinha, S. *et al.* Normal or abnormal? Machine learning for the leakage detection in carbon sequestration projects using pressure field data. *Int. J. Greenh. Gas Control* **103**, 103189 (2020).
30. Zhou, Z. *et al.* A data-driven CO₂ leakage detection using seismic data and spatial-temporal densely connected convolutional neural networks. *Int. J. Greenh. Gas Control* **90**, 102790 (2019).
31. Tartakovsky, A. M., Barajas-Solano, D. A. & He, Q. Physics-informed machine learning with conditional karhunen–Loève expansions. *J. Comput. Phys.* **426**, 109904 (2021).
32. Kaipio, J. & Somersalo, E. *Statistical and Computational Inverse Problems* Vol. 160 (Springer, 2006).
33. Baysal, E., Kosloff, D. D. & Sherwood, J. W. C. Reverse time migration. *Geophysics* **48**(11), 1514–1524 (1983).
34. Yilmaz, Ö. *Seismic Data Analysis* Vol. 1 (Society of Exploration Geophysicists Tulsa, 2001).
35. Barajas-Solano, D. A. & Tartakovsky, A. M. Approximate Bayesian model inversion for pdes with heterogeneous and state-dependent coefficients. *J. Comput. Phys.* **395**, 247–262 (2019).
36. Kadeethum, T. *et al.* A framework for data-driven solution and parameter estimation of pdes using conditional generative adversarial networks. (2021) [arXiv:2105.13136](https://arxiv.org/abs/2105.13136).
37. Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nat. Rev. Phys.* **3**(6), 422–440 (2021).
38. McGovern, A. *et al.* Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Am. Meteor. Soc.* **100**(11), 2175–2199 (2019).
39. Aggarwal, H. K., Mani, M. P. & Jacob, M. Modl: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imag.* **38**(2), 394–405 (2018).
40. He, Q. Z., Barajas-Solano, D., Tartakovsky, G. & Tartakovsky, A. M. Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Adv. Water Resour.* **141**, 103610 (2020).
41. Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D. & Barajas-Solano, D. Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems. *Water Resour. Res.* **56**(5), e2019WR026731 (2020).
42. Geneva, N. & Zabarar, N. Modeling the dynamics of pde systems with physics-constrained deep auto-regressive networks. *J. Comput. Phys.* **403**, 109056 (2020).
43. Zhu, Y., Zabarar, N., Koutsourelakis, P.-S. & Perdikaris, P. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *J. Comput. Phys.* **394**, 56–81 (2019).
44. Mohan, A. T., Tretiak, D., Chertkov, M. & Livescu, D. Spatio-temporal deep learning models of 3d turbulence with physics informed diagnostics. *J. Turbul.* **21**(9–10), 484–524 (2020).
45. Mohan, A. T., Lubbers, N., Livescu, D. & Chertkov, M. Embedding hard physical constraints in neural network coarse-graining of 3d turbulence. (2020) [arXiv:2002.00021](https://arxiv.org/abs/2002.00021).
46. Mumpower, M. R., Sprouse, T. M., Lovell, A. E. & Mohan, A. T. Physically interpretable machine learning for nuclear masses. (2022) [arXiv:2203.10594](https://arxiv.org/abs/2203.10594).
47. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).

48. Wu, H., O'Malley, D., Golden, J. K. & Vesselinov, V. V. Inverse analysis with variational autoencoders: A comparison of shallow and deep networks. *J. Mach. Learn. Model. Comput.* **3**(2), 47–70. <https://doi.org/10.1615/JMachLearnModelComput.2022042093> (2022).
49. Erichson, N.B., Muehlebach, M., & Mahoney, M.W. Physics-informed autoencoders for Lyapunov-stable fluid flow prediction. (2019). [arXiv:1905.10866](https://arxiv.org/abs/1905.10866).
50. Trujillo, D. M. & Busby, H. R. *Practical Inverse Analysis in Engineering* (CRC Press, 2017).
51. Lin, Y., Le, E. B., O'Malley, D., Vesselinov, V. V. & Bui-Thanh, T. Large-scale inverse model analyses employing fast randomized data reduction. *Water Resour. Res.* **53**(8), 6784–6801 (2017).
52. Innes, M., Edelman, A., Fischer, K., Rackauckas, C., Saba, E., Shah, V.B., & Tebbutt, W. A differentiable programming system to bridge machine learning and scientific computing. (2019). [arXiv:1907.07587](https://arxiv.org/abs/1907.07587).
53. O'Malley, D., Vesselinov, V.V., Greer, S., & Harp, D.R. DPFEHM.jl. <https://github.com/OrchardLANL/DPFEHM.jl> (2020).
54. Liu, D. C. & Nocedal, J. On the limited memory bfgs method for large scale optimization. *Math. Program.* **45**(1–3), 503–528 (1989).
55. Hager, W. W. & Zhang, H. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim.* **16**(1), 170–192 (2005).
56. Mogensen, P. K. & Riseth, A. N. Optim: A mathematical optimization package for Julia. *J. Open Source Softw.* **3**(24), 615 (2018).
57. Robbe, P.J. Gaussianrandomfields.jl. <https://github.com/PieterjanRobbe/GaussianRandomFields.jl> (2020).
58. Kingma, D.P. & Welling, M. Auto-encoding variational bayes. (2013). [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
59. Doersch, C. Tutorial on variational autoencoders. (2016). [arXiv:1606.05908](https://arxiv.org/abs/1606.05908).
60. Gunaydin, D., Peirce, A.P. & Bungler, A.P. Laboratory experiments contrasting growth of uniformly and nonuniformly spaced hydraulic fractures. *J. Geophys. Res. Solid Earth* **126**, e2020JB020107. <https://doi.org/10.1029/2020JB020107> (2021).
61. O'Malley, D. *et al.* Where does water go during hydraulic fracturing?. *Groundwater* **54**(4), 488–497 (2016).
62. Bonnet, E. *et al.* Scaling of fracture systems in geological media. *Rev. Geophys.* **39**(3), 347–383 (2001).
63. Dershowitz, W., Winberg, A., Hermanson, J., Byegård, J., Tullborg, E.L., Andersson, P. & Mazurek, M. Äspö hard rock laboratory. Äspö task force on modelling of groundwater flow and transport of solutes. task 6c. a semi-synthetic model of block scale conductive structures at the Äspö hrl. *International Progress Report IPR-03-13, Swedish Nuclear Fuel and Waste Management Co., Stockholm* (2003).
64. Hyman, J. D., Aldrich, G., Viswanathan, H., Makedonska, N. & Karra, S. Fracture size and transmissivity correlations: Implications for transport simulations in sparse three-dimensional discrete fracture networks following a truncated power law distribution of fracture size. *Water Resour. Res.* **52**(8), 6472–6489 (2016).
65. Marcellus shale energy and environment laboratory. <http://mseel.org/>. Accessed: 2022-06-01.

Acknowledgements

WH and DO acknowledge support from Los Alamos National Laboratory's Laboratory Directed Research and Development Early Career Award (20200575ECR). SG acknowledges support from the United States Department of Energy through the Computational Science Graduate Fellowship (DOE CSGF) under grant number DE-SC0019323.

Author contributions

H.W.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing-original draft, writing-review & editing, visualization, supervision, project administration. S.G.: Conceptualization, software, validation, writing-review & editing. D.O.: Conceptualization, resources, writing-review & editing, supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-26898-1>.

Correspondence and requests for materials should be addressed to H.W.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023