JGR Solid Earth

RESEARCH ARTICLE

Key Points:

- A variational autoencoder (VAE) may be used to effectively assemble a diverse set of patterns in a single prior for geophysical inversion
- Geologically consistent transformations can be used to improve pattern diversity when training the VAE
- A VAE assembled prior produces less biased geophysical images than those produced by smooth inversion or a VAE trained on a single pattern

Correspondence to:
J. Lopez-Alvis,
lopez.alvis@geociencias.unam.mx

Geophysical Inversion Using a Variational Autoencoder to Model an Assembled Spatial Prior Uncertainty

J. Lopez-Alvis[1,2] , F. Nguyen[1] , M. C. Looms[2] , and T. Hermans[3]

[1]Urban and Environmental Engineering, Applied Geophysics, University of Liege, Liege, Belgium, [2]Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark, [3]Department of Geology, Ghent University, Gent, Belgium

Abstract Prior information regarding subsurface spatial patterns may be used in geophysical inversion to obtain realistic subsurface models. Field experiments require prior information with sufficiently diverse patterns to accurately estimate the spatial distribution of geophysical properties in the sensed subsurface domain. A variational autoencoder (VAE) provides a way to assemble all patterns deemed possible in a single prior distribution. Such patterns may include those defined by different base training images and also their perturbed versions, for example, those resulting from geologically consistent operations such as erosion/dilation, local deformation, and intrafacies variability. Once the VAE is trained, inversion may be done in the latent space which ensures that inverted models have the patterns defined by the assembled prior. Gradient-based inversion with both a synthetic and a field case of cross-borehole GPR traveltime data shows that using the VAE assembled prior performs as good as using the VAE trained on the pattern with the best fit, but it has the advantage of lower computation cost and more realistic prior uncertainty. Moreover, the synthetic case shows an adequate estimation of most small-scale structures. The absolute values of wave velocity are computed by assuming a linear mixing model which involves two additional parameters that effectively shift and scale velocity values and are included in the inversion.

Plain Language Summary Obtaining realistic images of the subsurface is important for characterizing processes that are sensitive to small-scale structures such as solute transport. Geophysical methods usually require additional information concerning the spatial patterns of the subsurface materials to obtain such realistic images. If more than one kind of pattern is deemed likely, enforcing a set of patterns in the geophysical image is not straightforward and traditional methods often result in over-simplified representations of the subsurface. In this work, we propose a new method that is capable of enforcing a diverse set of spatial patterns. The method is based on a pair of convolutional neural networks that form a model called variational autoencoder (VAE). The VAE is trained with a large number of samples of all the possible patterns and then it is capable of generating new patterns that are consistent with those of the training samples. The geophysical images are then constrained only to those generated by the VAE. We show that our method effectively assembles the set of possible patterns and provides a more realistic and less biased image when compared to other methods or even a VAE trained with a single kind of pattern.

1. Introduction

Obtaining a spatial model of physical properties from sparse and noisy measurements is ubiquitous in geophysics and serves important goals such as process understanding and future state prediction. This may be quantitatively framed as the solution of an inverse problem and is often simply referred to as inversion. In brief, inversion estimates the values of the spatial model parameters by combining information regarding the model itself, the measured data and a forward operator, which gives a relation between model parameters and data by describing approximately the physical process by which the data arose. When data does not provide sufficiently independent information about the distribution of subsurface properties, inversion relies on regularization to stabilize the solution (Backus & Gilbert, 1967; Tikhonov & Arsenin, 1977). An alternative is to parameterize the model by means of a linear expansion of general basis functions (Davis & Li, 2011; Jafarpour, 2011; Jafarpour et al., 2009). Both of these approaches inherently bias the solution toward an a priori constraint which may not be realistic and therefore may hinder the use of the model for certain applications. Outcrops which are representative of the local geology may provide information on spatial patterns of certain subsurface characteristics. For example, the geometry of specific layer representative of a hydrofacies may be observed. Similarly, statistical information of

Writing – original draft: J. Lopez-Alvis, F. Nguyen, M. C. Looms, T. Hermans
Writing – review & editing: J. Lopez-Alvis, F. Nguyen, M. C. Looms, T. Hermans

clay lenses in terms of shape, proportion, and size may be estimated (see Figure 2 for an example used here). If information regarding such spatial patterns of the subsurface is available it may be used together with measured data in order to improve model realism (Tarantola & Valette, 1982). This information is typically obtained from independent knowledge about the subsurface structure, for example, outcrops, boreholes, analogs. To integrate this information with measured data, the patterns must be described by techniques that account for their spatial nature (Linde et al., 2015). This has been generally achieved by using traditional geostatistical techniques, which usually provide more realistic models than classical regularization by means of imposing a covariance structure (Franklin, 1970; Hermans et al., 2012; Maurer et al., 1998), and also by using low-rank linear parameterizations such as principal component analysis or singular value decomposition and their corresponding kernel forms (Khaninezhad et al., 2012; Oware et al., 2019; Sarma et al., 2008; Vo & Durlofsky, 2014). The choice of representation technique depends on both the complexity of the spatial patterns and the information content of the measured data (Mariethoz, 2018). In general, it is recognized that multiple-point geostatistics (MPS) is more suited to reproduce highly connected spatial structures than covariance-based (or Gaussian random field) methods (Journel & Zhang, 2007; Strebelle, 2002). Recently, deep generative models (DGMs) have been proposed as an alternative to MPS to reproduce such complex spatial patterns (Chan & Elsheikh, 2019; Laloy et al., 2017; You et al., 2021).

MPS and DGMs rely on a gridded (pixel) representation for generating high-resolution spatial realizations. An Euclidian space $\mathbb{R}^N$ may be assumed for this representation where N is the number of pixels, then models may be seen as points in a high-dimensional model space. Since the spatial patterns are restricted, however, the set of possible models will not cover the whole model space. This subset may be stated by a prior probability distribution (Tarantola & Valette, 1982). While both MPS and DGMs are able to approximate such prior distribution and generate new samples with patterns similar to those contained in a training data set (e.g., a large training image, TI), DGMs present some advantages for inversion. First, contrary to MPS which either saves the number of occurrences of patterns (Straubhaar et al., 2011; Strebelle, 2002) or queries them directly from the TI (Mariethoz et al., 2010), DGMs build a continuous prior probability distribution from which spatial realizations of the patterns are generated. This continuous probability distribution means that DGMs may provide (1) more diverse patterns, that is, they generate models whose patterns are not necessarily contained in the training image, effectively interpolating between training samples, (2) a direct continuous perturbation step while exploring the model space (Laloy et al., 2017), and (3) the possibility of assembling different kinds of patterns in a single prior probability distribution (Bergmann et al., 2017). Second, given certain conditions, DGMs may also allow for gradient information (of the objective function) to be used in inversion which may substantially reduce the computational cost (Laloy et al., 2019; Lopez-Alvis et al., 2021; Mosser et al., 2018). This is typically not available for inversion with MPS, for which other ways of exploring the model space have been used (Caers & Hoffman, 2006; Hansen et al., 2012; Hu et al., 2001; Linde et al., 2015).

There were two main advances that allowed for DGMs to be applicable to high-resolution images: (1) neural networks that preserve complex spatial information, and (2) inference algorithms that are able to train instances of these networks that specifically include a continuous probability distribution within their layers. A common type of neural network that fulfills the first point are (deep) convolutional neural networks (CNNs; Fukushima, 1980; LeCun et al., 1989). CNNs are widely used in image processing and computer vision and have shown to be able to process highly complex spatial patterns (Krizhevsky et al., 2017). DGMs may use CNNs as their generative mapping and therefore produce new high-resolution samples with the training spatial patterns (Radford et al., 2016). Given the high-dimensionality of the model space, the training of such models was only possible with the introduction of inference algorithms that were able to cope with such high-dimensionality. Two main algorithms are currently used to train DGMs: amortized variational inference (Kingma & Welling, 2014; Zhang et al., 2018) and adversarial learning (Goodfellow et al., 2014). The former gives rise to variational autoencoders (VAEs) while the latter produces generative adversarial networks (GANs).

Both VAEs and GANs may be used to generate samples that display the training patterns by sampling from a n-dimensional probability distribution (where typically $n \ll N$). However, when used for inversion, the concern is not only on pattern accuracy but also on the feasibility of efficiently exploring the possible models that fit the data, or in Bayesian terms, efficiently integrating model prior information with the measured data by means of the forward operator (Canchumuni et al., 2019; Jiang & Jafarpour, 2021; Laloy et al., 2019; Mosser et al., 2018). It was recently argued that with certain choice of parameters VAEs may control both the degree of nonlinearity
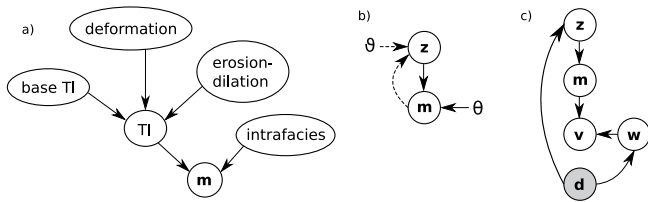
Figure 1. Probabilistic graphical models for: (a) generation of model samples using the original variables, (b) generation (continuous line) and encoding (dashed line) of model samples with the variational autoencoder (VAE) using the latent variables and (c) VAE-based inversion. **m**, **v**, **w**, **m**, **d**, and **z** refer to the model, velocity, mixing, data, and latent vectors, respectively, while θ and ϑ are the trained parameters of the VAE's decoder and encoder, respectively.

and the topological changes of their generative mapping, which in turn allows the gradient to be used in a computationally efficient inversion (Lopez-Alvis et al., 2021). Such choice of parameters is also useful in controlling the diversity of samples: instead of only generating samples very close to the training samples, the probability distribution expands or covers larger regions between the samples what can counterbalance the lack of diversity or finite nature of the training image. This improved diversity may be useful when the goal is to generate a prior probability distribution which is assembled from different types of patterns (e.g., different TIs), including the case when base patterns are perturbed by operations such as deformation, erosion-dilation, and intrafacies variability. This may be advantageous for field data because it increases the number of possible patterns in the subsurface which leads to a better representation of model prior information or uncertainty.

In this work, DGMs are used to impose spatial patterns during geophysical inversion. In particular, the ability of VAEs to build an assembled prior from different base TIs and their perturbed versions is tested. The impact of such assembled prior for modeling the subsurface is assessed by making use of gradient-based inversion for both synthetic and field cases of cross-bore-hole ground-penetrating radar (GPR) traveltime data. A recently published study by Jiang and Jafarpour (2021) in the subject of flow data matching was also aimed at representing a set of TIs with a single VAE. Using both 2D and 3D synthetic cases, they reached the similar conclusion that the VAE is able to build a prior distribution that encompasses all the patterns in the different TIs.

In this contribution, we provide an explicit comparison of VAEs performance, both quantitatively (by means of a connectivity metrics) and qualitatively, when trained using all TIs leading to a single VAE, or with individual TIs leading to their corresponding VAEs. The considered TIs have diverse spatial patterns which are automatically generated through morphological operations which is another novel aspect of this work providing a geologically consistent and continuous prior. In addition, we provide a benchmark using synthetic modeling and a validation on a real field study, which is to the best of our knowledge one of the first attempts to validate gradient-based inversion using a VAE-based prior on a real-world data set. Finally, in contrast to previous studies using DGM-based inversion (Laloy et al., 2017, 2018; Lopez-Alvis et al., 2021; Mosser et al., 2018) the values of the geophysical parameter (wave velocity) are assumed unknown and included in inversion by means of a mixing model.

The remainder of this work is structured as follows. In Section 2, an outline of the proposed framework including the underlying theory of VAEs and their use within gradient-based inversion is presented. In this section, the field data used to test the framework are also described. Section 3 presents and discusses results of the proposed approach: first, a synthetic case that mimics the field case is introduced and then results of the field case are presented. The discussion in Section 3.4 explains the importance of the proposed framework in the context of solid Earth imaging, compares it with previous studies and gives some suggestions for future work. Finally, concluding remarks of this work are presented in Section 4.

2. Methods

The framework proposed in this study is depicted in Figure 1 and may be summarized as follows:

1. Define a realistic generative model as prior distribution for the subsurface spatial patterns. The generative model may include operations that transform some base patterns such as erosion/dilation, local deformation and intrafacies variability (Figure 1a)
2. Train the VAE with samples from the generative model. Once trained, the VAE works as an assembled prior, that is, it is able to generate patterns similar to the training patterns including those transformed by the defined operations (Figure 1b)
3. Perform gradient-based inversion in the latent space of the VAE (Figure 1c)

All of the methods and concepts required in each of the previous steps are detailed in the following sections.
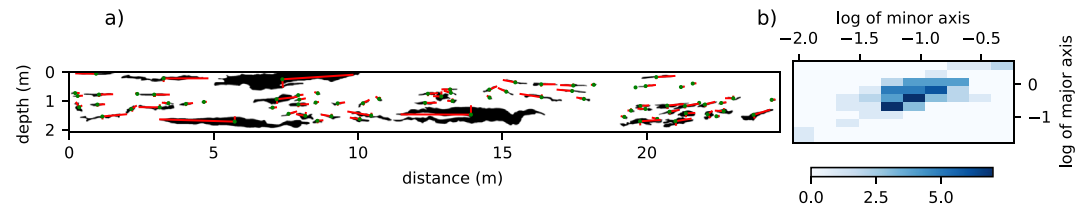
Figure 2. (a) Digitized outcrop from Kessler et al. (2013) showing sand bodies in black, background till in white, the axes of fitted ellipses for the sand bodies in red and centers of the ellipses in green. (b) Two-dimensional histogram of the major and minor axes lengths of the ellipses fitted in the outcrop.

When using methods that include geological uncertainty by means of examples of the spatial patterns, an important practical limitation is the number of available examples. The number of latent dimensions controls the capacity of the VAE to encode information regarding the spatial patterns, this means that for more diverse patterns one could try increasing the dimensionality of the latent space to improve the performance of the VAE. Ideally, the number of training examples must be also increased to cover the new space added with the extra dimensions. This means that a balance must be maintained between the number of added latent dimensions and the number of training examples available. To test this idea, we evaluate the impact of increasing the diversity of the patterns while keeping the same number of latent dimensions. This is done by considering VAEs trained on individual TIs and comparing their results with those of the VAE trained on all the TIs.

2.1. Variational Autoencoder: Approximating a Complex Probability Distribution

A VAE may be classified as a DGM. A DGM is a type of probabilistic model that relies on a relatively simple probability distribution $p(\mathbf{z})$ to approximate a more complex one $p(\mathbf{m})$ by passing the samples from the former through a (usually nonlinear) mapping, for example, a neural network (Dayan et al., 1995; Uria et al., 2014). This mapping is referred to as the generative mapping $\mathbf{g}_\theta(\mathbf{z})$ and may be represented more generally by a conditional distribution $p_\theta(\mathbf{m}|\mathbf{z})$ where θ denotes the parameters of the mapping, for example, the weights of the neural network. Here, $\mathbf{m}$ is defined in the original model space $\mathbb{R}^N$ while $\mathbf{z}$ is defined in a space $\mathbb{R}^n$. The space $\mathbb{R}^n$ is usually referred to as the latent space and $\mathbf{z}$ is called the code or latent vector. In general, samples $\mathbf{m}$ exhibit some order or structure which means they are confined to a subset $\mathcal{M} \subset \mathbb{R}^N$. This assumption is known as the "manifold hypothesis" (Fefferman et al., 2016) and means that in general it should be possible to define $\mathbb{R}^n$ with $n < N$, for which n is at minimum the dimension of the subset (or manifold) $\mathcal{M}$. This also means that the probability distribution $p(\mathbf{m})$ only needs to be defined over $\mathcal{M}$.

Assuming a large data set $\mathbf{M} = \left\{\mathbf{m}^{(i)}\right\}_{i=1}^P$ containing P samples from the complex probability distribution $p(\mathbf{m})$ is available, DGMs are trained by estimating the parameters θ of the generative mapping given a fixed $p(\mathbf{z})$. In this way, one is able to generate new samples similar to those of the training data set $\mathbf{M}$ by sampling from $p(\mathbf{z})$ and passing through the generative mapping, that is, sampling according to $p(\mathbf{z})p_\theta(\mathbf{m}|\mathbf{z})$. However, when the training samples $\mathbf{m}^{(i)}$ are high-dimensional, nonstandard inference methods are required to efficiently estimate the parameters θ of the generative mapping. VAEs use a neural network as generative mapping and rely on amortized variational inference to estimate its parameters (Kingma & Welling, 2014; Rezende et al., 2014). This inference technique requires another mapping to approximate a recognition (or variational) probability distribution $q_\vartheta(\mathbf{z}|\mathbf{m})$. In this way the generative mapping may take the output of the recognition mapping as input and vice-versa, which resembles a neural network architecture known as autoencoder (Kramer, 1991), with the generative mapping as decoder and the recognition mapping as encoder. In this work the choices proposed by Kingma and Welling (2014) regarding the probability distributions involved in a VAE are followed. The resulting framework for the VAE is detailed in Appendix A. In the rest of this work, we drop the subindex θ in $\mathbf{g}(\mathbf{z})$ to simplify notation and also because once the DGM is trained, the parameters θ do not change, that is, they are fixed for the subsequent inversion.

Note that the training data set $\mathbf{M}$ may contain different kinds of patterns which allow the VAE to effectively learn what is here termed an assembled prior, that is, a continuous prior distribution which generates not only patterns similar to those in the training set but also those corresponding to the transitions between the training patterns. Bergmann et al. (2017) propose a similar idea for GANs. One may also picture this process as changing

or substituting the original (probabilistic) generative model by the VAE, that is, the latent variables now include jointly the effects of the original variables (Figure 1).

In this work we consider a VAE in which both encoder and decoder (see Figure A1) are based on CNNs. The size of the latent vector $n = 40$ was chosen by testing a set of increasing values ($n = 20, 40,$ and 60) whose range was based on previous studies for similar patterns (Lopez-Alvis et al., 2021) and selecting the one that provides accurate reconstruction of the training samples without degrading the similarity of the generated patterns. This was assessed by visualizing a set of randomly generated models and by using the connectivity measure described in Section 2.5. We found for example, that $n = 60$ provides only a slight improvement in reconstruction of the training samples but causes a noticeable degradation of generated patterns.

2.2. Objective Function for Inversion With VAE

As mentioned above, a VAE using CNNs provides a powerful tool to represent complex probability distributions. Therefore, if one has a large data set containing examples of spatial patterns, the VAE allows to approximate complex prior probability distributions in the context of geophysical inversion. Following a Bayesian approach (to be consistent with the one used to derive the VAE), inversion may be considered as the conjunction of information regarding the model, the measured data and their relation given by a forward operator (Tarantola & Valette, 1982). The latter relation may be expressed as:

$$\mathbf{d} = \mathbf{f}(\mathbf{m}) \tag{1}$$

where $\mathbf{d}$ is a Q-dimensional vector representing the data and $\mathbf{f} : \mathbb{R}^N \to \mathbb{R}^Q$ is the geophysical forward operator. Since both the measurements and the forward operator typically have some error, the relation in Equation 1 may be represented with a conditional probability distribution $p(\mathbf{d}|\mathbf{m})$. Then, inversion is stated as:

$$p(\mathbf{m}|\mathbf{d}) = k\, p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m}) \tag{2}$$

where $p(\mathbf{m}|\mathbf{d})$ is the posterior distribution, $p(\mathbf{m})$ is the model prior distribution, $p(\mathbf{d}|\mathbf{m})$ is termed the likelihood function and k is a constant.

When the prior distribution is approximated with a VAE, inversion may be restated in terms of the latent vector $\mathbf{z}$ as:

$$p(\mathbf{m}, \mathbf{z}|\mathbf{d}) = k\, p(\mathbf{d}|\mathbf{m})\, p(\mathbf{z})\, p(\mathbf{m}|\mathbf{z})$$

$$p(\mathbf{z}|\mathbf{d}) = k\, p(\mathbf{z}) \int p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m}|\mathbf{z})\, d\mathbf{m} \tag{3}$$

where $p(\mathbf{z})$ is the latent prior distribution and $p(\mathbf{m}|\mathbf{z})$ is the generative mapping (or decoder), as defined in Section 2.1. Further, as mentioned above when considering only the mean of the decoder then $p(\mathbf{m}|\mathbf{z}) = \delta(\mathbf{m} - \mathbf{g}(\mathbf{z}))$ and Equation 3 may be written as:

$$p(\mathbf{z}|\mathbf{d}) = k\, p(\mathbf{z}) \int p(\mathbf{d}|\mathbf{m})\, \delta(\mathbf{m} - \mathbf{g}(\mathbf{z}))\, d\mathbf{m}$$

$$= k\, p(\mathbf{z})\, p(\mathbf{d}|\mathbf{g}(\mathbf{z})) \tag{4}$$

Equation 4 may be used to solve an inverse problem in which a VAE (or some other DGM) is used to state the prior model distribution. For instance, one may apply Markov chain Monte Carlo to Equation 4 and get the posterior distribution of the latent variables (Laloy et al., 2017, 2018). When appropriate values to train the VAE are used, $\mathbf{g}$ is expected to be only mildly nonlinear (Lopez-Alvis et al., 2021). If we further assume that $\mathbf{f}$ is also mildly nonlinear and that errors in the data (with respect to forward predictions) are independent and Gaussian, the likelihood $p(\mathbf{d}|\mathbf{g}(\mathbf{z}))$ will be approximately independent and Gaussian (Holm-Jensen & Hansen, 2019). Given these conditions, minimizing the following objective function $\zeta(\mathbf{z})$ should provide a good approximation for maximum likelihood model parameters:

$$\zeta(\mathbf{z}) = \|\mathbf{f}(\mathbf{g}(\mathbf{z})) - \mathbf{d}\|^2 + \lambda\|\mathbf{z}\|^2 \tag{5}$$

where $\mathbf{f}(\mathbf{g}(\mathbf{z}))$ is the composition of the forward operator after the generative mapping, $\|\mathbf{z}\|^2$ is a regularization term which enforces the search to be consistent with the multivariate Gaussian distribution $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, I_n)$ and

λ is a regularization weight (Bora et al., 2017). To minimize $\zeta(\mathbf{z})$ we take advantage of the gradient, which is computed following the chain rule as:

$$\nabla_{\mathbf{z}}\zeta(\mathbf{z}) = \nabla_{\mathbf{z}}\|\mathbf{f}(\mathbf{g}(\mathbf{z})) - \mathbf{d}\|^2 + \lambda\nabla_{\mathbf{z}}\|\mathbf{z}\|^2$$

$$= \mathbf{S}(\mathbf{z})^T\nabla_{\mathbf{m}}\|\mathbf{f}(\mathbf{m}) - \mathbf{d}\|^2 + 2\lambda\mathbf{z} \tag{6}$$

with the Jacobian $\mathbf{S}(\mathbf{z})$ of size $N \times n$ obtained directly by the autodifferentiation used to trained the VAE (Paszke et al., 2017) and whose elements are:

$$[\mathbf{S}(\mathbf{z})]_{i,j} = \frac{\partial g_i(\mathbf{z})}{\partial z_j} \tag{7}$$

The gradient with respect to the data misfit may be computed by linearization of the forward operator:

$$\nabla_{\mathbf{m}}\|\mathbf{f}(\mathbf{m}) - \mathbf{d}\|^2 = -\mathbf{J}(\mathbf{m})^T(\mathbf{d} - \mathbf{f}(\mathbf{m})) \tag{8}$$

where is $\mathbf{J}(\mathbf{m})$ is the $Q \times N$ Jacobian (or sensitivity) matrix of the forward operator whose elements are:

$$[\mathbf{J}(\mathbf{m})]_{i,j} = \frac{\partial f_i(\mathbf{m})}{\partial m_j} \tag{9}$$

Finally, inversion is done by minimizing the objective function in Equation 5 and whose gradient is computed according to Equations 6 and 8 as:

$$\nabla_{\mathbf{z}}\zeta(\mathbf{z}) = -\mathbf{S}(\mathbf{z})^T\left(\mathbf{J}(\mathbf{m})^T(\mathbf{d} - \mathbf{f}(\mathbf{m}))\right) + 2\lambda\mathbf{z} \tag{10}$$

2.3. Inversion of Traveltime Data Using a VAE as Prior

In this work, we illustrate the proposed approach with a cross-borehole GPR traveltime field data set. In order to approximate the propagation of waves, a forward operator that relies on the eikonal equation:

$$|\nabla\tau|^2 = v^{-2} \tag{11}$$

is used, where τ denotes the traveltime and v is the velocity of the subsurface materials. Note that Equation 11 is not limited to GPR but may also be applied to for example, seismic traveltime. A numerical solution is typically required, where after discretization one obtains the forward operator that relates the vector of traveltimes $\mathbf{d} = \tau$ to the slowness (which is the reciprocal of velocity) vector $\mathbf{m} = \mathbf{v}^{-1}$ in Equation 1. The Fast-Marching method and a factorized version of the eikonal equation are used herein (Treister & Haber, 2016). The factorized equation helps to reduce the error induced by spatial discretization in the proximity of the sources. It is important to note that this forward operator may still result in noticeable error when used for field data since effects related to the finite-frequency or scattering are not considered. When a proper discretization is chosen and a moderate velocity contrast is assumed, the magnitude of this error is comparable to the one of measurement error (Hansen et al., 2014) which should allow for data misfit error only a bit higher than with more realistic operators. Though, a nonnegligible bias remains which must be considered when analyzing inversion results. The same implementation allows one to efficiently compute the product $\mathbf{J}(\mathbf{m})^T(\mathbf{d} - \mathbf{f}(\mathbf{m}))$ which is given by the solution of a triangular system exploiting the Fast-Marching sort order of the forward operator (Treister & Haber, 2016). The choice of such forward operator is motivated by the need to keep computational demand low, as inversions usually require a significant amount of both forward simulations and the above sensitivity product.

In contrast to previous studies where synthetic cases assumed that the mean velocity values in each facies were known (Canchumuni et al., 2019; Laloy et al., 2017, 2018; Mosser et al., 2018), here the inversion of these velocity (or slowness) values is done by assuming a linear mixing model that shifts and scales the spatial models obtained from the VAE according to $\mathbf{v} = w_1 + w_2\mathbf{m}$. This is helpful for field cases since typically there is uncertainty in these values. The inversion will then include two extra parameters (w_1 and w_2). If these parameters are assumed independent of the latent vector $\mathbf{z}$, one may compute the gradient of the objective function with respect to them:

$$\frac{\partial \zeta(\mathbf{w})}{\partial w_i} = \nabla_{\mathbf{v}} \|\mathbf{f}(\mathbf{v}) - \mathbf{d}\|^2 \frac{\partial \mathbf{v}}{\partial w_i} \tag{12}$$

and for the two w_i parameters we have:

$$\frac{\partial \mathbf{v}}{\partial w_1} = \mathbf{1}, \ \frac{\partial \mathbf{v}}{\partial w_2} = \mathbf{m} \tag{13}$$

Similarly, the first term on the right of Equation (6) should now be computed using $\mathbf{v}$ instead of $\mathbf{m}$. This term also needs to be multiplied by the derivative with respect to $\mathbf{m}$, which is equal to w_2. Since these two parameters cause a stronger impact on traveltime values than the latent variables, their step is multiplied by a factor equal to 10^{-4} to make the inversion stable.

In this work, inversion is done by optimizing the objective function in Equation 5 using stochastic gradient descent (SGD) and Equation 10. SGD provides two main advantages: (1) it is less prone to get trapped in local minima, especially if the objective function has the shape of a global basin of attraction, and (2) the computational cost of each iteration is reduced by simulating only a subset of the data (also called a data batch). Decreasing of the step size (or learning rate) is also employed as it has been shown to further aid in reaching the neighborhood of the global minimum (Kleinberg et al., 2018). Such decreasing was done only every three iterations and using 0.99 as decreasing factor.

2.4. Field Site and Data Description

The field site is located at the Kallerup gravel pit, Denmark. The local geology is composed by a glacial till with several elongated sand bodies (Kessler et al., 2012). Till is composed of particle sizes from clay to gravel, while sand bodies have a more narrow grain size distribution. Further, shapes of the sand bodies display varying degrees of deformation characteristic of basal till. This type of geology results in highly contrasting subsurface, as may be seen in Figure 2a. After the data was acquired, the field site was excavated which allows to compare with inversion results, at least qualitatively (Bording et al., 2019; Larsen et al., 2016).

The field data set is the cross-borehole traveltime data presented by Looms et al. (2018) and is available at Looms et al. (2021). Measurements were collected with 100 MHz borehole antennas and a PulseEKKO system (Sensors & Software, ON, Canada). The two boreholes are located 3.25 m apart and are 8 m deep. Data was acquired forming a multioffset gather (MOG) with all source positions in one borehole and receiver positions in the other. Spacing for both sources and receivers was 0.25 m and data was collected from 1.0 to 7.0 m deep, for a total of 625 traces. First arrivals were picked with a semi-automatic procedure (Looms et al., 2018). Data for sources and receivers with depth less than 1.5 m were removed to avoid error from refraction at the air-ground interface. For similar reasons, since the boreholes are located in the unsaturated zone, data offsets with angles greater than 30° were not considered to avoid error from borehole refraction. Estimated measurement error is 0.47 ns while average traveltime is 41.5 ns.

2.5. Training VAE With Realistic Patterns Based on an Outcrop

The size of the spatial domain to be modeled was selected according to the region sensed by the acquisition setup (see details on Section 2.4). A uniform cell discretization of 5 cm was chosen to model high-resolution details. Although CNNs may be set to the desired dimensions by selecting the correct size for the filters, stride and padding, one could also consider a slightly larger size and then crop the cells outside the domain since they do not affect the data misfit. In this work, some cells close to the surface are retained even if they are outside the sensed volume because they allow a qualitative assessment of the effect of the prior pattern information in the absence of data. Therefore, the spatial domain was discretized by $65 \times 129 = 8,385$ cells, corresponding to a 3.25×6.45 m section.

The training patterns used to train the VAE are constructed by a hierarchical model that allows for the transformation of an initial set of TIs (Figure 1a). The sensed subsurface was assumed to be mainly composed by two different materials: till and sand. Two initial object-based TIs (BTI_1 and BTI_2) were built according to information on local geology and a quantitative analysis of an outcrop close to the investigated cross-borehole section (Kessler
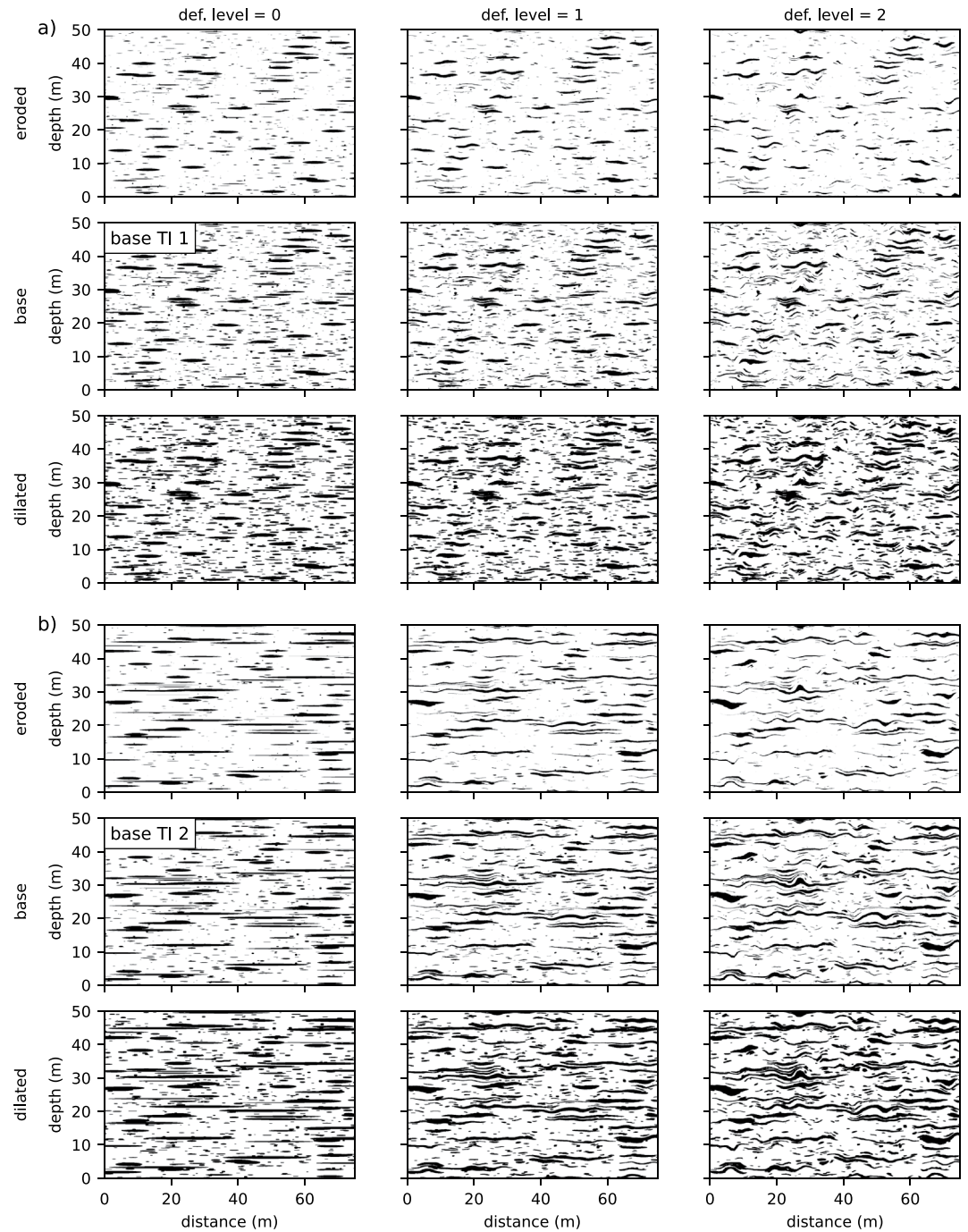
Figure 3. $1,500 \times 1,000$ pixel croppings of the nine modified TIs corresponding to both (a) BTI_1 and (b) BTI_2.

et al., 2013). These two TIs were mainly chosen because there is uncertainty in the presence of sand sheets (the most elongated sand bodies) in the sensed region: they were not present in the outcrop used in the analysis but they were present in other outcrops. All of the sand bodies were assumed to be approximated with ellipses of different sizes and eccentricity (Figure 2a). For this, the statistical distribution of the major and minor axes of the sand bodies was approximated from the outcrop by a two-dimensional histogram (Figure 2b). Then, BTI_1 is directly constructed by sampling ellipses sizes according to the histogram, placing them randomly in the domain (overlapping is allowed to partially account for the more complex shapes) while maintaining a facies proportion similar to the one in the outcrop which is 0.17 (Figure 3a). BTI_2 is built similarly but includes the sand sheets

(Figure 3b) whose size distribution was based on the one reported by Kessler et al. (2012). The size of these TIs was chosen in order to include many repetitions of the patterns for the target size to be simulated (65×129), therefore TIs with a size of $4,762 \times 4,762$ are used.

To account for more diverse and realistic shapes for the sand bodies (as those seen in the outcrop) two main transformations were applied to the initial TIs: erosion/dilation and local deformation. Erosion/dilation here refers to the image morphological operation for which pixels are removed/added to the limits of objects by setting a pixel to the minimum/maximum over all pixels in a neighborhood centered at that pixel (Soille, 2004). Though erosion/dilation may refer to either of the two facies, here we will refer to that of the sand bodies to avoid confusion. One step for dilation and one for erosion was done using a neighborhood with six pixels in the x direction and two pixels in the y direction. The local deformation was done by a piecewise affine transformation (van der Walt et al., 2014) which requires defining a uniform grid of nodes and a corresponding mesh by Delaunay triangulation. Then, the positions of the nodes were perturbed according to two Gaussian random fields (one for the x coordinates and one for the y coordinates) and finally a local affine transformation is done to the pixels inside each triangle of the original mesh in order to fit the new deformed mesh. Deformation was applied with two different amplitudes in the perturbation of the grid, resulting in two different levels of deformation. Considering all the combinations of erosion-dilation and deformation (including the ones with no erosion-dilation and zero deformation) a total of nine different cases or modified TIs for each base TI are built. The patterns of each of all 18 modified TIs obtained are shown in Figure 3. The size of each of these modified TIs is a bit smaller ($4,722 \times 4,722$) than for the base TIs since cropping was needed in the edges after deformation. Given this setting a significant overlapping between the different TIs is expected, so that there are no regions with very low probability between the different modes of the prior, that is, the prior distribution is approximately unimodal.

Finally, intrafacies variability was considered by means of using Gaussian field simulations with different means and anisotropy for each facies: both facies use a Gaussian covariance function with correlation length of 1.0 m but the channels facies uses an anisotropy factor of 0.2 and a mean of 0.35 (prior to transforming to velocity values) while the background facies uses a factor 0.25 and a mean of 0.7. This variability was added following a "cookie cutter" approach where each of the simulations is only set in pixels with the corresponding facies value. Values were log-transformed in order to prevent negative values. This step is done after the sample is cropped from the modified TI to train the VAE to allow more variability in the patterns. The overall hierarchical model from where training samples for the VAE are taken is shown in Figure 1a. Note that the transformations are coherent with the geological processes and one could also easily include others such as faulting.

To aid in assessing the accuracy of the VAE generated samples, a measure of connectivity is used. Connectivity is an important property of geological media (Renard & Allard, 2013) and its estimation at different scales has been proposed as a useful measure of similarity between different models of the subsurface (Laloy et al., 2017; Lemmens et al., 2019). We used the connectivity function (Torquato et al., 1988) that has been previously used to evaluate the generation performance of DGMs (Laloy et al., 2017, 2018). In brief, this connectivity function gives the probability of two points of a certain phase or material to be connected for a given lag distance. Considering a number of lag distances (in number of pixels) a connectivity curve is obtained, which characterizes connectivity at different scales. The connectivity is dependent on the orientation and for our study we considered three different orientations: vertical (parallel to y axis), horizontal (parallel to x axis) and diagonal (45° between x axis and y axis).

To assess the performance of our proposed inversion approach, a synthetic case is first analyzed with the same acquisition settings as those of the field data. A synthetic model was built with the same statistical distribution of BTI_2 but with a higher proportion of sand to till proportion (0.32) and different degree of deformation (an amplitude just in the middle between 1 and two in Figure 3). The model was cropped from a TI of the same size as the ones used for training but its random spatial realization was different, that is, the ellipses and its positions were randomly set, therefore one should expect different patterns may be present than those in the TI used for training. Then, synthetic data were generated using the forward operator and Gaussian noise with the same magnitude as the error estimated for the field data was added (0.47 ns). Note that in this case, there is no error due to the forward operator. In this way, the synthetic case should provide an idea of how performant is the VAE-based inversion in obtaining patterns that deviate from the ones used for training.

3. Results

3.1. Training the VAE

The VAE for the assembled prior is trained by randomly selecting from any of the 18 modified TIs, then randomly sampling a cropped piece (with the appropriate size of the spatial domain) and adding the intrafacies variability. Examples of the cropped samples are shown in Figure 5a. The VAE was implemented and trained using PyTorch (Paszke et al., 2017). The training used a total of $P = 10^7$ cropped samples and took around ~4.5 hr on a Nvidia GPU RTX 2060 (~3 hr without the intrafacies). Note that deformation and erosion-dilation may have been done directly while feeding the samples to train the VAE (similar to the intrafacies), however, this would have likely resulted in prohibiting computational time (while erosion-dilation is typically fast, the local deformation is generally much slower). Qualitative performance of the VAE capability is assessed by checking the reconstruction (i.e., encoding and then decoding) of some training samples. This is shown in Figure 5b where one can notice some degradation in terms of both the sharpness of the limits and some flattening of highly deformed bodies. Once trained, samples are generated according to the graphical model in Figure 1b (following the process defined by Figure A1). A few examples of random samples generated from the trained VAE are shown in Figure 5c, these are samples from the assembled prior distribution approximated by the VAE.

The connectivity function in three different orientations (Section 2.5) is also used to quantitatively evaluate the generative performance of the VAE. Figure 4a shows statistics of connectivity curves for 360 VAE generated model samples together with those of 360 training samples. The connectivity in the three orientations follows approximately the same trend: the 25th percentile curve from the generated samples matches that of the training samples across all lag distances while both 50th and 75th percentile curve from the generated samples underestimates that of the training samples. Notably, the 75th percentile curve is significantly lower which means that generated samples tends to under-represent high connectivities at all scales. This is consistent with what is observed when comparing generated model samples (Figure 5e) with training samples (Figure 5a): a higher proportion of shorter (horizontally) and narrower (vertically) sand (black) bodies seems to be present in the generated samples. This indicates that while the assembled VAE is able to generate relatively accurate patterns, the diversity in the generated samples is slightly reduced when compared to that of the TIs. To check the impact of increasing the number of latent dimensions, we also trained an assembled VAE using 60 latent dimensions. Surprisingly, the samples generated with this VAE cause a stronger reduction in diversity; a possible explanation is that there is a larger space between training samples in latent space and the VAE fills it with models of moderate connectivity.

To assess the performance of the VAE in reconstructing samples out of the training set, a validation set was cropped from a new set of 18 modified TIs obtained from two different base TIs. These two base TIs are built following the same process as that for BTI_1 and BTI_2 but with different samples for the sizes and positions of the ellipses. Some examples of these validation samples are shown in Figure 5d and their reconstructions using the VAE are shown in Figure 5e. A visual assessment of the validation samples and their reconstructions shows that the VAE has a similar reconstruction accuracy than that of the training samples which indicates the ability of the VAE to generalize, that is, work correctly beyond training samples.

As mentioned in Section 2, we also train VAEs on each of the 18 individual TIs to test the impact of increasing diversity in the training patterns without increasing the number of latent dimensions. In this way, one can picture the case of the multi-TI VAE as using the same volume (in latent space) as the single-TI VAE to encode a more diverse set of patterns. This causes the training samples to be represented closer together in latent space for the multi-TI VAE that is, we have to encode 18 times more samples in the same volume of latent space. The statistics of the connectivity curves for this case are shown in Figure 4b. It can be seen that the accuracy and diversity of the generated patterns for this case is very similar to that of the assembled VAE and there is only a slight improvement in the match of the 75th percentile curve which means a slightly higher proportion of wider and thicker sand bodies. Also, as a comparison the statistics of the connectivity curves for the VAE trained using only TI_1 are shown in Figure 4c where a significantly lower diversity is evidenced by tighter percentile curves.

The VAE-based generated patterns may fail to adequately represent the patterns of heterogeneity encountered in the field for three main reasons: (1) sufficiently similar patterns are not included during training, (2) patterns are filtered or simplified by the VAE, and (3) the diversity of the patterns was not sufficient to simulate new consistent patterns. In general, these three reasons play a role to different degrees. The first is unavoidably present in any study that aims to use information from nearby outcrops or local geology to constrain the subsurface patterns in
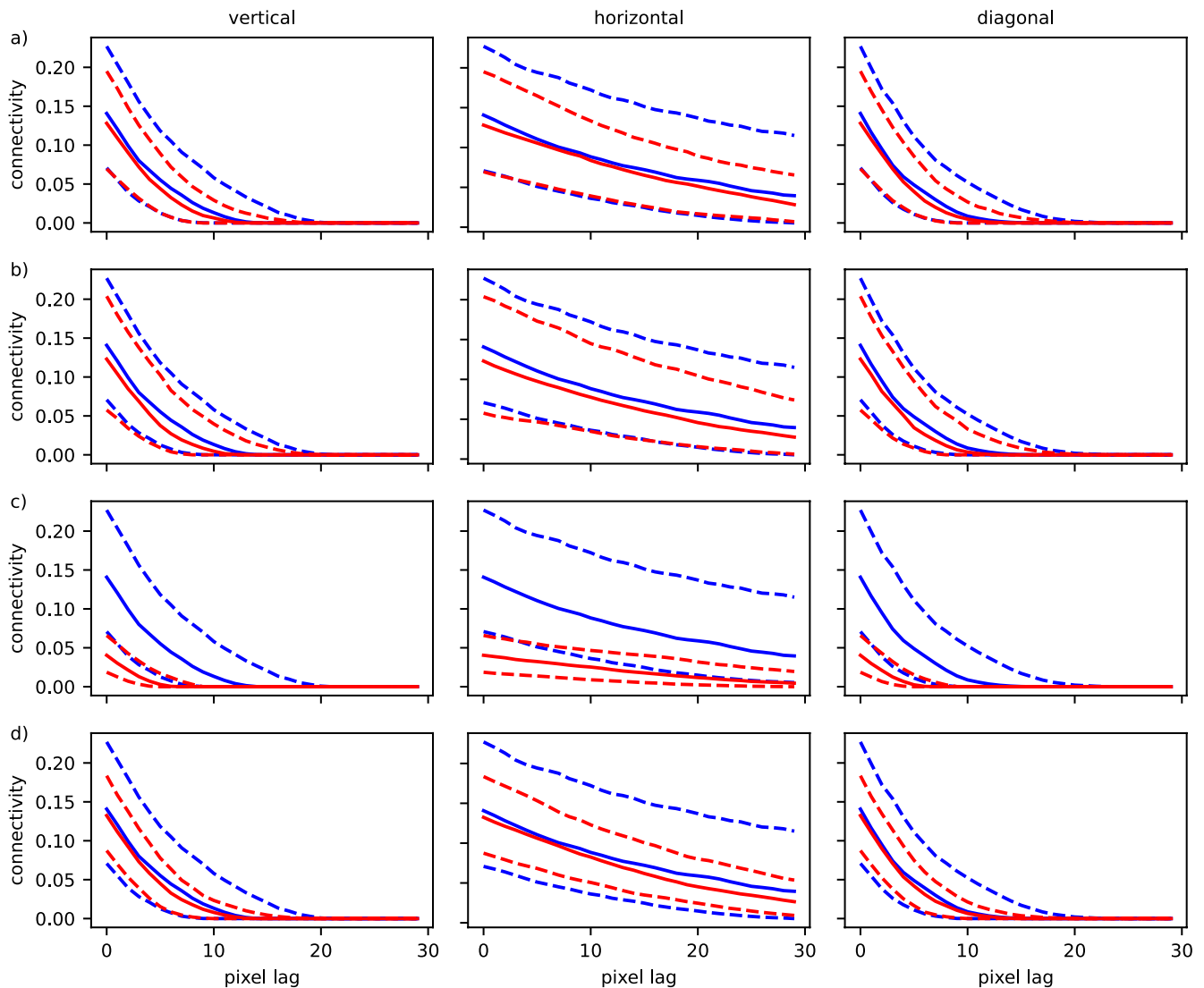
Figure 4. Statistical performance of the variational autoencoders (VAEs) in terms of connectivity: Percentiles 50 (continuous line), 25 and 75 (dashed line) of the connectivity curves computed from 360 model samples. The red lines denote curves from VAE generated model samples and the blue lines curves from the training samples. Generated samples come from (a) assembled VAE, (b) all 18 individual VAEs, (c) individual VAE for TI_1, and (d) assembled VAE with 60 latent dimensions.

the sensed domain. However, this may be partially accounted for by considering different base patterns and their perturbed versions (obtained by morphological operations and local transformations) which may all be attributed to a similar environment. Note, however, that this strategy will not add new materials (lithologies). A prior consistency check before training may indicate if the VAE fails due to the first reason. In this work, this check was done using a methodology based on a low-dimensional representation of the data (Hermans et al., 2015; Lopez-Alvis et al., 2019; Park et al., 2013; Scheidt et al., 2018) according to which none of the TIs is falsified, that is, all the proposed patterns are likely to have generated the data. The details are shown in Supporting Information S1. The effect arising from the second reason is directly related to generative accuracy and is captured for example, in Figures 5b, 5c, and 5e where the generated and reconstructed samples seem to have filtered out patterns with very high curvature. Finally, the third reason, which is somewhat tied to the first, is related to how the VAE is able to interpolate between training patterns. As noted above, the statistics of the connectivity curves of generated models (4a) give a first indication that diversity is being slightly reduced. This may also be checked by visualizing a set of training images as in Figure 5c and also making a latent traversal as shown in Figure 6, which makes steps along two of the dimensions of the latent space and fixes the rest. This should also be supplemented by an assessment of how much the generated patterns depart from the training samples while retaining consistent patterns.
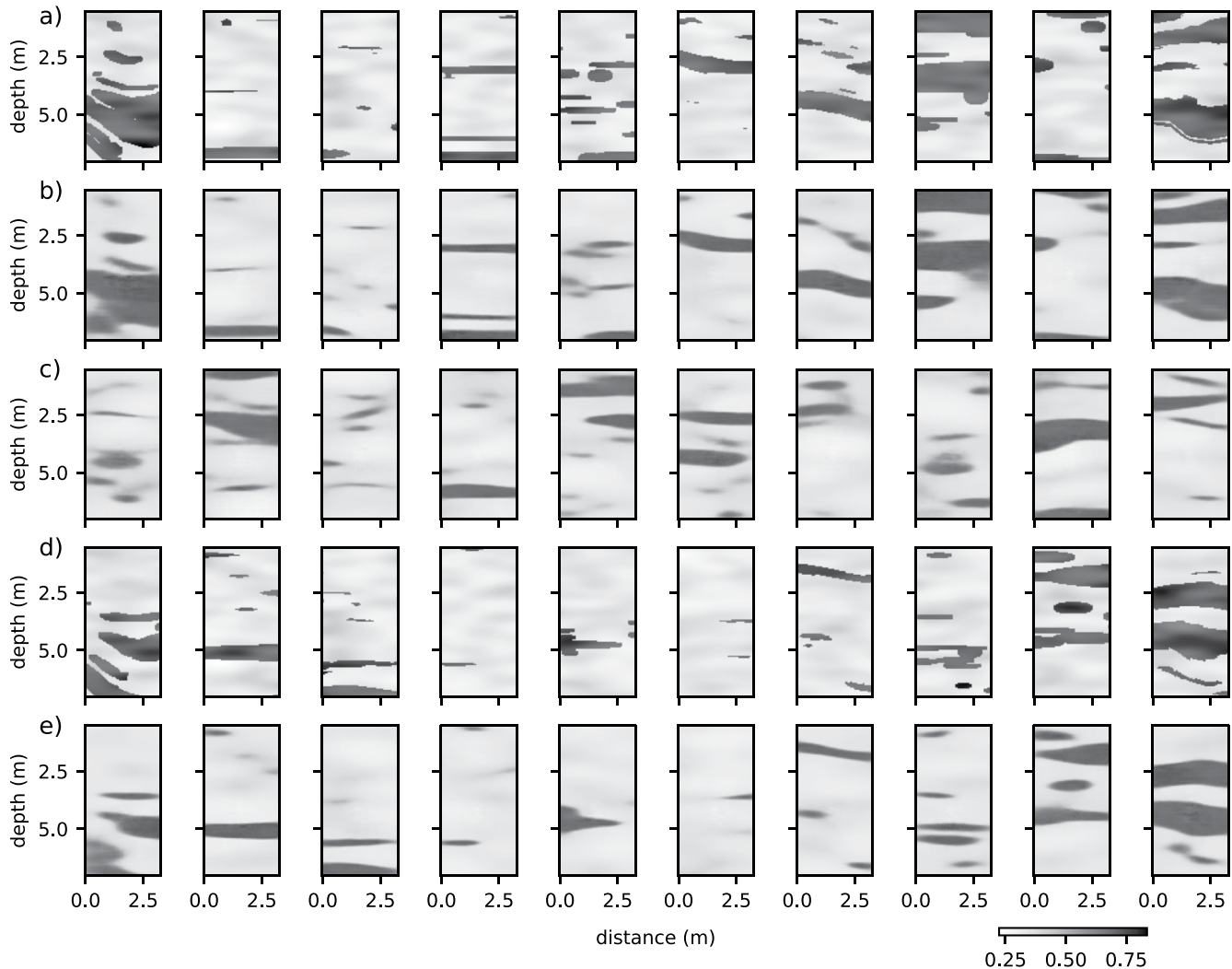
Figure 5. Examples of training samples (a), reconstruction of these training samples (b), samples generated from the trained variational autoencoder (VAE) (c), examples of validation samples (d), and reconstruction of these validation samples (e). The gray scale is with respect to the model variable **m** prior to its transformation to velocity values.

In recent work, Lopez-Alvis et al. (2021) show that VAEs are able to deviate from training patterns while still preserving realistic patterns through breaking continuous channels from the original training image. There have been some recent efforts to quantitatively measure diversity in DGMs (Lucic et al., 2018; Sajjadi et al., 2018) however, it remains an open question whether useful departures (such as breaking channels) would be adequately captured by these measures. A possible solution to this issue would be to directly include such departures in the training set. In summary, the proposed approach is not intended to generate perfectly accurate patterns but to allow the generated patterns to deviate from training patterns in order to both improve diversity and fit the data without compromising the patterns' realism.

3.2. SGD-Based Inversion of Synthetic Data With VAE as Prior

Once the VAE is trained, the assembled prior may be used directly in inversion to impose the diverse patterns. It is worth noticing that the latent parameters **z** of the VAE have effectively substituted the parameters related to the original hierarchical model (the new generative model is shown in Figure 1b). We assume that the latent parameter distribution now includes most of the discrete and intractable operations (i.e., different base TIs, erosion-dilation, deformation, and intrafacies variability) in a continuous and searchable space. This allows for optimization to be performed by continuously stepping in the latent space. Moreover, such steps can take advantage of the
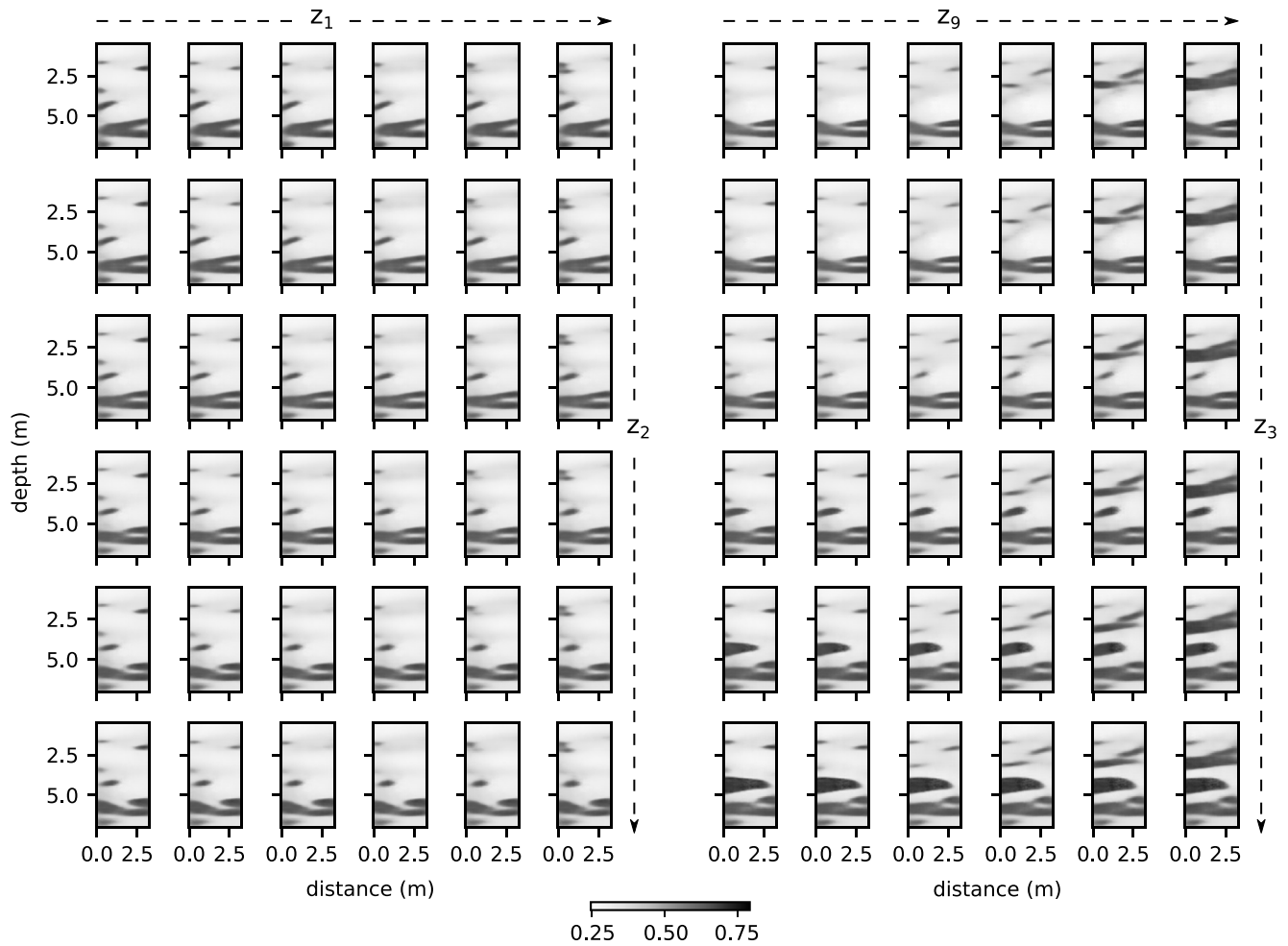
Figure 6. Examples of variational autoencoder (VAE) latent traversals (stepping in two latent dimensions while keeping the rest fixed) for: latent dimensions z_1 and z_2 (left) and latent dimensions z_9 and z_3 (right). The gray scale is with respect to the model variable **m** prior to its transformation to velocity values.

gradient (as detailed in Section 2.3) which generally would not be the case if one sought to directly estimate the original parameters.

The results of our proposed inversion approach are first assessed using the synthetic data presented above. Figures 7a–7c shows the real synthetic model, the reconstruction (encode-decode operation) of this model with the trained VAE and an inverted model with a regularized least squares solution, respectively. The regularized inversion is done with a regularization factor of 10^4 which represents the use of generic prior information; the effect of smoothing is noticeable and the model misses some spatial features when compared against VAE-based inversion (Seo et al., 2019 report a similar comparison). In the case of VAE-based inversion, since we are using SGD which is a stochastic optimization method, inversion is done for 10 different starting models. Inversion results for four different starting models are shown in Figures 7e–7h. For the model in Figure 7e, the initial model (decoding of the initial **z**) is shown in Figure 7d. For the same inverted model, the behavior of the data misfit (RMSE), the Euclidian distance between the current model and the real model, the norm of **z** and the velocity parameters as the inversion progresses are shown in Figures 7l–7o. The norm of **z** is useful to check that the algorithm does not diverge from the prior. This is because the prior $p(\mathbf{z})$ is multivariate Gaussian $\mathcal{N}(\mathbf{0}, I_n)$, then models consistent with the prior should not be far from the origin and also models with the most common patterns should be centered according to a χ-distribution with d degrees of freedom. To assess the impact of the assembled prior compared to VAEs trained on individual TIs, inversion is done also for each of the individual cases. Considering 10 different starting models for each case, the mean and standard deviation of data RMSE, norm of **z**, and velocity parameters are computed. These values are shown in Table 1 for the best, median and worst individual TIs in terms of mean RMSE together
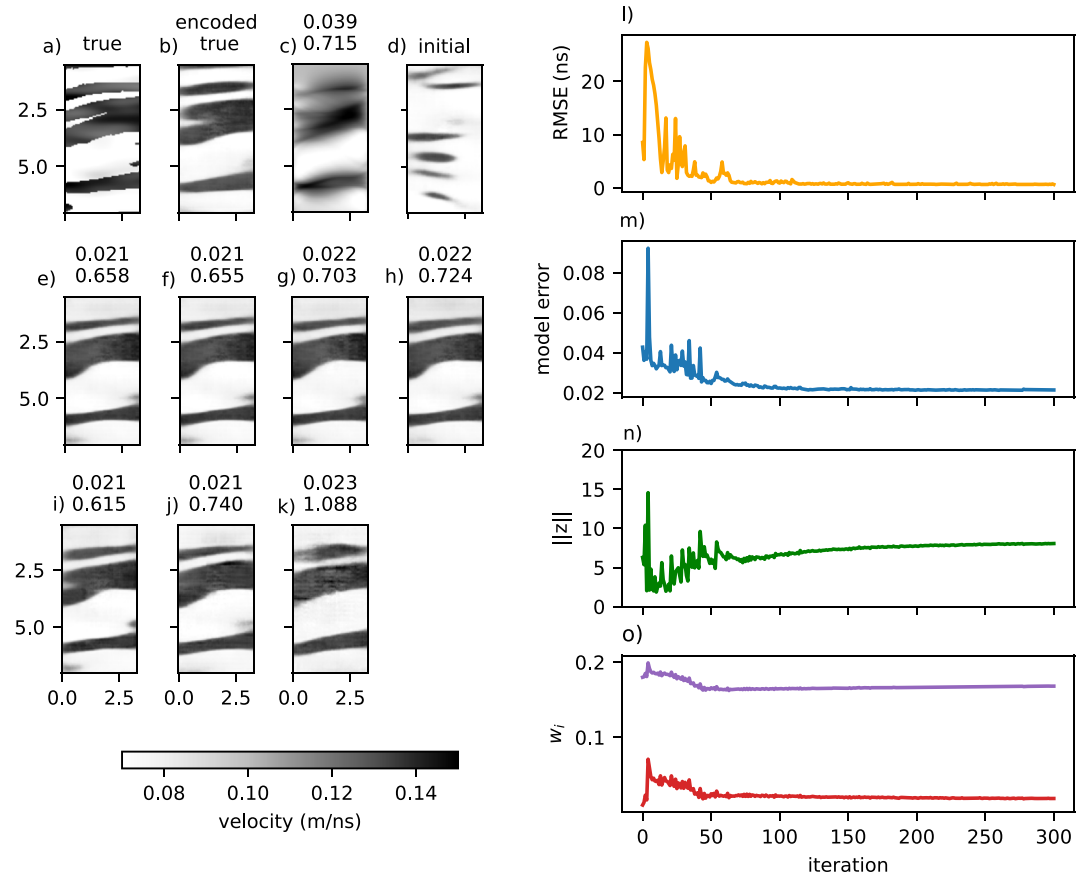
Figure 7. Inversion results for the synthetic case: (a) True model, (b) reconstructed true model, (c) regularized inverted model, (d) initial model, (e–h) variational autoencoder-stochastic gradient descent (VAE-SGD) inverted models for four different starting models using the assembled prior. VAE-SGD inverted models for prior with individual TIs using one random starting model: best (i), median (j), and worst (k) in terms of RMSE (see Figure 8a). For all inverted models, model RMSE and data RMSE are shown at the top. For the model in (e), the values in each iteration for: data RMSE (l), model RMSE (m), norm of $\mathbf{z}$ (n), and linear mixing parameters (o).

Table 1
Mean and Standard Deviation Values of Inversions Using 10 Different Initial Models

TI[a]	Data RMSE (ns)	$\|\mathbf{z}\|$	v_1	v_2
Synthetic case				
All	0.655 ± 0.050	8.004 ± 0.309	0.017 ± 0.005	0.17 ± 0.007
Best	0.632 ± 0.017	7.767 ± 0.124	0.019 ± 0.001	0.166 ± 0.002
Median	0.728 ± 0.011	8.325 ± 0.072	0.017 ± 0.001	0.171 ± 0.001
Worst	1.058 ± 0.018	10.097 ± 0.326	0.015 ± 0.001	0.175 ± 0.003
Field case				
All	0.634 ± 0.008	5.342 ± 0.244	0.031 ± 0.001	0.157 ± 0.004
Best	0.623 ± 0.013	5.194 ± 0.124	0.029 ± 0.001	0.157 ± 0.004
Median	0.674 ± 0.041	5.155 ± 0.294	0.033 ± 0.003	0.148 ± 0.010
Worst	0.732 ± 0.035	5.371 ± 0.214	0.031 ± 0.001	0.150 ± 0.004

[a]The labels indicate best, median and worst in terms of data RMSE from all 18 TIs.
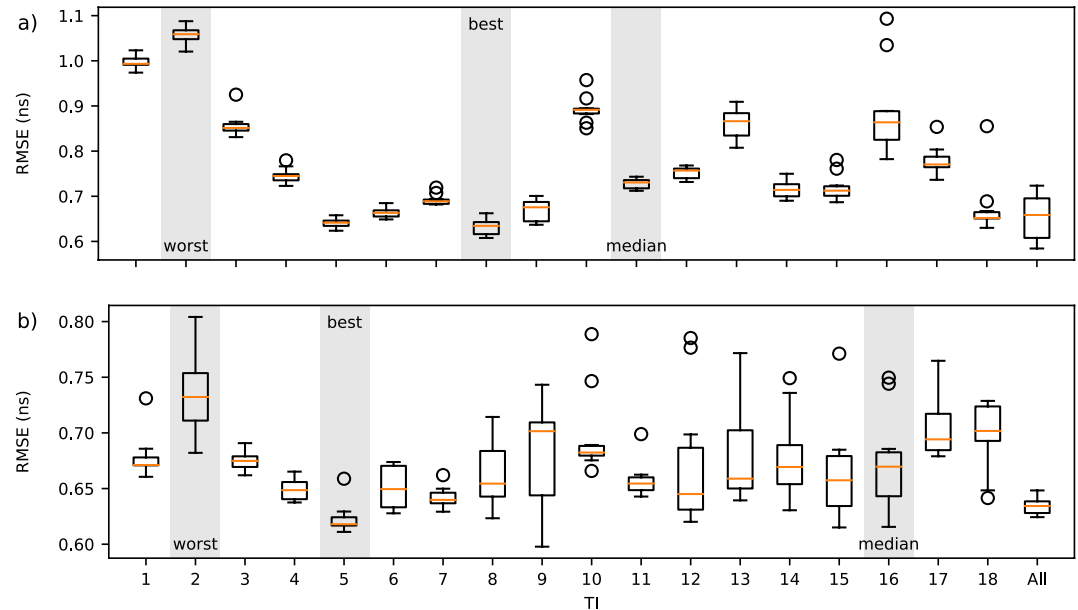
Figure 8. RMSE boxplots for synthetic (a) and field case (b) for individual priors (variational autoencoders (VAEs) trained on each of the 18 TIs) and the assembled prior (labeled "All"). Shaded areas indicate best, median and worst performing individual prior in terms of mean data RMSE.

with those of the assembled prior. Boxplots of the data RMSE for all individual TIs and the assembled prior are shown in Figure 8a. Notice that the mean data RMSE for the assembled prior (0.655 ns) is only slightly higher than the magnitude of the added noise.

To test our VAE-based inversion with different synthetic models, 10 true models are taken from the training samples (Figure 9a) and 10 true models are taken from the validation samples (Figure 9d). For this case, only one initial model was used for each inversion. Inversion results models are shown in Figures 9b and 9e. Similarly, to allow for a comparison, inverted models using all the single-TI VAEs are shown Figures 9c and 9f. The inverted models match the true models in most large-scale features (although these sometimes have differing width and thickness) but seem to miss some small-scale features. Overall there are no significant changes in inversion results due to the presence of true models in the training samples nor due to the use of the assembled VAE versus the individual VAEs.

To analyze the impact of prior information (as represented by the VAE) on inversion results, one must also consider how much information content is provided by the data, that is, how much the data constrains the posterior distribution. In this work, the cross-hole traveltime data set is considered informative enough to produce relatively similar inverted models, however, since a high-resolution model is desired, the choice of prior information (and the way it is imposed) still causes noticeable variations in inverted models (Day-Lewis et al., 2005). The inversion results for the synthetic case in Figure 7e show that although reconstruction is not perfect, the method is able to identify most of the structural characteristics of the real model. The inverted model is noticeably better than traditional regularized inversion (Figure 7c), which shows higher data RMSE and from which it is not possible to identify small features (at 5 m depth in the right and close to 7 m depth on the left in Figure 7a). On the other hand, both inversion methods miss a low velocity structure (at 3 m depth on the left of Figure 7a) and most of the intrafacies variability. The VAE-SGD inversion even locally biases the model in order to account for the lack of intrafacies variability (note a more pronounced bend of the lower part of the sand body at ~4 m depth to make up for a low velocity intrafacies zone). Most likely this comes from the fact that the synthetic model is not exactly within the prior. Since no error in the forward operator model is introduced for the synthetic case, the RMSE value higher than the noise level indicates that deviations in the inverted model are mainly due to the prior, whose accuracy slightly degrades due to a joint effect of the three reasons mentioned in Section 3.1. The synthetic case shows that our proposed inversion still provides useful results even when the patterns of the real model differ slightly from those of the TIs used to train the VAE.
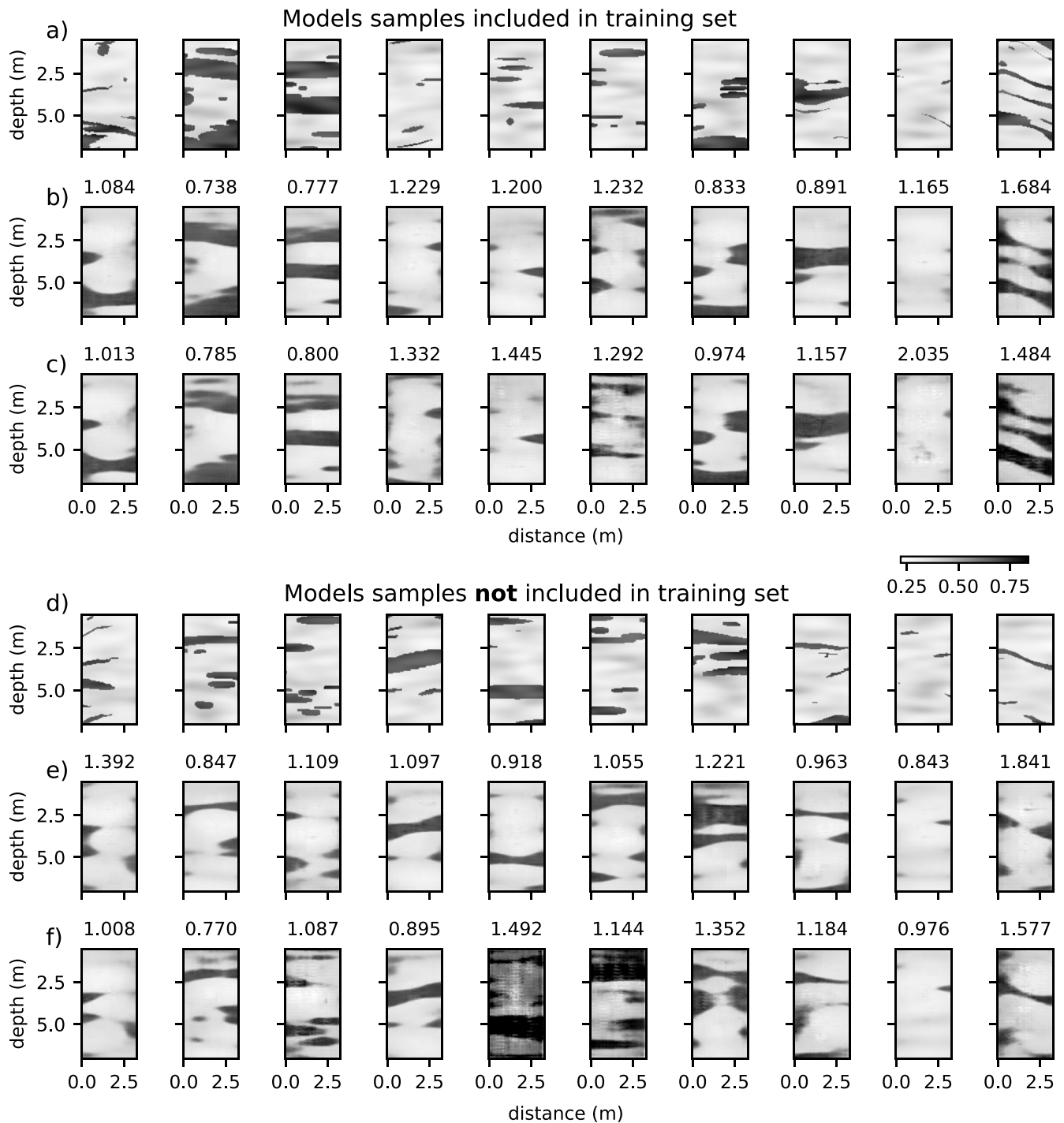
Figure 9. Variational autoencoder (VAE)-based inversion with true models in the training samples (a) and true models in the validation samples (d). Inverted models using the assembled VAE (b, e) and inverted models using the VAEs trained on individual TIs (c, f). Data RMSE are shown at the top of each inverted model. The gray scale is with respect to the model variable **m** prior to its transformation to velocity values.

The impact of noise on inversion results was also tested by considering two different noise levels: one half (0.235 ns) and the other twice (0.94 ns) the original noise level. Five inverted models with each of these two noise levels are shown in Supporting Information S1. Inverted models with lower noise show very similar results to those with the original noise level and the data RMSE only improves marginally, which is consistent with our conclusion above that the misfit is mainly due to prior approximation error. Inversion results for the higher noise
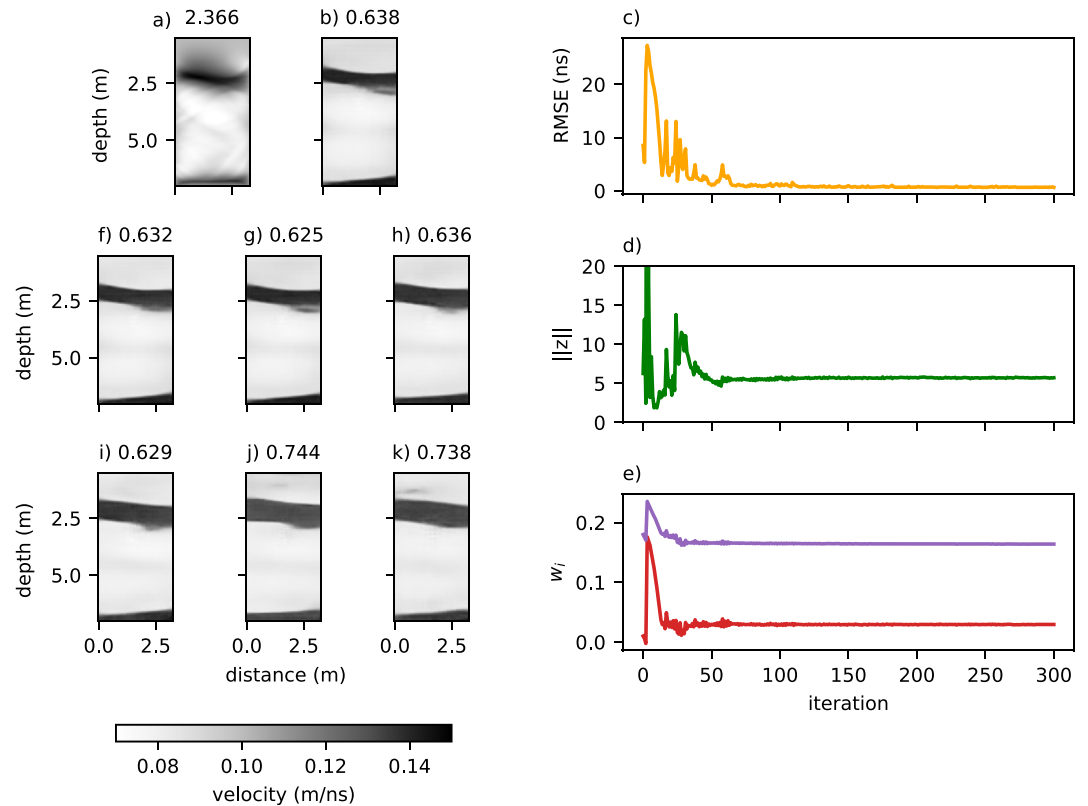
Figure 10. Inversion results for the field case: (a) regularized inverted model, (b) variational autoencoder-stochastic gradient descent (VAE-SGD) inverted model for one random starting model using the assembled prior. For the model in (b), the values in each iteration for: data RMSE (c), norm of **z** (d) and linear mixing parameters (e). VAE-SGD inverted modes for three different starting models using the assembled prior (f–h). VAE-SGD inverted models for prior with individual TIs using one random starting model: best (i), median (j), and worst (k) in terms of RMSE (see Figure 8b). For all inverted models, data RMSE is shown at the top.

level show data RMSEs much closer to the noise level of 0.94 ns and still a good estimation of the true model, however, there seems to be a higher variability between models inverted with different initial models which indicates SGD is converging to a wider area around the global minimum.

3.3. SGD-Based Inversion of Field Data With VAE as Prior

Inversion for field data is done similarly to the synthetic case. The smooth regularization inverted model and the VAE inverted model are shown in Figures 10a and 10b, respectively. The behavior of RMSE, norm of **z**, and velocity parameters during optimization is shown in Figures 10c–10e. The RMSE follows a behavior consistent with the chosen SGD scheme: an initial phase with very large oscillations followed by a more stable decreasing behavior. The behavior of the norm of **z** indicates that during the initial phase the search covers very large range of radial distances from the origin while for the end it is constrained to small radial changes. VAE inverted models with different initial starting models are shown in Figures 10f–10h. Again, to check if assembling the prior from many different TIs is advantageous, we compare it with the results of using the individual TIs. Boxplots of the data RMSE for 10 inversions (each with a different initial model) for all VAEs trained with each of the TIs and the assembled prior are shown in Figure 8b. From the inversions with individual TIs we choose the best, median and worst in terms of average data RMSE and together with the assembled VAE, compute the mean and standard deviation values of their final data RMSE, norm of **z**, and the velocity parameters from their corresponding 10 inverted models (Table 1). The models inverted for one starting model with the TIs corresponding to the best, median, and worst average data RMSE are shown in Figures 10i–10k.

Inversion results for the field data (Figure 10) show a behavior very similar to the synthetic case. However, the inverted model indicates a simpler structure when compared to the synthetic case. This is consistent with evidence from the excavation and even inclination trends of both the upper sand and lower sand bodies seem to match those observed in excavated profiles close to the GPR sensed domain (Larsen et al., 2016; Bording et al., 2019). Regarding the performance of the assembled prior for inversion, Table 1 shows that training the VAE with all the TIs at the same time performs better than the median individual TI and results in approximately equal values of average RMSE compared to inversion with the best individual TI. This indicates that it may be better to build an assembled wide prior than to consider many TIs individually for inversion (Hermans et al., 2015). Note that results of the best individual TI have only slightly lower values of RMSE. The assembled prior also has the advantage of a lower computational demand: one does not have to train a VAE and do the inversion for each individual TI. In the presented field case, for instance, the computational demand is 18 times higher if the TIs are considered individually. Moreover, prior uncertainty tends to be larger in field cases therefore a wider prior distribution, such as the one modeled by the VAE with all the TIs, is preferable. This wider prior distribution may indeed help in reducing bias arising when highly informative prior information is used.

3.4. Discussion

Solid Earth research strongly relies on our ability to observe it. Unfortunately, this is impeded by the sparse direct observations and the indirect measurements that help us inform it.

Starting from the description of spatial patterns of the subsurface (e.g., statistics of clay lenses embedded in a matrix) that are only sparsely visible, we combine them with geophysical data (e.g., traveltimes) which provide a greater coverage at the cost of being indirect information in a novel imaging framework which honors both the geophysical data and the prior knowledge of the subsurface. In addition, our method allows to enlarge the observed prior through morphological operators, which are geologically consistent and provide a continuous prior trained by the DGM. The implications of this work go beyond the scale of observation or the measurements that are used. This is supported by the similar conclusions that Jiang and Jafarpour (2021) obtained considering a larger scale and different kind of data (i.e., flow data versus traveltime data).

Compared with other methods used to integrate geological information and geophysical data, our approach aims to create a low-dimensional representation of a diverse set of patterns (represented e.g., by different TIs) that is (1) continuous and (2) easily handled by gradient-based inversion. Other methods fail in at least one of these two points, for example, MPS is not able to create a continuous low-dimensional representation neither is directly explored by means of gradient-based inversion and GANs can create a continuous low-dimensional representation but gradient-based inversion may not behave properly with such representation (Laloy et al., 2019). While not all inversion problems are based on a continuous prior distribution, an important subset actually do. This will be the case, for example, when the uncertainty regarding the geological scenario involves similar environments for which one may expect a continuous transition due to overlapping of spatial patterns.

Improving diversity performance of VAE while generating geologically realistic models is an important issue when comparing with other methods. Consider for example, the mechanism by which the VAE generates new samples of the patterns to equivalent mechanisms in MPS. While the departure of new patterns from training patterns in a VAE depends mainly in training parameters such as regularization weights α and β (see Appendix A) which in turn impact the approximation of the continuous prior in model space, MPS may control the diversity of patterns by relaxing the conditioning, for example, by changing the number of conditioning pixels or by defining distances to the conditioning event. Further study of these differences should enlighten under which circumstances it is better to use either of these strategies to produce more diverse patterns or even if it is possible to combine them to better represent prior uncertainty in the most realistic way possible (see e.g., Bai & Tahmasebi, 2020). It is worth mentioning that the problem of using multiple TIs with MPS seems to have received little attention (Scheidt et al., 2016; Silva & Deutsch, 2012) perhaps because most studies focus on discrete aspects (e.g., different depositional environments) rather than continuous aspects as in this study (i.e., deformation, erosion-dilation, and intrafacies variability). In several cases, however, one should be able to frame inversion problems for subsurface models in terms of continuous variables (e.g., two depositional environments may have transitional environments between them).

In this work we considered a normal multivariate Gaussian distribution to model the prior in latent space (i.e., as input to the generative function of the VAE), however, other types of distributions may also be used, for example, a Gaussian mixture model (Makhzani et al., 2015). A Gaussian mixture model may be used as latent prior distribution, for instance, when we have clearly separated modes, such as two training images that most likely do not have overlapping patterns. These other types of distributions may provide two main advantages: (1) they may provide a better approximation of the prior since they get rid of some pathologies of traditional VAEs (Yacoby et al., 2021), and (2) they are more directly related to the prior distribution in model space and therefore cause less nonlinearity and/or topological changes which is favorable for the performance of gradient-based inversion (Lopez-Alvis et al., 2021). However, sampling from these distributions in latent space is not as straightforward as for a multivariate Gaussian. This means that one would have to rely on either different regularization terms in latent space or more advanced (but potentially more computationally demanding) ways of sampling, such as Markov chain Monte Carlo.

4. Conclusions

When prior information is expressed by a set of TIs, a VAE may be used to approximate a prior distribution that effectively assembles all the possible spatial patterns. Variability of the spatial patterns included in the TIs may be further increased by using geologically consistent transformations such as erosion/dilation, local deformation and intrafacies variability which result in a set of patterns that represent similar geological environments. The VAE is capable of producing patterns that deviate from training patterns but remain realistic, therefore increasing pattern diversity. The cross-borehole GPR traveltime synthetic case demonstrates that inversion with SGD in the latent space of the VAE is able to obtain a realistic model while remaining computationally efficient. Even though the final misfit is higher than the noise level, most structural features are correctly inverted. By assuming a linear mixing model (two additional parameters), the absolute values of velocity may be also estimated in the inversion. This allows for inversion using a VAE as prior to be more readily applied to a field data set. Results from the field case validate VAE-based inversion since they show a realistic inverted model with misfit only slightly higher than the estimated noise and therefore provide one of the first successful applications of DGM-based inversion. A comparison of VAEs trained on individual TIs and the VAE trained with all the TIs at the same time shows that the latter performs as good as the best individual TIs. Moreover, it has the advantage of lower computational demand and a more adequate (wider) prior uncertainty, which in turn may reduce bias from highly informative prior information. Finally, future work may include extending the proposed method to handle more than two subsurface materials, testing new geologically consistent transformations, considering more general distributions in the latent space and using it in combination with MPS to improve the accuracy and diversity of patterns.

Appendix A: Variational Autoencoder

The starting point is to pose the VAE's training as maximizing the sum of the evidence (or marginal likelihood) lower bound of each individual sample $\mathbf{m}^{(i)}$. The evidence lower bound for each sample can be written as:

$$\mathcal{L}\left(\theta, \vartheta; \mathbf{m}^{(i)}\right) = \mathcal{L}^m + \mathcal{L}^z \tag{A1}$$

with

$$\mathcal{L}^m = \mathbb{E}_{q_\vartheta(\mathbf{z}|\mathbf{m}^{(i)})}\left[\log\left(p_\theta\left(\mathbf{m}^{(i)}|\mathbf{z}\right)\right)\right] \tag{A2}$$

and

$$\mathcal{L}^z = -D_{KL}\left(q_\vartheta\left(\mathbf{z}|\mathbf{m}^{(i)}\right)\|p(\mathbf{z})\right) \tag{A3}$$

where $p_\theta(\mathbf{m}|\mathbf{z})$ is the (probabilistic) decoder, $q_\vartheta(\mathbf{z}|\mathbf{m})$ is the (probabilistic) encoder, $\mathbb{E}$ denotes the expectation operator, D_{KL} denotes the Kullback-Leibler divergence and, θ and ϑ are the parameters (weights and biases) of the neural networks for the decoder and encoder, respectively.

In order to maximize the evidence lower bound in Equation A1, an estimator for $\mathcal{L}$ is used. This estimator is based on a so called reparameterization trick of the random variable $\tilde{\mathbf{z}} \sim q_\vartheta(\mathbf{z}|\mathbf{m})$ which uses an auxiliary noise ϵ. In the case of a VAE, the encoder is defined as a multivariate Gaussian with diagonal covariance:

$$q_\vartheta(\mathbf{z}|\mathbf{m}) = \mathcal{N}\left(\mathbf{h}_\vartheta(\mathbf{m}), \mathbf{u}_\vartheta(\mathbf{m}) \cdot I_n\right) \tag{A4}$$

where $\mathbf{h}_\vartheta(\mathbf{m})$ and $\log \mathbf{u}_\vartheta(\mathbf{m})$ are modeled with neural networks and I_n is a $n \times n$ diagonal matrix. Then, the encoder and the auxiliary noise ϵ are used in the following way during training:

$$\tilde{\mathbf{z}} = \mathbf{h}_\vartheta(\mathbf{m}) + \mathbf{u}_\vartheta(\mathbf{m}) \odot \epsilon, \quad \epsilon \sim \mathcal{N}\left(\mathbf{0}, \alpha \cdot I_n\right) \tag{A5}$$

where $\odot$ denotes an element-wise product and α defines the magnitude of the variance of ϵ. Often Equation A3 has an analytical solution, then only Equation A2 is approximated with the estimator as:

$$\tilde{\mathcal{L}}^m = \frac{1}{L} \sum_{j=1}^{L} \log\left(p_\theta\left(\mathbf{m}^{(i)}|\tilde{\mathbf{z}}^{(i,j)}\right)\right) \tag{A6}$$

where $\tilde{\mathbf{z}}^{(i,j)} = \mathbf{h}_\vartheta\left(\mathbf{m}^{(i)}\right) + \mathbf{u}_\vartheta\left(\mathbf{m}^{(i)}\right) \odot \epsilon^{(j)}$ and L is the number of samples used for the estimator. Further, if we set the decoder $p_\theta(\mathbf{m}|\mathbf{z})$ as a multivariate Gaussian with diagonal covariance structure, then

$$p_\theta(\mathbf{m}|\mathbf{z}) = \mathcal{N}\left(\mathbf{g}_\theta(\mathbf{z}), \mathbf{v}_\theta(\mathbf{z}) \cdot I_N\right) \tag{A7}$$

where $\mathbf{g}_\theta(\mathbf{z})$ and $\log \mathbf{v}_\theta(\mathbf{z})$ are modeled with neural networks and I_N is a $N \times N$ diagonal matrix. In this work, we consider only the mean of the decoder $p_\theta(\mathbf{m}|\mathbf{z})$ which is just the (deterministic) generator $\mathbf{g}_\theta(\mathbf{z})$. Then, the corresponding (mean-squared error) loss function may be written as

$$\tilde{\mathcal{L}}^m = \frac{1}{L} \sum_{j=1}^{L} \|\mathbf{g}_\theta\left(\tilde{\mathbf{z}}^{(i,j)}\right) - \mathbf{m}^{(i)}\|^2 \tag{A8}$$

The described setting allows for the gradient to be computed with respect to both θ and ϑ and then stochastic gradient descent is used to maximize the lower bound in Equation A1.

As previously mentioned, it is often possible to analytically integrate the Kullback-Leibler divergence in Equation A3. In this work, we consider that $p(\mathbf{z})$ and $q_\vartheta(\mathbf{z}|\mathbf{m})$ are both Gaussian therefore Equation A3 may be rewritten as (Kingma & Welling, 2014):

$$\mathcal{L}^z = \frac{1}{2} \sum_{i=1}^{n} \left(1 + \log\left((u_i)^2\right) - (h_i)^2 - (u_i)^2\right) \tag{A9}$$

where the sum is done for the n output dimensions of the encoder.

Note that the term in Equations A2, A6, and A8 may be interpreted as a reconstruction term that causes the outputs of the encode-decode operation to look similar to the training samples, while the term in Equations A3 and A9 may be considered a regularization term that enforces the encoder $q_\vartheta(\mathbf{z}|\mathbf{m})$ to be close to a prescribed distribution $p(\mathbf{z})$. In practice, one may add a weight to the second term (Higgins et al., 2017) of the lower bound as:

$$\tilde{\mathcal{L}}\left(\theta, \vartheta; \mathbf{m}^{(i)}\right) = \tilde{\mathcal{L}}^m + \beta \mathcal{L}^z \tag{A10}$$

to prevent samples to be encoded far from each other in the latent space, which may cause overfitting of the reconstruction term and degrade the VAE's generative performance. The overall process of training and generation for a VAE is depicted in Figure A1.
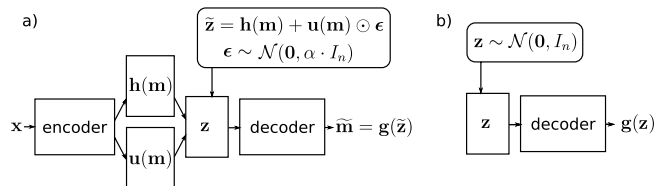

Figure A1. A diagram for a VAE: (a) steps needed for training and (b) steps needed for generation.

Data Availability Statement

The cross-borehole GPR data set data used for the field case in this work is available at Pangaea via https://doi.org/10.1594/PANGAEA.934056 with a Creative Commons Attribution 4.0 International (CC-BY-4.0) license. Version 1.0 of the VAE_SGD_field code used for VAE training and VAE-based inversion is preserved at https://doi.org/10.5281/zenodo.4915277 and available via MIT License.

References

Backus, G. E., & Gilbert, J. F. (1967). Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal of the Royal Astronomical Society*, *13*(1–3), 247–276. https://doi.org/10.1111/j.1365-246X.1967.tb02159.x

Bai, T., & Tahmasebi, P. (2020). Hybrid geological modeling: Combining machine learning and multiple-point statistics. *Computers & Geosciences*, *142*, 104519. https://doi.org/10.1016/j.cageo.2020.104519

Bergmann, U., Jetchev, N., & Vollgraf, R. (2017). *Learning Texture manifolds with the Periodic spatial GAN*. ArXiv:1705.06566 [cs, stat]. Retrieved from http://arxiv.org/abs/1705.06566

Bora, A., Jalal, A., Price, E., & Dimakis, A. G. (2017). *Compressed sensing using generative models*. ArXiv:1703.03208 [cs, math, stat]. Retrieved from http://arxiv.org/abs/1703.03208

Bording, T. S., Fiandaca, G., Maurya, P. K., Auken, E., Christiansen, A. V., Tuxen, N., & Larsen, T. H. (2019). Cross-borehole tomography with full-decay spectral time-domain induced polarization for mapping of potential contaminant flow-paths. *Journal of Contaminant Hydrology*, *226*, 103523. https://doi.org/10.1016/j.jconhyd.2019.103523

Caers, J., & Hoffman, T. (2006). The probability perturbation method: A new look at Bayesian inverse modeling. *Mathematical Geology*, *38*(1), 81–100. https://doi.org/10.1007/s11004-005-9005-9

Canchumuni, S. W., Emerick, A. A., & Pacheco, M. A. C. (2019). Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers & Geosciences*, *128*, 87–102. https://doi.org/10.1016/j.cageo.2019.04.006

Chan, S., & Elsheikh, A. H. (2019). *Parametrization and generation of geological models with generative adversarial networks*. ArXiv:1708.01810 [physics, stat]. Retrieved from http://arxiv.org/abs/1708.01810

Davis, K., & Li, Y. (2011). Fast solution of geophysical inversion using adaptive mesh, space-filling curves and wavelet compression. *Geophysical Journal International*, *185*(1), 157–166. https://doi.org/10.1111/j.1365-246X.2011.04929.x

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*(5), 889–904. https://doi.org/10.1162/neco.1995.7.5.889

Day-Lewis, F. D., Singha, K., & Binley, A. M. (2005). Applying petrophysical models to radar travel time and electrical resistivity tomograms: Resolution-dependent limitations. *Journal of Geophysical Research*, *110*, B08206. https://doi.org/10.1029/2004JB003569

Fefferman, C., Mitter, S., & Narayanan, H. (2016). Testing the manifold hypothesis. *Journal of the American Mathematical Society*, *29*(4), 983–1049. https://doi.org/10.1090/jams/852

Franklin, J. N. (1970). Well-posed stochastic extensions of ill-posed linear problems. *Journal of Mathematical Analysis and Applications*, *31*(3), 682–716.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/BF00344251

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. ArXiv:1406.2661 [cs, stat]. Retrieved from http://arxiv.org/abs/1406.2661

Hansen, T. M., Cordua, K. S., Jacobsen, B. H., & Mosegaard, K. (2014). Accounting for imperfect forward modeling in geophysical inverse problems—Exemplified for crosshole tomography. *Geophysics*, *79*(3), H1–H21. https://doi.org/10.1190/geo2013-0215.1

Hansen, T. M., Cordua, K. S., & Mosegaard, K. (2012). Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. *Computational Geosciences*, *16*(3), 593–611. https://doi.org/10.1007/s10596-011-9271-1

Hermans, T., Nguyen, F., & Caers, J. (2015). Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: Workflow and case study. *Water Resources Research*, *51*, 5332–5352. https://doi.org/10.1002/2014WR016460

Hermans, T., Vandenbohede, A., Lebbe, L., Martin, R., Kemna, A., Beaujean, J., & Nguyen, F. (2012). Imaging artificial salt water infiltration using electrical resistivity tomography constrained by geostatistical data. *Journal of Hydrology*, *438–439*, 168–180. https://doi.org/10.1016/j.jhydrol.2012.03.021

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., et al. (2017). *Beta-VAE: Learning basic visual concepts with a constrained variational framework*.

Holm-Jensen, T., & Hansen, T. M. (2019). Linear Waveform tomography inversion using machine learning algorithms. *Mathematical Geosciences*, *52*, 31–51. https://doi.org/10.1007/s11004-019-09815-7

Hu, L. Y., Blanc, G., & Noetinger, B. (2001). Gradual deformation and iterative calibration of sequential stochastic simulations. *Mathematical Geology*, *33*(4), 475–489.

Jafarpour, B. (2011). Wavelet reconstruction of geologic facies from nonlinear dynamic flow measurements. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(5), 1520–1535. https://doi.org/10.1109/TGRS.2010.2089464

Jafarpour, B., Goyal, V. K., McLaughlin, D. B., & Freeman, W. T. (2009). Transform-domain sparsity regularization for inverse problems in geosciences. *Geophysics*, *74*(5), R69–R83. https://doi.org/10.1190/1.3157250

Jiang, A., & Jafarpour, B. (2021). Deep convolutional autoencoders for robust flow model calibration under uncertainty in geologic continuity. *Water Resources Research*, *57*, e2021WR029754. https://doi.org/10.1029/2021WR029754

Journel, A., & Zhang, T. (2007). The necessity of a multiple-point prior model. *Mathematical Geology*, *38*(5), 591–610. https://doi.org/10.1007/s11004-006-9031-2

Kessler, T. C., Comunian, A., Oriani, F., Renard, P., Nilsson, B., Klint, K. E., et al. (2013). Modeling Fine-scale geological heterogeneity-examples of sand lenses in tills. *Ground Water*, *51*(5), 692–705. https://doi.org/10.1111/j.1745-6584.2012.01015.x

Kessler, T. C., Klint, K. E. S., Nilsson, B., & Bjerg, P. L. (2012). Characterization of sand lenses embedded in tills. *Quaternary Science Reviews*, *53*, 55–71. https://doi.org/10.1016/j.quascirev.2012.08.011

Khaninezhad, M. M., Jafarpour, B., & Li, L. (2012). Sparse geologic dictionaries for subsurface flow model calibration: Part I. Inversion formulation. *Advances in Water Resources*, *39*, 106–121. https://doi.org/10.1016/j.advwatres.2011.09.002

Kingma, D. P., & Welling, M. (2014). *Auto-encoding variational Bayes*. ArXiv:1312.6114 [cs, stat]. Retrieved from http://arxiv.org/abs/1312.6114

Kleinberg, R., Li, Y., & Yuan, Y. (2018). *An alternative view: When does SGD Escape local minima?* arXiv:1802.06175 [cs]. Retrieved from https://arxiv.org/abs/1802.061752

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, *37*(2), 233–243. https://doi.org/10.1002/aic.690370209

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Laloy, E., Hérault, R., Jacques, D., & Linde, N. (2018). Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, *54*, 381–406. https://doi.org/10.1002/2017WR022148

Laloy, E., Hérault, R., Lee, J., Jacques, D., & Linde, N. (2017). Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Advances in Water Resources*, *110*, 387–405. https://doi.org/10.1016/j.advwatres.2017.09.029

Laloy, E., Linde, N., Ruffino, C., Hérault, R., Gasso, G., & Jacques, D. (2019). Gradient-based deterministic inversion of geophysical data with generative adversarial networks: Is it feasible? *Computers & Geosciences*, *133*, 104333. https://doi.org/10.1016/j.cageo.2019.104333

Larsen, T. H., Palstrøm, P., & Larsen, L. (2016). *Kallerup Grusgrav: Kortlægning af sandlinser i moræneler* (Tech. Rep. 3641400021). Orbicon A/S.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Lemmens, L., Rogiers, B., Jacques, D., Huysmans, M., Swennen, R., Urai, J. L., et al. (2019). Nested multiresolution hierarchical simulated annealing algorithm for porous media reconstruction. *Physical Review E*, *100*(5), 053316. https://doi.org/10.1103/PhysRevE.100.053316

Linde, N., Renard, P., Mukerji, T., & Caers, J. (2015). Geological realism in hydrogeological and geophysical inverse modeling: A review. *Advances in Water Resources*, *86*, 86–101. https://doi.org/10.1016/j.advwatres.2015.09.019

Looms, M. C., Klotzsche, A., van der Kruk, J., Larsen, T. H., Edsen, A., Tuxen, N., et al. (2018). Mapping sand layers in clayey till using crosshole ground-penetrating radar. *Geophysics*, *83*(1), A21–A26. https://doi.org/10.1190/geo2017-0297.1

Looms, M. C., Klotzsche, A., van der Kruk, J., Larsen, T. H., Edsen, A., Tuxen, N., et al. (2021). Crosshole ground penetrating radar data collected in clayey till with sand occurrences. *Pangaea*. https://doi.org/10.1594/PANGAEA.934056

Lopez-Alvis, J., Hermans, T., & Nguyen, F. (2019). A cross-validation framework to extract data features for reducing structural uncertainty in subsurface heterogeneity. *Advances in Water Resources*, *133*, 103427. https://doi.org/10.1016/j.advwatres.2019.103427

Lopez-Alvis, J., Laloy, E., Nguyen, F., & Hermans, T. (2021). Deep generative models in inversion: The impact of the generator's nonlinearity and development of a new approach based on a variational autoencoder. *Computers & Geosciences*, *152*, 104762. https://doi.org/10.1016/j.cageo.2021.104762

Lucic, M., Kurach, K., Michalski, M., Gelly, S., & Bousquet, O. (2018). *Are GANs created equal? A large-scale study*. arXiv:1711.10337 [cs, stat]. Retrieved from http://arxiv.org/abs/1711.10337

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., & Frey, B. (2015). *Adversarial autoencoders*. ArXiv:1511.05644 [cs]. Retrieved from http://arxiv.org/abs/1511.05644

Mariethoz, G. (2018). When should we use multiple-point geostatistics? In B. Daya Sagar, Q. Cheng, & F. Agterberg (Eds.), *Handbook of mathematical geosciences: Fifty years of IAMG* (pp. 645–653). Springer International Publishing. https://doi.org/10.1007/978-3-319-78999-6_31

Mariethoz, G., Renard, P., & Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations: Performing multiple-points simulations. *Water Resources Research*, *46*, W11536. https://doi.org/10.1029/2008WR007621

Maurer, H., Holliger, K., & Boerner, D. (1998). Stochastic regularization: Smoothness or similarity. *Geophysical Research Letters*, *25*(15), 2889–2892.

Mosser, L., Dubrule, O., & Blunt, M. J. (2018). *Stochastic seismic waveform inversion using generative adversarial networks as a geological prior*. arXiv:1806.03720 [physics, stat]. Retrieved from http://arxiv.org/abs/1806.03720

Oware, E. K., Irving, J., & Hermans, T. (2019). Basis-constrained Bayesian Markov-chain Monte Carlo difference inversion for geoelectrical monitoring of hydrogeologic processes. *Geophysics*, *84*(4), A37–A42. https://doi.org/10.1190/geo2018-0643.1

Park, H., Scheidt, C., Fenwick, D., Boucher, A., & Caers, J. (2013). History matching and uncertainty quantification of facies models with multiple geological interpretations. *Computational Geosciences*, *17*(4), 609–621. https://doi.org/10.1007/s10596-013-9343-5

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., & Lerer, A. (2017). *Automatic differentiation in PyTorch*.

Radford, A., Metz, L., & Chintala, S. (2016). *Unsupervised representation learning with deep convolutional generative adversarial networks*. arXiv:1511.06434 [cs]. Retrieved from http://arxiv.org/abs/1511.06434

Renard, P., & Allard, D. (2013). Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*, *51*, 168–196. https://doi.org/10.1016/j.advwatres.2011.12.001

Rezende, D. J., Mohamed, S., & Wierstra, D. (2014). *Stochastic backpropagation and approximate inference in deep generative models*. arXiv:1401.4082 [cs, stat]. Retrieved from http://arxiv.org/abs/1401.4082

Sajjadi, M. S. M., Bachem, O., Lucic, M., Bousquet, O., & Gelly, S. (2018). *Assessing generative models via Precision and Recall*. arXiv:1806.00035 [cs, stat]. Retrieved from http://arxiv.org/abs/1806.00035

Sarma, P., Durlofsky, L. J., & Aziz, K. (2008). Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Mathematical Geosciences*, *40*(1), 3–32. https://doi.org/10.1007/s11004-007-9131-7

Scheidt, C., Fernandes, A. M., Paola, C., & Caers, J. (2016). Quantifying natural delta variability using a multiple-point geostatistics prior uncertainty model: Delta variability and geostatistics. *Journal of Geophysical Research: Earth Surface*, *121*, 1800–1818. https://doi.org/10.1002/2016JF003922

Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying uncertainty in subsurface systems* (No. 236). John Wiley and Sons & American Geophysical Union.

Seo, J. K., Kim, K. C., Jargal, A., Lee, K., & Harrach, B. (2019). A learning-based method for solving ill-posed nonlinear inverse problems: A simulation study of lung EIT. *SIAM Journal on Imaging Sciences*, *12*(3), 1275–1295. https://doi.org/10.1137/18M1222600

Silva, D. A., & Deutsch, C. V. (2012). Multiple point statistics with multiple training images. *CCG Annual Report*, *14*, 8.

Soille, P. (2004). *Morphological image analysis*. Springer. https://doi.org/10.1007/978-3-662-05088-0

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., & Besson, O. (2011). An improved parallel multiple-point Algorithm using a list approach. *Mathematical Geosciences*, *43*(3), 305–328. https://doi.org/10.1007/s11004-011-9328-7

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, *34*(1), 1–21. https://doi.org/10.1023/A:1014009426274

Tarantola, A., & Valette, B. (1982). Inverse problems = quest for information. *Journal of Geophysics*, *50*(3), 159–170.

Tikhonov, A. N., & Arsenin, V. I. A. (1977). *Solutions of ill-posed problems*. Winston and Sons.

Torquato, S., Beasley, J. D., & Chiew, Y. C. (1988). Two-point cluster function for continuum percolation. *The Journal of Chemical Physics*, *88*(10), 6540–6547. https://doi.org/10.1063/1.454440

Treister, E., & Haber, E. (2016). A fast marching algorithm for the factored eikonal equation. *Journal of Computational Physics*, *324*, 210–225. https://doi.org/10.1016/j.jcp.2016.08.012

Uria, B., Murray, I., & Larochelle, H. (2014). *A deep and tractable density estimator*. arXiv:1310.1757 [cs, stat]. Retrieved from http://arxiv.org/abs/1310.1757

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., et al. (2014). scikit-image: Image processing in Python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

Vo, H. X., & Durlofsky, L. J. (2014). A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Mathematical Geosciences*, *46*(7), 775–813. https://doi.org/10.1007/s11004-014-9541-2

Yacoby, Y., Pan, W., & Doshi-Velez, F. (2021). *Failure modes of variational autoencoders and their effects on downstream tasks*. (_eprint: 2007.07124).

You, N., Li, Y. E., & Cheng, A. (2021). 3D Carbonate digital rock reconstruction using progressive growing GAN. *Journal of Geophysical Research: Solid Earth*, *126*, e2021JB021687. https://doi.org/10.1029/2021JB021687

Zhang, C., Butepage, J., Kjellstrom, H., & Mandt, S. (2018). *Advances in variational inference*. ArXiv:1711.05597 [cs, stat]. Retrieved from http://arxiv.org/abs/1711.05597