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

Jorge Lopez-Alvis[1], Frederic Nguyen[2], Thomas Hermans[3], and Majken C. Looms[4]

[1]University of Liege
[2]University of Liège
[3]Ghent University
[4]University of Copenhagen

November 23, 2022

Abstract

Prior information regarding subsurface 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, e.g. 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. 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. Estimation of absolute values of wave velocity is also possible by assuming a linear mixing model and including two additional parameters in the inversion.

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

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

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

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.

Corresponding author: Jorge Lopez-Alvis, `jlopez@uliege.be`

Abstract

Prior information regarding subsurface 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, e.g. 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. 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. Estimation of absolute values of wave velocity is also possible by assuming a linear mixing model and including two additional parameters 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.

Keywords: prior information, geophysical inversion, variational autoencoder, deep learning, ground-penetrating radar, traveltime tomography

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) but this inherently biases the solution towards an a priori constraint which may not be realistic and therefore may hinder the use of the model for certain applications. If information regarding 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, e.g. outcrops which are representative of the local geology (Linde et al., 2015). To integrate this information with measured data, the patterns must be described by techniques that account for their spatial nature. 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; Maurer et al., 1998; Hermans et al., 2012). The choice of geostatistical 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 (Strebelle, 2002; Journel & Zhang, 2007). Recently, deep generative models (DGMs) have been proposed as an alternative to MPS to reproduce such complex spatial patterns (Laloy et al., 2017; Chan & Elsheikh, 2019; 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 dataset (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 (Strebelle, 2002; Straubhaar et al., 2011) 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, i.e. 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; Mosser et al., 2018; Lopez-Alvis et al., 2021). This is typically not available for inversion with MPS, for which other ways of exploring the model space have been used (Hu et al., 2001; Caers & Hoffman, 2006; Hansen et al., 2012; 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 $n << N$). 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 (Mosser et al., 2018; Laloy et al., 2019; Canchumuni et al., 2019). It was recently argued that with certain choice of parameters VAEs may control both the degree of nonlinearity 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-borehole ground penetrating radar (GPR) traveltime data. The novelty of this work lies first in the training of the VAE using perturbed images based on transformations which result in a set of patterns that represent similar geological environments. Second, to the best of our knowledge, the current work constitutes one of the first efforts to successfully validate DGM-based inversion with a field dataset. Finally, the methodology is not restricted to GPR dataset but could easily be extended to seismic traveltimes or more generally to other geophysical inverse problems. Also, in contrast to previous studies (Laloy et al., 2017, 2018; Mosser et al., 2018; Lopez-Alvis et al., 2021) the values of the geophysical parameter (wave velocity) are assumed unknown and included in inversion by means of a mixing model.

The remaining 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. In this section, the relation of the proposed framework with previous studies is also highlighted and suggestions for future work are given. 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, i.e. 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.

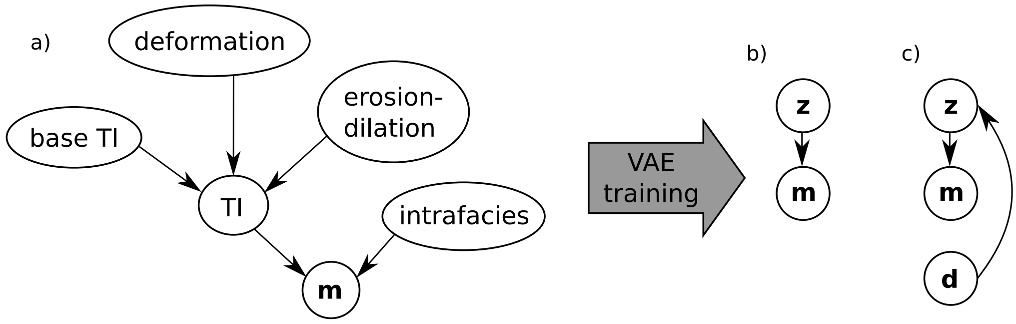

Figure 1. Probabilistic graphical models for: (a) generation of model samples using the original variables, (b) generation of new model samples from the VAE using the latent variables and (c) VAE-based inversion. $\mathbf{m}$, $\mathbf{d}$ and $\mathbf{z}$ refer to the model, data and latent vectors, respectively.

2.1 Variational autoencoder: approximating a complex probability distribution

A variational autoencoder (VAE) may be classified as a deep generative model (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, e.g. 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, e.g. 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 dataset $\mathbf{M} = \{\mathbf{m}^{(i)}\}_{i=1}^P$ 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 dataset $\mathbf{M}$ by sampling from $p(\mathbf{z})$ and passing through the generative mapping, i.e. sampling according to $p(\mathbf{z})p_\theta(\mathbf{m}|\mathbf{z})$. However, when the training samples $\mathbf{m}^{(i)}$ are high-dimensional, non-standard 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 & 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, i.e. they are fixed for the subsequent inversion.

Note that the training dataset $\mathbf{M}$ may contain different kinds of patterns which allow the VAE to effectively learn what is here termed an assembled prior, i.e. 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, i.e. 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). We found e.g. that $n = 60$ provides only a slight improvement in reconstruction of the training samples but causes a significant degradation of generated patterns.

2.2 Inversion of traveltime data using a VAE as prior

As mentioned above, a VAE using CNNs provides a powerful tool to represent complex probability distributions. Therefore if one has a large dataset containing examples of spatial patterns, the VAE allows to approximate complex prior probability distributions in the context of geophysical inversion. Following the derivation in Appendix B, inversion is done by minimizing the objective function in equation B5 and whose gradient is computed according to equations B6 and B8 as:

$$\nabla_{\mathbf{z}}\zeta(\mathbf{z}) = -\mathbf{S}(\mathbf{z})^T\left(\,\mathbf{J}(\mathbf{m})^T(\mathbf{d} - \mathbf{f}(\mathbf{m}))\,\right) + 2\lambda\mathbf{z} \tag{1}$$

where $\mathbf{S}(\mathbf{z})$ is the Jacobian matrix of the generative mapping, $\mathbf{f}$ is the geophysical forward operator, $\mathbf{J}(\mathbf{m})$ is the Jacobian matrix of the forward operator and λ is a regularization factor in the latent space.

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

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

is used, where τ denotes the traveltime and v is the velocity of the subsurface materials. Note that equation 2 is not limited to GPR but may also be applied to e.g. seismic traveltime. A numerical solution is typically required, where after discretization one obtains the forward operator that relates the vector of traveltimes $\mathbf{d} = \boldsymbol{\tau}$ to the slowness (which is the reciprocal of velocity) vector $\mathbf{m} = \mathbf{v}^{-1}$ in equation B1. 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 non-negligible 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 (Laloy et al., 2017, 2018; Mosser et al., 2018; Canchumuni et al., 2019), 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:

$$\frac{\partial\zeta(\mathbf{w})}{\partial w_i} = \nabla_{\mathbf{v}}\gamma(\mathbf{v})\frac{\partial\mathbf{v}}{\partial w_i} \tag{3}$$

where $\nabla_{\mathbf{v}}\gamma(\mathbf{v})$ is given by equation B8 but computed using the values of $\mathbf{v}$ instead of $\mathbf{m}$. For the two w_i parameters we have:

$$\frac{\partial \mathbf{v}}{\partial w_1} = \mathbf{1},\ \frac{\partial \mathbf{v}}{\partial w_2} = \mathbf{m} \tag{4}$$

Similarly, the first term on the right of equation B6 should now be computed using $\mathbf{v}$ instead of $\mathbf{m}$. Strictly, this term should also include a derivative with respect to $\mathbf{m}$, however, this is a constant and it has no impact since the step size of the optimization is scaled in every iteration. 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, stochastic gradient descent (SGD) and equation 1 are used for optimization of the objective function (Lopez-Alvis et al., 2021). 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).

2.3 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 $65 \times 129 = 8385$ cells, corresponding to a 3.25 m $\times$ 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 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 2c). BTI_2 is built similarly but includes the sand sheets (Figure 2d) 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 4762×4762 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 which is 6×2 pixels. 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- 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 the nine modified TIs obtained from BTI_2 are shown in Figure 2e. The size of each of these modified TIs is a bit smaller (4722×4722) than for the base TIs since cropping was needed in the edges after deformation.

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.

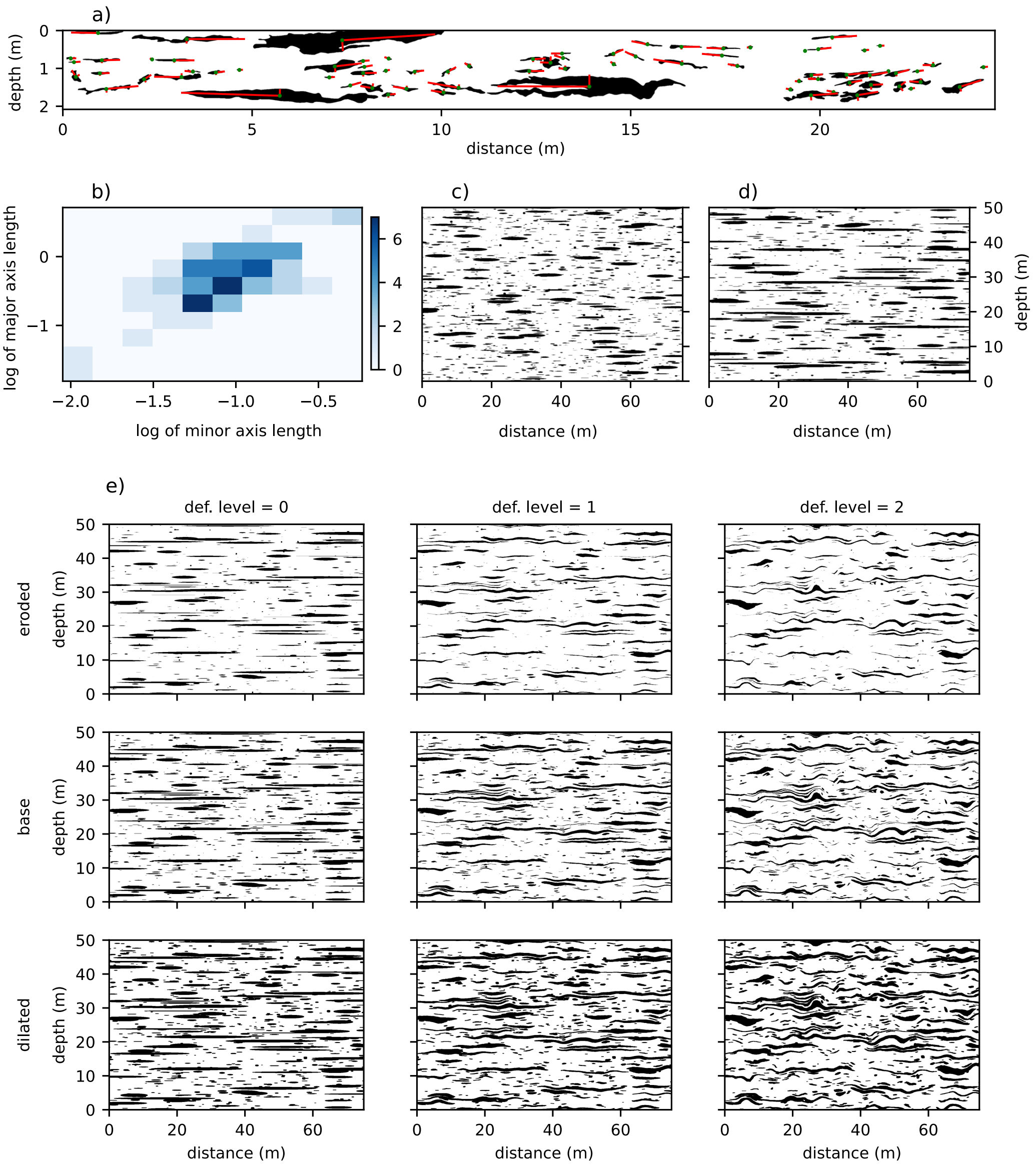

Figure 2. (a) Digitized outcrop from Kessler et al. (2013) showing sand bodies in black, background till in white, the axes of fitted ellipses for the sand bodies in red and centers of the ellipses in green. (b) Two-dimensional histogram of the major and minor axes lengths of the ellipses fitted in the outcrop. 1500 $\times$ 1000 pixel croppings of: (c) base image BTI_1, (d) base image BTI_2 and (e) the nine modified TIs corresponding to BTI_2.

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 (Larsen et al., 2016; Bording et al., 2019).

The field dataset is the cross-borehole traveltime data presented by Looms et al. (2018). 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 multi-offset 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 m 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 degrees were not considered to avoid error from borehole refraction. Estimated measurement error is 0.47 ns while average traveltime is 41.5 ns.

To assess the performance of our proposed inversion, a synthetic case is first analyzed with the same acquisition settings than 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 2 in Figure 2e). The model was cropped from a TI of the same size as the ones used for training but its random spatial realization was different, i.e. 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 was 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

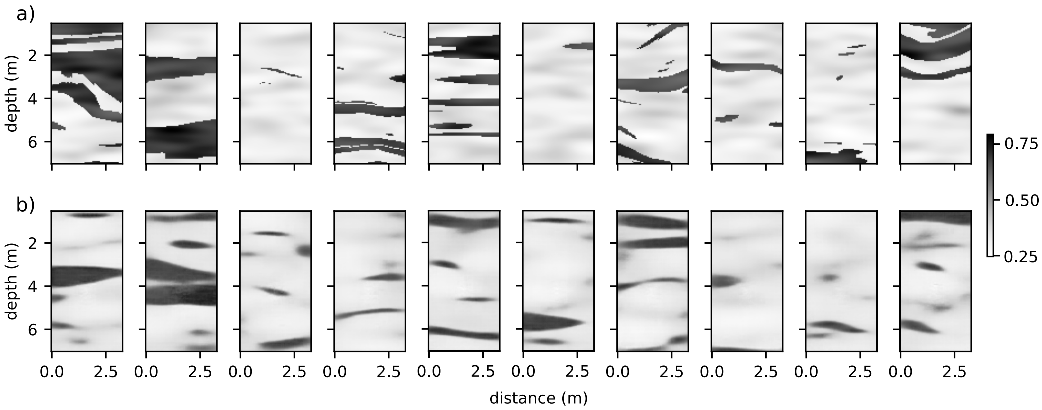

Figure 3. Examples of training samples (a) and samples generated from the trained VAE (b). The grey scale is with respect to the model variable $\mathbf{m}$ prior to its transformation to velocity values.

size of the spatial domain) and adding the intrafacies variability. Examples of the cropped samples are shown in Figure 3a. 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 hrs on a Nvidia GPU RTX 2060 (~ 3 hrs 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). 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 3b, these are samples from the assembled prior distribution approximated by the VAE. Also, a VAE is trained for each individual TI to make a comparison with the assembled prior.

The VAE-based generated patterns may fail to adequatly 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 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 (Park et al., 2013; Hermans et al., 2015; Scheidt et al., 2018; Lopez-Alvis et al., 2019) according to which none of the TIs is falsified, i.e. all the proposed patterns are likely to have generated the data. The details are shown in Supplementary material S1. The effect arising from the second reason is directly related to generative accuracy and is captured e.g. in Figure 3 where the generated 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. This may be checked by visualizing a set of training images as in Figure 3 and also making a latent traversal as shown in Figure 4, 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. 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 the breaking channels) would be adequately captured by these measures. 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 $\mathbf{z}$ of the VAE have effectively substituted the parameters related to the original hierarchical model (the substitution is denoted by the grey arrow in Figure 1). The latent parameter distribution now includes all 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 gradient (as detailed in section 2.2) 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. Figure 5a,b,c shows the real synthetic model, an inverted model with traditional smooth regularization and a VAE inverted model (for one randomly chosen starting model), respectively. The smooth inversion is done with a low regularization factor (10^{-9}), so it mainly represents the information content of the data and therefore is prone to artefacts due to noise (e.g. ray artefacts in Figure 5b). In contrast, due to the use of prior information that is geologically coherent, the VAE inverted model is artefact-free. For the model in Figure 5c, the behavior of the data misfit (RMSE), the Euclidian distance between the current model and the real model, the norm of $\mathbf{z}$ and the velocity parameters as the inversion progresses are shown in

Figure 4. Examples of VAE latent traversals (stepping in two latent dimensions while keeping the rest xed) 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.

Table 1. Mean and standard deviation values of inversions using 10 different initial models

TI*	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

*The labels indicate best, median and worst in terms of data RMSE from all 18 TIs.

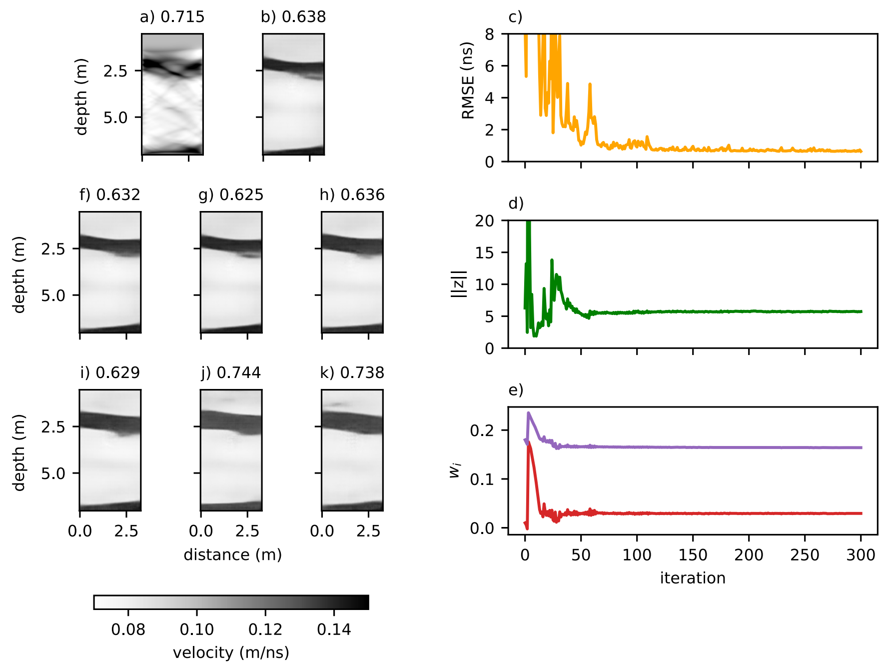

Figure 7. Inversion results for the field case: (a) smooth inverted model, (b) VAE-SGD inverted model for one random starting model using the assembled prior. For the model in (b), the values in each iteration for: data RMSE (c), norm of $\mathbf{z}$ (d) and linear mixing parameters (e). VAE-SGD inverted modes for three different starting models using the assembled prior (f,g,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 6b). For all inverted models, data RMSE is shown at the top.

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 slightly lower values of RMSE. This may be partially explained by the fact that a constant dimensionality $n = 40$ for the latent vector is used. A better strategy might be to slightly increase n when more diversity in the patterns is considered. 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.

It is interesting to contrast 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, e.g. by changing the number of conditioning pixels or by defining distances to the conditioning event. Further study of this relation 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 (Silva & Deutsch, 2012; Scheidt et al., 2016) 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 some 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), so further study of this subject may prove beneficial.

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, e.g. a Gaussian mixture model (Makhzani et al., 2015). These other types of distributions may provide two main advantages: (1) they may produce more accurate patterns, and (2) they are more directly related to the prior distribution in model space and therefore cause less nonlinearity and/or topological changes. 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.

4 Conclusions

When prior information is expressed by a set of TIs and their perturbed versions, a VAE may be used to approximate a prior distribution that effectively assembles all the possible spatial patterns. The perturbations are key to reproduce the expected heterogeneity and may include geologically consistent operations 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 dataset. 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:

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

with

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

and

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

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 distance 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 $\widetilde{\mathbf{z}} \sim q_\vartheta(\mathbf{z}|\mathbf{m})$ which uses an auxiliary noise $\boldsymbol{\epsilon}$. In the case of a VAE, the encoder is defined as a multivariate Gaussian with diagonal covariance:

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

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 $\boldsymbol{\epsilon}$ are used in the following way during training:

$$\widetilde{\mathbf{z}} = \mathbf{h}_\vartheta(\mathbf{m}) + \mathbf{u}_\vartheta(\mathbf{m}) \odot \boldsymbol{\epsilon}, \;\; \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \alpha \cdot I_n) \tag{A5}$$

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

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

where $\widetilde{\mathbf{z}}^{(i,j)} = \mathbf{h}_\vartheta(\mathbf{m}^{(i)}) + \mathbf{u}_\vartheta(\mathbf{m}^{(i)}) \odot \boldsymbol{\epsilon}^{(j)}$ 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

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

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

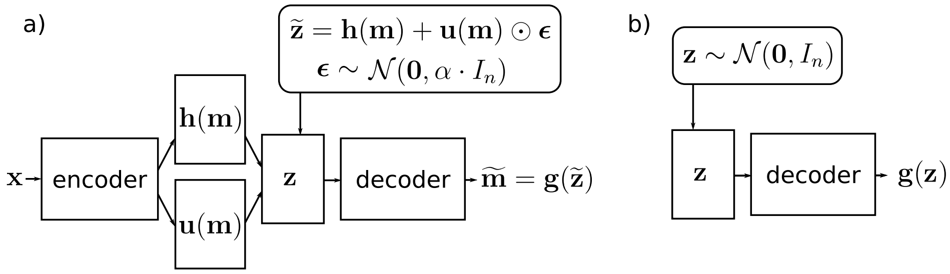

Figure A1. A diagram for a VAE: (a) steps needed for training and (b) steps needed for generation.

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

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 distance 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):

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

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:

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

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.

Appendix B Objective function for inversion with VAE

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:

$$\mathbf{d} = \mathbf{f}(\mathbf{m}) \tag{B1}$$

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 B1 may be represented with a conditional probability distribution $p(\mathbf{d}|\mathbf{m})$. Then, inversion is stated as:

$$p(\mathbf{m}|\mathbf{d}) = k\, p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m}) \tag{B2}$$

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

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

$$\begin{aligned} p(\mathbf{m}, \mathbf{z}|\mathbf{d}) &= k\, p(\mathbf{d}|\mathbf{m})\, p(\mathbf{z})\, p(\mathbf{m}|\mathbf{z}) \\ p(\mathbf{z}|\mathbf{d}) &= k\, p(\mathbf{z}) \int p(\mathbf{d}|\mathbf{m})\, p(\mathbf{m}|\mathbf{z})\, d\mathbf{m} \end{aligned} \tag{B3}$$

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 only considering the mean of the decoder then $p(\mathbf{m}|\mathbf{z}) = \delta(\mathbf{m} - \mathbf{g}(\mathbf{z}))$ and equation B3 may be written as:

$$\begin{aligned} p(\mathbf{z}|\mathbf{d}) &= k\, p(\mathbf{z}) \int p(\mathbf{d}|\mathbf{m})\, \delta(\mathbf{m} - \mathbf{g}(\mathbf{z}))\, d\mathbf{m} \\ &= k\, p(\mathbf{z})\, p(\, \mathbf{d}|\mathbf{g}(\mathbf{z})\,) \end{aligned} \tag{B4}$$

Equation B4 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 B4 and get the posterior distribution of the latent variables (Laloy et al., 2017, 2018). When appropriate values to train the VAE are used (Lopez-Alvis et al., 2021), $\mathbf{g}$ is expected to be only mildly nonlinear. 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:

$$\zeta(\mathbf{z}) = \|\mathbf{f}(\mathbf{g}(\mathbf{z})) - \mathbf{d}\|^2 + \lambda\|\mathbf{z}\|^2 \tag{B5}$$

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:

$$\begin{aligned}\nabla_{\mathbf{z}}\zeta(\mathbf{z}) &= \nabla_{\mathbf{z}}\|\mathbf{f}(\mathbf{g}(\mathbf{z})) - \mathbf{d}\|^2 + \lambda\nabla_{\mathbf{z}}\|\mathbf{z}\|^2 \\ &= \mathbf{S}(\mathbf{z})^T\nabla_{\mathbf{m}}\|\mathbf{f}(\mathbf{m}) - \mathbf{d}\|^2 + 2\lambda\mathbf{z}\end{aligned} \tag{B6}$$

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:

$$[\mathbf{S}(\mathbf{z})]_{i,j} = \frac{\partial g_i(\mathbf{z})}{\partial z_j} \tag{B7}$$

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

$$\nabla_{\mathbf{m}}\|\mathbf{f}(\mathbf{m}) - \mathbf{d}\|^2 = -\mathbf{J}(\mathbf{m})^T(\mathbf{d} - \mathbf{f}(\mathbf{m})) \tag{B8}$$

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

$$[\mathbf{J}(\mathbf{m})]_{i,j} = \frac{\partial f_i(\mathbf{m})}{\partial m_j} \tag{B9}$$

Acknowledgments

This work has received funding from the European Unions Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement number 722028 (ENIGMA ITN).

The cross-borehole GPR dataset data used for the field case in this work will be available at Pangaea via [DOI, persistent identifier link] with [license, access conditions].

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*(13), 247–276. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.1967.tb02159.x` doi: https://doi.org/10.1111/j.1365-246X.1967.tb02159.x

Bai, T., & Tahmasebi, P. (2020, September). Hybrid geological modeling: Combining machine learning and multiple-point statistics. *Computers & Geosciences*, *142*, 104519. Retrieved 2020-08-10, from `https://linkinghub.elsevier.com/retrieve/pii/S0098300419309124` doi: 10.1016/j.cageo.2020.104519

Bergmann, U., Jetchev, N., & Vollgraf, R. (2017, September). Learning Texture Manifolds with the Periodic Spatial GAN. *arXiv:1705.06566 [cs, stat]*. Retrieved 2020-11-23, from `http://arxiv.org/abs/1705.06566` (arXiv: 1705.06566)

Bora, A., Jalal, A., Price, E., & Dimakis, A. G. (2017, March). Compressed Sensing using Generative Models. *arXiv:1703.03208 [cs, math, stat]*. Retrieved 2020-07-17, from `http://arxiv.org/abs/1703.03208` (arXiv: 1703.03208)

Bording, T. S., Fiandaca, G., Maurya, P. K., Auken, E., Christiansen, A. V., Tuxen, N., … Larsen, T. H. (2019, October). Cross-borehole tomography with full-decay spectral time-domain induced polarization for mapping of potential contaminant flow-paths. *Journal of Contaminant Hydrology*, *226*, 103523. Retrieved 2020-11-23, from `https://linkinghub.elsevier.com/retrieve/pii/S0169772218303620` doi: 10.1016/j.jconhyd.2019.103523

Caers, J., & Hoffman, T. (2006, January). The Probability Perturbation Method: A New Look at Bayesian Inverse Modeling. *Mathematical Geology*, *38*(1), 81–100. Retrieved 2019-03-27, from `http://link.springer.com/10.1007/s11004-005-9005-9` doi: 10.1007/s11004-005-9005-9

Canchumuni, S. W., Emerick, A. A., & Pacheco, M. A. C. (2019, July). Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers & Geosciences*, *128*, 87–102. Retrieved 2019-11-20, from `https://linkinghub.elsevier.com/retrieve/pii/S0098300419300378` doi: 10.1016/j.cageo.2019.04.006

Chan, S., & Elsheikh, A. H. (2019, April). Parametrization and generation of geological models with generative adversarial networks. *arXiv:1708.01810 [physics, stat]*. Retrieved 2020-06-25, from `http://arxiv.org/abs/1708.01810` (arXiv: 1708.01810)

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995, September). The Helmholtz Machine. *Neural Computation*, *7*(5), 889–904. Retrieved 2020-

11-23, from https://doi.org/10.1162/neco.1995.7.5.889 (Publisher: MIT Press) doi: 10.1162/neco.1995.7.5.889

Fefferman, C., Mitter, S., & Narayanan, H. (2016, February). Testing the manifold hypothesis. *Journal of the American Mathematical Society*, *29*(4), 983–1049. Retrieved 2020-05-10, from http://www.ams.org/jams/2016-29-04/S0894-0347-2016-00852-4/ doi: 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. Retrieved 2016-05-06, from http://www.sciencedirect.com/science/article/pii/0022247X7090017X

Fukushima, K. (1980, April). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. Retrieved 2020-11-16, from http://link.springer.com/10.1007/BF00344251 doi: 10.1007/BF00344251

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014, June). Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*. Retrieved 2020-06-25, from http://arxiv.org/abs/1406.2661 (arXiv: 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. Retrieved from https://doi.org/10.1190/geo2013-0215.1 doi: 10.1190/geo2013-0215.1

Hansen, T. M., Cordua, K. S., & Mosegaard, K. (2012, June). Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling. *Computational Geosciences*, *16*(3), 593–611. Retrieved 2020-06-02, from http://link.springer.com/10.1007/s10596-011-9271-1 doi: 10.1007/s10596-011-9271-1

Hermans, T., Nguyen, F., & Caers, J. (2015, July). Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: Workflow and case study. *Water Resources Research*, *51*(7), 5332–5352. doi: 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. doi: 10.1016/j.jhydrol.2012.03.021

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., ... Lerchner, A. (2017). beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework..

Holm-Jensen, T., & Hansen, T. M. (2019, June). Linear Waveform Tomography Inversion Using Machine Learning Algorithms. *Mathematical Geosciences*. Retrieved 2019-11-27, from http://link.springer.com/10.1007/s11004-019-09815-7 doi: 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.

Journel, A., & Zhang, T. (2007, May). The Necessity of a Multiple-Point Prior Model. *Mathematical Geology*, *38*(5), 591–610. Retrieved 2020-08-28, from http://link.springer.com/10.1007/s11004-006-9031-2 doi: 10.1007/s11004-006-9031-2

Kessler, T. C., Comunian, A., Oriani, F., Renard, P., Nilsson, B., Klint, K. E., & Bjerg, P. L. (2013, September). Modeling Fine-Scale Geological Heterogeneity-Examples of Sand Lenses in Tills. *Groundwater*, *51*(5), 692–705. Retrieved 2017-07-11, from http://doi.wiley.com/10.1111/j.1745-6584.2012.01015.x doi: 10.1111/j.1745-6584.2012.01015.x

Kessler, T. C., Klint, K. E. S., Nilsson, B., & Bjerg, P. L. (2012, October). Characterization of sand lenses embedded in tills. *Quaternary Science Reviews*, *53*, 55–71. Retrieved 2020-07-02, from https://linkinghub.elsevier.com/retrieve/pii/S0277379112003125 doi: 10.1016/j.quascirev.2012.08.011

Kingma, D. P., & Welling, M. (2014, May). Auto-Encoding Variational Bayes. *arXiv:1312.6114 [cs, stat]*. Retrieved 2019-11-04, from http://arxiv.org/abs/1312.6114 (arXiv: 1312.6114)

Kleinberg, R., Li, Y., & Yuan, Y. (2018, August). An Alternative View: When Does SGD Escape Local Minima? *arXiv:1802.06175 [cs]*. Retrieved 2020-03-12, from http://arxiv.org/abs/1802.06175 (arXiv: 1802.06175)

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

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017, May). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. Retrieved 2020-11-16, from https://dl.acm.org/doi/10.1145/3065386 doi: 10.1145/3065386

Laloy, E., Hrault, R., Jacques, D., & Linde, N. (2018, January). Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network. *Water Resources Research*, *54*(1), 381–406. Retrieved 2018-07-13, from http://doi.wiley.com/10.1002/2017WR022148 doi: 10.1002/2017WR022148

Laloy, E., Hrault, R., Lee, J., Jacques, D., & Linde, N. (2017, December). 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. Retrieved 2018-07-13, from https://linkinghub.elsevier.com/retrieve/pii/S0309170817306243 doi: 10.1016/j.advwatres.2017.09.029

Laloy, E., Linde, N., Ruffino, C., Hrault, R., Gasso, G., & Jacques, D. (2019, December). Gradient-based deterministic inversion of geophysical data with generative adversarial networks: Is it feasible? *Computers & Geosciences*, *133*, 104333. Retrieved 2019-11-20, from https://linkinghub.elsevier.com/retrieve/pii/S009830041831207X doi: 10.1016/j.cageo.2019.104333

Larsen, T. H., Palstrm, P., & Larsen, L. (2016, November). *Kallerup Grusgrav: Kortlgning af sandlinser i morneler* (Tech. Rep. No. 3641400021). Roskilde, Danmark: Orbicon A/S.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989, December). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.*, *1*(4), 541–551. Retrieved from `https://doi.org/10.1162/neco.1989.1.4.541` (Place: Cambridge, MA, USA Publisher: MIT Press) doi: 10.1162/neco.1989.1.4.541

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

Looms, M. C., Klotzsche, A., van der Kruk, J., Larsen, T. H., Edsen, A., Tuxen, N., ... Nielsen, L. (2018, January). Mapping sand layers in clayey till using crosshole ground-penetrating radar. *GEOPHYSICS*, *83*(1), A21–A26. Retrieved 2019-11-20, from `https://library.seg.org/doi/10.1190/geo2017-0297.1` doi: 10.1190/geo2017-0297.1

Lopez-Alvis, J., Hermans, T., & Nguyen, F. (2019, November). A cross-validation framework to extract data features for reducing structural uncertainty in subsurface heterogeneity. *Advances in Water Resources*, *133*, 103427. Retrieved 2020-03-12, from `https://linkinghub.elsevier.com/retrieve/pii/S0309170819300661` doi: 10.1016/j.advwatres.2019.103427

Lopez-Alvis, J., Laloy, E., Nguyen, F., & Hermans, T. (2021, July). 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. Retrieved 2021-04-30, from `https://linkinghub.elsevier.com/retrieve/pii/S0098300421000698` doi: 10.1016/j.cageo.2021.104762

Lucic, M., Kurach, K., Michalski, M., Gelly, S., & Bousquet, O. (2018, October). Are GANs Created Equal? A Large-Scale Study. *arXiv:1711.10337 [cs, stat]*. Retrieved 2020-11-23, from `http://arxiv.org/abs/1711.10337` (arXiv: 1711.10337)

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., & Frey, B. (2015, November). Adversarial Autoencoders. *arXiv:1511.05644 [cs]*. Retrieved 2019-06-27, from `http://arxiv.org/abs/1511.05644` (arXiv: 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). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-319-78999-6_31` doi: 10.1007/978-3-319-78999-6_31

Mariethoz, G., Renard, P., & Straubhaar, J. (2010, November). The Direct Sampling method to perform multiple-point geostatistical simulations: PERFORMING MULTIPLE-POINTS SIMULATIONS. *Water Resources Research*, *46*(11). Retrieved 2017-07-06, from `http://doi.wiley.com/10.1029/`

2008WR007621 doi: 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, June). Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. *arXiv:1806.03720 [physics, stat]*. Retrieved 2020-02-09, from http://arxiv.org/abs/1806.03720 (arXiv: 1806.03720)
Park, H., Scheidt, C., Fenwick, D., Boucher, A., & Caers, J. (2013, August). History matching and uncertainty quantification of facies models with multiple geological interpretations. *Computational Geosciences*, *17*(4), 609–621. doi: 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, January). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434 [cs]*. Retrieved 2020-06-25, from http://arxiv.org/abs/1511.06434 (arXiv: 1511.06434)
Rezende, D. J., Mohamed, S., & Wierstra, D. (2014, May). Stochastic Backpropagation and Approximate Inference in Deep Generative Models. *arXiv:1401.4082 [cs, stat]*. Retrieved 2020-03-31, from http://arxiv.org/abs/1401.4082 (arXiv: 1401.4082)
Sajjadi, M. S. M., Bachem, O., Lucic, M., Bousquet, O., & Gelly, S. (2018, October). Assessing Generative Models via Precision and Recall. *arXiv:1806.00035 [cs, stat]*. Retrieved 2020-11-23, from http://arxiv.org/abs/1806.00035 (arXiv: 1806.00035)
Scheidt, C., Fernandes, A. M., Paola, C., & Caers, J. (2016, October). Quantifying natural delta variability using a multiple-point geostatistics prior uncertainty model: DELTA VARIABILITY AND GEOSTATISTICS. *Journal of Geophysical Research: Earth Surface*, *121*(10), 1800–1818. Retrieved 2018-01-16, from http://doi.wiley.com/10.1002/2016JF003922 doi: 10.1002/2016JF003922
Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying Uncertainty in Subsurface Systems* (No. 236). Hoboken, NJ & Washington D.C.: John Wiley and Sons & American Geophysical Union.
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*. Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved 2020-12-14, from http://link.springer.com/10.1007/978-3-662-05088-0 doi: 10.1007/978-3-662-05088-0
Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., & Besson, O. (2011, April). An Improved Parallel Multiple-point Algorithm Using a List Approach. *Mathematical Geosciences*, *43*(3), 305–328. Retrieved 2018-03-27, from http://link.springer.com/10.1007/s11004-011-9328-7 doi:

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. Retrieved 2017-07-17, from http://www.springerlink.com/index/8G2MEAGU5K0U07PK.pdf doi: 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.* New York: Winston and Sons.

Treister, E., & Haber, E. (2016, November). A fast marching algorithm for the factored eikonal equation. *Journal of Computational Physics*, *324*, 210–225. Retrieved 2020-03-23, from http://arxiv.org/abs/1607.00973 (arXiv: 1607.00973) doi: 10.1016/j.jcp.2016.08.012

Uria, B., Murray, I., & Larochelle, H. (2014, January). A Deep and Tractable Density Estimator. *arXiv:1310.1757 [cs, stat]*. Retrieved 2020-11-23, from http://arxiv.org/abs/1310.1757 (arXiv: 1310.1757)

van der Walt, S., Schnberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., … contributors, t. s.-i. (2014). scikit-image: image processing in Python. *PeerJ*, *2*, e453. Retrieved from https://doi.org/10.7717/peerj.453 doi: 10.7717/peerj.453

You, N., Li, Y. E., & Cheng, A. (2021, May). 3D Carbonate Digital Rock Reconstruction Using Progressive Growing GAN. *Journal of Geophysical Research: Solid Earth*, *126*(5). Retrieved 2021-05-31, from https://onlinelibrary.wiley.com/doi/10.1029/2021JB021687 doi: 10.1029/2021JB021687

Zhang, C., Butepage, J., Kjellstrom, H., & Mandt, S. (2018, October). Advances in Variational Inference. *arXiv:1711.05597 [cs, stat]*. Retrieved 2020-04-07, from http://arxiv.org/abs/1711.05597 (arXiv: 1711.05597)

Supporting Information for "Geophysical Inversion Using a Variational Autoencoder to Model an Assembled Spatial Prior Uncertainty"

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

[1]Urban and Environmental Engineering, Applied Geophysics, University of Liege, Liege, Belgium

[2]Department of Geology, Ghent University, Gent, Belgium

[3]Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark

Contents of this file

1. Text S1
2. Figures S1 to S2

Text S1. Checking the prior consistency

When the inversion described in section 2.2 is applied to a field case, it is important to check that the chosen prior is consistent with the data (Scheidt et al., 2018). Further, when considering an assembled prior, this check may allow to falsify some of the patterns before training the VAE, potentially improving the accuracy of the generated patterns and/or allowing for a lower dimensionality to be used for the latent space. This prior consistency or falsification step is done using the original generative model (Fig. 1a). The method applied here relies on approximating the marginal conditional distribution with respect

---

Corresponding author: J. Lopez-Alvis, (jlopez@uliege.be)

to the TI as $p(\mathbf{d}|TI) \approx p(\mathbf{d}^*|TI)$ where $\mathbf{d}^*$ refers to a lower-dimensional or compressed version of the data $\mathbf{d}$. Here, a number of samples from each TI and their corresponding simulations (using the forward operator) are obtained, then principal component analysis (PCA) is used to perform the dimensionality reduction. The conditional $p(\mathbf{d}^*|TI)$ is then approximated with adaptive kernel density estimation (KDE) (Park et al., 2013). Finally, the value of $p(\mathbf{d}^*|TI)$ at the observed data is compared to the probability density value at the 99 percent confidence region of a multivariate Gaussian distribution with the same dimension as $\mathbf{d}^*$. If the density value at the observed data is lower than the density value of the multivariate Gaussian, the TI is falsified or deemed inconsistent with the data.

The prior consistency check is performed for both the synthetic and field data (see sections 3.2 and 3.3). For this, 300 model samples (generated as in Fig. 1a) and their corresponding forward simulations are obtained for each training image. Then, the first three PCA components of these simulations and the data are used to compute the value of $p(\mathbf{d}^*|TI)$. The first three components were considered because they account for about 84 percent of data variability (explained variance). The density value at the contour of the 99 percent confidence region of a three-dimensional multivariate Gaussian distribution is equal to 2.2×10^{-4}, so any TI with a conditional density value lower than this is deemed non-consistent or very unlikely to have generated the data. Figure S1 shows the $p(\mathbf{d}^*|TI)$ for each TI. For both the synthetic case and the field case, all TIs show a conditional probability above the defined threshold, i.e. none of the TIs is falsified. Note that for the field data, TIs 3 and 5 are very close to the threshold. An additional visual check for these two TIs is performed by plotting of the data point together with the simulated data points (Figure S2), which confirms that the data point is in a low density region but it is still likely to be produced by each of the two TIs.

Note that the prior falsification step for the field data gives a rather low probability value for the best performing inversion case (compare TI_5 in Fig. 6b and Fig. S1b). This may be caused by: (1) the low number of samples used for the prior falsification (300 forward runs for each TI) and (2) the enhanced diversity caused by the VAE, i.e. even if the patterns in TI_5 did not produced sufficiently similar patterns to those giving rise to the field data, the VAE trained with this TI does produce such patterns.

References

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

Scheidt, C., Li, L., & Caers, J. (2018). *Quantifying Uncertainty in Subsurface Systems* (No. 236). Hoboken, NJ & Washington D.C.: John Wiley and Sons & American Geophysical Union.

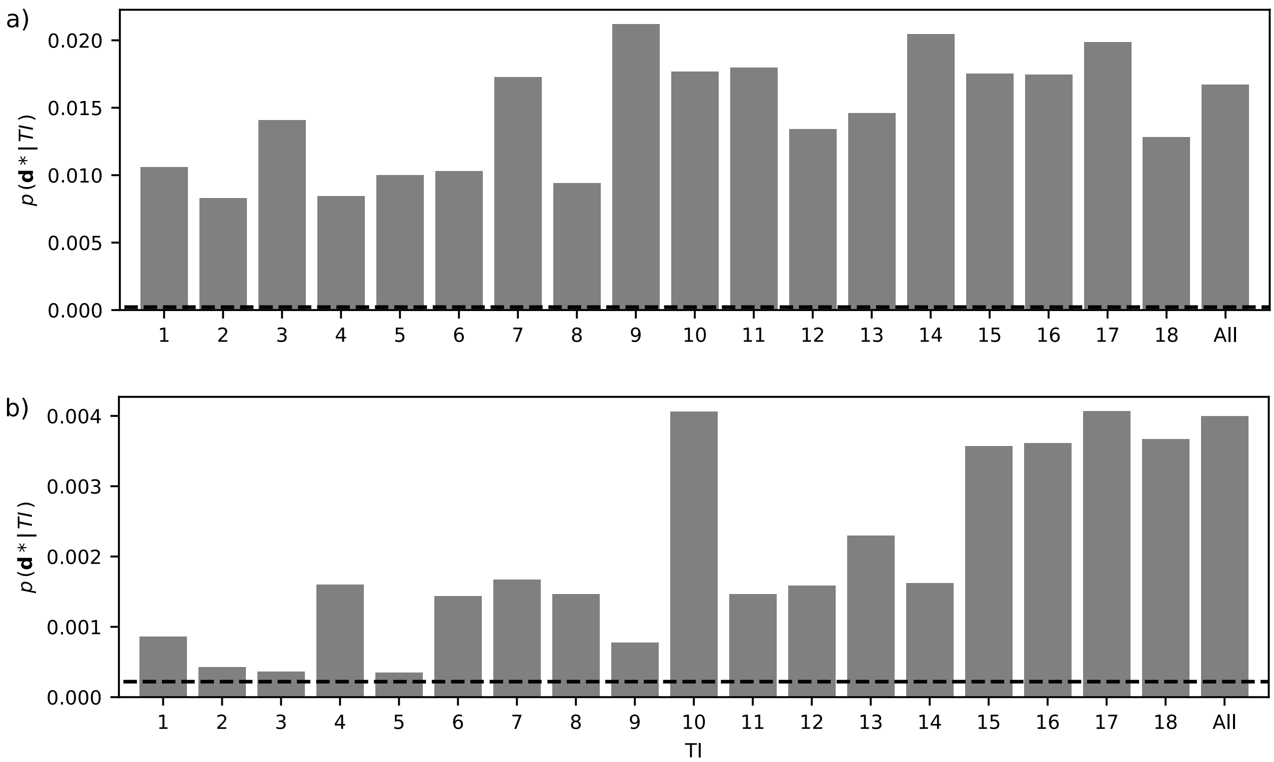

Figure S1. Prior falsification results for synthetic (a) and field case (b) for individual priors (VAEs trained on each of the 18 TIs) and the assembled prior (labeled "All"). Dashed line is the threshold for falsification.

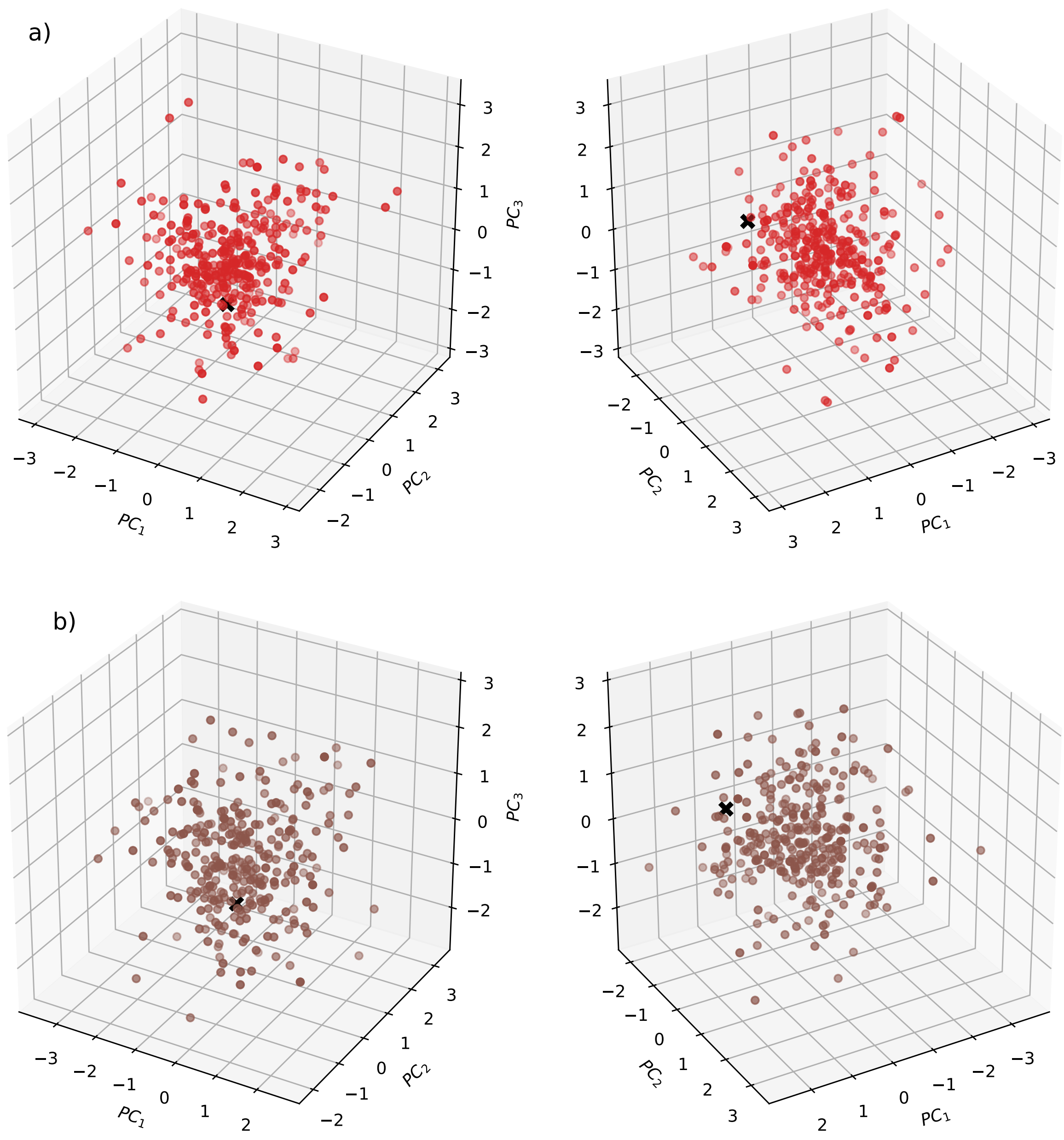

Figure S2. Principal components of simulated data and field data for TI_3 (a) and TI_5 (b). Simulated data is in colored dots and field data is denoted by the '×' symbol.