



Stochastic Seismic Waveform Inversion Using Generative Adversarial Networks as a Geological Prior

Lukas Mosser¹  · Olivier Dubrule¹ ·
Martin J. Blunt¹

Received: 23 October 2018 / Accepted: 30 September 2019 / Published online: 15 November 2019
© The Author(s) 2019

Abstract We present an application of deep generative models in the context of partial differential equation constrained inverse problems. We combine a generative adversarial network representing an a priori model that generates geological heterogeneities and their petrophysical properties, with the numerical solution of the partial-differential equation governing the propagation of acoustic waves within the earth's interior. We perform Bayesian inversion using an approximate Metropolis-adjusted Langevin algorithm to sample from the posterior distribution of earth models given seismic observations. Gradients with respect to the model parameters governing the forward problem are obtained by solving the adjoint of the acoustic wave equation. Gradients of the mismatch with respect to the latent variables are obtained by leveraging the differentiable nature of the deep neural network used to represent the generative model. We show that approximate Metropolis-adjusted Langevin sampling allows an efficient Bayesian inversion of model parameters obtained from a prior represented by a deep generative model, obtaining a diverse set of realizations that reflect the observed seismic response.

Keywords Ill-posed problems · Bayes · Deep learning · Earth modeling

Mathematics Subject Classification 86A22 · 62F15 · 86A32

✉ Lukas Mosser
lukas.mosser15@imperial.ac.uk

Olivier Dubrule
o.dubrule@imperial.ac.uk

Martin J. Blunt
m.blunt@imperial.ac.uk

¹ Imperial College London, London, UK

1 Introduction

Solving an inverse problem means finding a set of model parameters that best fit observed data (Tarantola 2005). The observed data or measurements are often noisy and/or sparse, and therefore lead to an ill-posed inverse problem where numerous realizations of the underlying model parameters may lead to a model response that matches observed data (Kabanikhin 2008). Additionally, the model used to describe how the observed data are generated, the so-called forward model, may be uncertain (Hansen and Cordua 2017).

Based on natural observations or an understanding of the underlying data generating process we may have a preconception about possible or impossible states of the model parameters. We may formulate this knowledge as a prior probability distribution function (PDF) of our model parameters and use Bayesian inference to obtain a posterior PDF of the model parameters given the observations (Tarantola 2005).

Seismic inversion involves modeling the physical process of waves radiating through the earth's interior (Fig. 1). By comparing the simulated synthetic measurements to actual acoustic recordings of reflected waves, we can modify model parameters and minimize the misfit between synthetic data and measurements. The adjoint of the partial differential equation (PDE) represents the gradient of the data mismatch with respect to the parameters, leading to a gradient-based optimization of the model parameters (Plessix 2006). In the most general case, which has been used in this study, these gradients are obtained by back-propagating the full wavefield in time, an approach commonly referred to as full-waveform inversion (FWI). The set of parameters represented by the spatial distribution of the acoustic velocity of the rocks within the earth can easily exceed 10^6 values, depending on the resolution of the simulation grid and the observed data. Large three-dimensional seismic observations may require millions of parameters to be inverted for, demanding enormous computational resources (Akcelik et al. 2003).

For direct observations of the earth's interior, boreholes may have been drilled for hydrocarbon exploration/development or hydrological measurements. These represent a quasi-one-dimensional source of information of spatially sparse nature. Typical borehole sizes are on the order of tens of centimeters in diameter, whereas the lateral resolution of seismic observations is usually on the order of tens of meters.

We can deduce prior knowledge of the earth's interior from observations of analog outcrops or subsurface reservoirs. This geological knowledge can be incorporated into prior distributions of physical properties of rocks, such as the acoustic P-wave velocity, or into the distribution of geological features such as geological facies and fault distributions within the earth.

Efficient parameterizations (Akcelik et al. 2002; Kadu et al. 2016) that enable a dimensionality-reduced representation of the high-dimensional parameter space of possible models have been shown to reduce computational cost and increase spatial resolution. Because of the high computational cost incurred by full-waveform inversion (Modrak and Tromp 2015; Akcelik et al. 2003), probabilistic ensembles of models that match observed data are rarely generated, and often only a single model that satisfies predefined quality criteria is created and used for interpretation and decision-making processes.

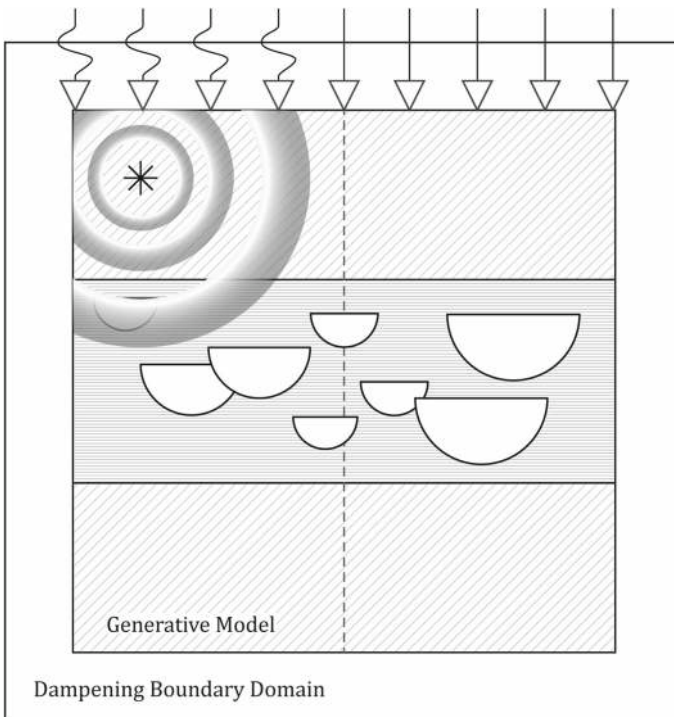


Fig. 1 Computational domain for the acoustic inversion problem. Acoustic recording devices are placed on the surface (∇) and record incoming acoustic waves reflected from geological structures emanating from an artificial source (*). The computational domain is embedded within a dampened boundary domain to emulate lateral and vertical dissipation of the wave-source. The generative model $G_\theta(\mathbf{z})$ creates the underlying spatially distributed P-wave velocity. Additional lower-dimensional constraints (dashed vertical line representing a well) can be placed on the generative model, by incorporating loss terms. The vertical axis of the computational domain has been rescaled by a factor of 10 for visualization purposes

We parameterize the earth model by a deep generative model that creates stochastic realizations of possible model parameters. The probabilistic distribution of model parameters is parameterized by a lower-dimensional set of multi-Gaussian-distributed latent variables. Combined with a generative deep neural network, this represents a differentiable prior on the possible model parameters. We combine this differentiable generative model with the numerical solution of the acoustic wave equation to produce synthetic acoustic observations of the earth's interior (Louboutin et al. 2017). Using the adjoint method (Plessix 2006), we compute a gradient of the mismatch between real and synthetic data with respect to model parameters not only in the high-dimensional model space, but also in the much smaller set of latent variables. These gradients are required to perform a Metropolis-adjusted Langevin (MALA) sampling of the posterior of the model parameters given the observed seismic data. Performing MALA sampling allows us to obtain a diverse ensemble of model parameters that match the observed seismic data. Additional constraints on the generative model, such as information located at existing boreholes, are readily incorporated and included in the MALA sampling procedure.

We summarize our contributions as follows:

- (i) We combine a differentiable generative model controlled by a set of latent variables with the solution of a PDE-constrained numerical solution of a physical forward problem.
- (ii) We use gradients obtained from the adjoint method and from neural network back-propagation to perform approximate MALA sampling of the posterior in the lower-dimensional set of latent variables.
- (iii) We illustrate the proposed inversion framework using a simple synthetic seismic inversion problem and evaluate the resulting ensemble of model parameters.
- (iv) The framework allows integration of additional information, such as the knowledge of geological facies along one-dimensional vertical boreholes.
- (v) The proposed approach may be readily extended to a number of inverse problems where gradients of the objective function with respect to input parameters can be calculated.

The code, data and trained weights of the neural networks have been made available under an open-source license.¹

2 Related Work

Tarantola (2005) cast the geophysical seismic inversion problem in a Bayesian framework. Mosegaard and Tarantola (1995) presented a general methodology to perform probabilistic inversion using Monte Carlo sampling. They used a Metropolis rule combined with a sampling of the prior to obtain the posterior distribution. In a similar manner, Sen and Stoffa (1996) evaluated the use of Gibbs sampling to obtain *a posteriori* model parameters and evaluate parameter uncertainties. Mosegaard (1998) showed that the general Bayesian inversion approach of Mosegaard and Tarantola (1995) also gives information on the ability to resolve geological features. Geostatistical models enable spatial relationships and dependencies of the petrophysical parameters to be modeled and incorporated into a stochastic inversion framework (Bortoli et al. 1993; Haas and Dubrule 1994). Bayesian linear inversion has been successfully applied to infer petrophysical property distributions (Grana and Della Rossa 2010). Buland and Omre (2003) developed an approach to perform Bayesian inversion for elastic petrophysical properties in a linearized setting. Grana et al. (2017) used a Gaussian mixture model for Bayesian linear inversion from seismic and well data. Stochastic sampling of petrophysical properties conditioned to well-log data allows petrophysical property distributions to be inferred using an appropriate sampling strategy such as Markov chain Monte Carlo (MCMC) (Bosch et al. 2009). A fully integrated stochastic inversion method that allows direct inversion from seismic amplitude-versus-angle (AVA) data creates a direct link between observed seismic data and underlying rock physics models (Azevedo et al. 2018). Geological modeling using multi-point statistics (Guardiano and Srivastava 1993) can be employed for inversion from seismic data (González et al. 2007) where geological features are represented by a set of representative training images. For a more extensive review of statistical inversion approaches

¹ Code Repository: https://github.com/LukasMosser/Stochastic_Seismic_Waveform_Inversion.

we refer to Bosch et al. (2010) and the comprehensive overviews of Dubrule (2003), Doyen (2007), Azevedo and Soares (2017).

In the case of nonlinear physics-based inversion schemes such as FWI, computation of the solution to the forward problem is very expensive. Therefore, computationally efficient approximations to the full solution of the wave equation may allow efficient solutions to complex geophysical inversion problems. Neural networks have been shown to be universal function approximators (Hornik et al. 1989) and as such lend themselves to use as possible proxy models for solutions to the geophysical forward and inverse problem (Hansen and Cordua 2017).

The early work by Röth and Tarantola (1994) presents an application of neural networks to invert from acoustic time-domain seismic amplitude responses to a depth profile of acoustic velocity in a supervised setting. They used pairs of synthetic data and velocity models to train a multi-layer feed-forward neural network with the goal of predicting acoustic velocities from recorded data only. They showed that neural networks can produce high-resolution approximations to the solution of the inverse problem based on representations of the input model parameters and resulting synthetic waveforms alone. In addition, they showed that neural networks can invert for geophysical parameters in the presence of significant levels of acoustic noise.

Representing the geophysical model parameters at each point in space quickly leads to a large number of model parameters, especially in the case of three-dimensional problems. Berg and Nyström (2017) represented the spatially varying coefficients that govern the solution of a PDE by a neural network. The neural network acts as an approximation to the spatially varying coefficients characterized by the weights of the neural network. The weights of the individual neurons are modified by leveraging the adjoint-state equation in the reduced-dimensional space of network parameters rather than at each spatial location of the computational grid.

Hansen and Cordua (2017) replaced the solution of the partial differential equation by a neural network, enabling fast computation of forward models and facilitating a solution to the inversion problem by Monte Carlo sampling. Araya-Polo et al. (2018) used deep neural networks to perform a mapping between seismic features and the underlying P-wave velocity domain; they validated their approach based on synthetic examples. A number of applications of deep generative priors have recently been presented in the context of computer vision for image reconstruction, linear (Chang et al. 2017) and bilinear (Asim et al. 2018) inverse problems, and compressed sensing (Bora et al. 2017). Mosser et al. (2017) proposed GANs to generate three-dimensional stochastic realizations of porous media from binary and grayscale computed tomography images (Mosser et al. 2018b). These deep generative models can be further conditioned to honor lower-dimensional features such as cross-sections or borehole data (Dupont et al. 2018; Mosser et al. 2018a; Chan and Elsheikh 2018). For more general subsurface inverse problems, Laloy et al. (2017) used a GAN to create geological models for hydrological inversion. Inversion was performed using an adapted Markov chain Monte Carlo (MCMC) (Laloy and Vrugt 2012) algorithm where the generative model was used as an unconditional prior to sample hydrological model parameters. Chan and Elsheikh (2017) evaluated the applicability of Wasserstein-GANs to parameterize geological models for uncertainty propagation.

Mosser et al. (2018c) used a generative adversarial network with cycle constraints (cycleGAN) (Zhu et al. 2017) to perform seismic inversion, formulating the inversion task as a domain-transfer problem. Their work used a cycleGAN to map between the seismic amplitude domain and P-wave velocity models. The cycle constraint ensures that models obtained by transforming from the amplitude to P-wave velocity representation and back to the amplitude domain are consistent. Because the P-wave velocity models and seismic amplitudes are represented as a function of depth rather than depth and time, respectively, this approach lends itself to stratigraphic inversion, where a pre-existing velocity model is used to perform time-depth conversion of the seismic amplitudes. Richardson (2018) showed that a quasi-Newtonian method can optimize model parameters in the latent space of a pre-trained GAN for a synthetic salt-body benchmark dataset.

3 Problem Definition

3.1 Bayesian Inversion

In the Bayesian framework of inverse problems, we aim to find the posterior of latent variables \mathbf{z} given the observed data \mathbf{d}_{obs} (Fig. 2). The joint probability of the latent variables \mathbf{z} and observed data \mathbf{d}_{obs} is

$$p(\mathbf{z}, \mathbf{d}_{\text{obs}}) = p(\mathbf{d}_{\text{obs}}|\mathbf{z})p(\mathbf{z}) \quad (1)$$

Furthermore, by applying Bayes rule, we define the posterior over the latent variables \mathbf{z} given the observed seismic data \mathbf{d}_{obs}

$$p(\mathbf{z}|\mathbf{d}_{\text{obs}}) = \frac{p(\mathbf{d}_{\text{obs}}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{d}_{\text{obs}})} \propto p(\mathbf{d}_{\text{obs}}|\mathbf{z})p(\mathbf{z}) \quad (2)$$

We express the observed data by assuming conditional independence between the observed seismic data \mathbf{d}_{seis} and data observed at the wells \mathbf{d}_{well}

$$p(\mathbf{d}_{\text{obs}}|\mathbf{z}) = p(\mathbf{d}_{\text{seis}}|\mathbf{z})p(\mathbf{d}_{\text{well}}|\mathbf{z}) \quad (3)$$

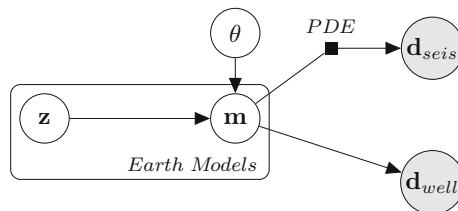


Fig. 2 Graphical model of the geological inversion problem. The set of possible earth models is represented by a generative model with parameters θ (the parameters of the generator $\mathbf{m} \sim G_{\theta}(\mathbf{z})$). We obtain model observations of the acoustic waves \mathbf{d}_{seis} via the deterministic PDE, as well as partial observation of the model parameters \mathbf{m} from local information, for example, at boreholes \mathbf{d}_{well}

We represent the observed seismic data by

$$\mathbf{d}_{\text{seis}} = S(\mathbf{m}) + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma_{\text{seis}}^2 \mathbf{I}) \tag{4}$$

where $S(\mathbf{m}) = S(m(\mathbf{x})) = S(G_\theta(\mathbf{z}))$, denoting the spatial model coordinates by \mathbf{x} , the seismic forward modeling operator by S , and the generative model by $G_\theta(\mathbf{z})$ with parameters θ . We assume a normally distributed noise term ε with zero mean and standard deviation σ_{seis} equal to 25% of the standard deviation of the reference model seismic amplitude data. The geological facies $\mathbf{m}^{\text{facies}}$, the P-wave velocity \mathbf{m}^{V_p} , and the rock density \mathbf{m}^ρ represent the set of model parameters \mathbf{m} . The model parameter $\mathbf{m}^{\text{facies}}$ represents the probability of a geological facies occurring at a spatial location \mathbf{x} .

The aim is to generate samples of the posterior $\mathbf{z} \sim p(\mathbf{z}|\mathbf{d}_{\text{obs}})$. We reformulate the approach using an iterative approximate Metropolis-adjusted Langevin sampling rule (MALA-approx) with iteration number t as follows (Roberts and Tweedie 1996; Roberts and Rosenthal 1998; Nguyen et al. 2016)

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \gamma_t \nabla \log p(\mathbf{z}_t|\mathbf{d}_{\text{obs}}) + \eta_t \tag{5a}$$

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \gamma_t \nabla \log [p(\mathbf{d}_{\text{seis}}|\mathbf{z}_t)p(\mathbf{d}_{\text{well}}|\mathbf{z}_t)p(\mathbf{z}_t)] + \eta_t \tag{5b}$$

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \gamma_t \{\nabla \log p(\mathbf{d}_{\text{seis}}|\mathbf{z}_t) + \nabla \log p(\mathbf{d}_{\text{well}}|\mathbf{z}_t) + \nabla \log p(\mathbf{z}_t)\} + \eta_t \tag{5c}$$

where $\eta_t \sim \mathcal{N}(0, 2\gamma_t \mathbf{I})$ is a sample from a Gaussian distribution with variance proportional to the step size γ_t at MALA iteration t . Assuming a Gaussian log-likelihood of the seismic data given the latent variables $\log p(\mathbf{d}_{\text{seis}}|\mathbf{z}_t) \propto -\|S(G_\theta(\mathbf{z}_t)) - \mathbf{d}_{\text{seis}}\|_2^2$ leads to the proposal rule of the MALA approximation (Nguyen et al. 2016) for the case when only seismic observations \mathbf{d}_{seis} are considered

$$\mathbf{z}_{t+1} = \mathbf{z}_t - \gamma_t \frac{\partial \|S(G_\theta(\mathbf{z}_t)) - \mathbf{d}_{\text{seis}}\|_2^2}{\partial G_\theta(\mathbf{z}_t)} \frac{\partial G_\theta(\mathbf{z}_t)}{\partial \mathbf{z}_t} + \gamma_t \nabla \log p(\mathbf{z}_t) + \eta_t \tag{6}$$

Using this sampling approach requires gradients of the data mismatch with respect to model parameters, which are obtained by the adjoint-state method which will be presented in the following section. The gradients of the model parameters $\frac{\partial G_\theta(\mathbf{z}_t)}{\partial \mathbf{z}_t}$ with respect to the latent variables are obtained by traditional neural network back-propagation. The gradient of the log-probability of the Gaussian prior distribution of latent variables $\nabla \log p(\mathbf{z}_t)$ can be interpreted as a regularization of the latent variables against deviation from the Gaussian prior assumption (Creswell and Bharath 2018).

We follow the MALA step-proposal algorithm using an initial step size $\gamma_{t=0} = 10^{-2}$ for every model inference (Xifara et al. 2013). To obtain valid samples of the posterior, we furthermore anneal the step size from the initial value of $\gamma_{t=0} = 10^{-2}$ to $\gamma_{t=200} = 10^{-5}$ over 200 iterations.

Where lower-dimensional information is available, such as at boreholes, the geological models should honor both the seismic response and this additional lower-dimensional information. In this study, we additionally find samples of the posterior that reflect observed geological facies indicators $\mathbf{d}_{\text{well}} = \mathbf{m}_{\text{well}}^{\text{facies}}$ at a one-dimensional

borehole. When including borehole information, the step-proposal corresponds to

$$\begin{aligned} \mathbf{z}_{t+1} = \mathbf{z}_t - \gamma_t & \frac{\partial \|S(G_\theta(\mathbf{z}_t)) - \mathbf{d}_{\text{seis}}\|_2^2}{\partial G_\theta(\mathbf{z}_t)} \frac{\partial G_\theta(\mathbf{z}_t)}{\partial \mathbf{z}_t} \\ & + \gamma_t \frac{\partial \log p(\mathbf{d}_{\text{well}} = \mathbf{m}_{\text{well}}^{\text{facies}} | \mathbf{z}_t)}{\partial \mathbf{z}_t} + \gamma_t \nabla \log p(\mathbf{z}_t) + \eta_t \end{aligned} \quad (7)$$

where we obtain samples of the posterior given the observed seismic data \mathbf{d}_{seis} and geological facies at the wells $\mathbf{d}_{\text{well}} = \mathbf{m}_{\text{well}}^{\text{facies}}$.

The additional term $\log p(\mathbf{d}_{\text{well}} = \mathbf{m}_{\text{well}}^{\text{facies}} | \mathbf{z}_t)$ in Eq. 7 represents the assumption of a Bernoulli distribution for the facies as derived from the generator and observed at the borehole.

3.2 Adjoint-State Method

We perform numerical solutions of the time-dependent acoustic wave equation given a set of model parameters

$$F(\mathbf{u}, \mathbf{m}^{V_p}) = \frac{1}{m^{V_p}(\mathbf{x})^2} \frac{d^2 u(\mathbf{x}, t)}{dt^2} - \Delta u(\mathbf{x}, t) + \eta \frac{du(\mathbf{x}, t)}{dt} - q(\mathbf{x}, \mathbf{x}_s, t) = 0 \quad (8)$$

where $u(\mathbf{x}, t)$ is the unknown wave-field and $m^{V_p}(\mathbf{x})$ is the acoustic P-wave velocity. The damping term $\eta \frac{du(\mathbf{x}, t)}{dt}$ prevents reflections from domain boundaries and ensures that waves dissipate laterally. We refer to the evaluation of $F(\mathbf{u}, \mathbf{m}^{V_p}) = 0$ (Eq. 8) as the forward problem.

Time-dependent source wavelets $q(\mathbf{x}, \mathbf{x}_s, t)$ are introduced at locations \mathbf{x}_s . We emulate the seismic acquisition process by placing regularly spaced acoustic receivers that record the incoming wave-field at the top edge of the simulation domain (Fig. 1). To show the impact of adding additional information from the acoustic forward problem to the posterior PDF of models, we perform Bayesian inversion using the proposed approach in a number of scenarios where we increase the number of acoustic shot data from 2 to 27 acoustic sources.

To perform sampling according to the MALA algorithm presented in Eq. 6, we seek to obtain a gradient of the following functional

$$J(m^{V_p}(\mathbf{x})) = \sum_{i=1}^{n_{\text{sources}}} \|\mathbf{d}_{\text{seis}}^{\text{pred}}(m^{V_p}(\mathbf{x}), q_i) - \mathbf{d}_{\text{seis}}\|_2^2 \quad (9)$$

where $\mathbf{d}_{\text{seis}}^{\text{pred}}$ and \mathbf{d}_{seis} are the predicted and observed seismic observations, respectively.

We augment the functional $J(m^{V_p}(\mathbf{x}))$ by forming the Lagrangian

$$\mathcal{L}(\mathbf{m}^{V_p}, \mathbf{u}, \lambda) = J(\mathbf{m}^{V_p}) - \langle \lambda, F(\mathbf{u}, \mathbf{m}^{V_p}) \rangle \quad (10)$$

Differentiating $\mathcal{L}(\mathbf{m}^{V_p}, \mathbf{u}, \lambda)$ with respect to λ leads to the state equation 8, but differentiation with respect to the acoustic wave-field \mathbf{u} leads to the adjoint state equations (Plessix 2006)

$$\left(\frac{\partial F(\mathbf{u}, \mathbf{m}^{V_p})}{\partial \mathbf{u}}\right)^T \lambda = (\mathbf{d}_{\text{seis}}^{\text{pred}} - \mathbf{d}_{\text{seis}}) \tag{11}$$

showing that we obtain a similar back-propagation equation as that used to derive gradients in neural networks (LeCun et al. 1988): the data mismatch is back-propagated thanks to a linear equation in the adjoint state vector λ . By differentiating the Lagrangian in Eq. 10 with respect to $m(\mathbf{x})$ we obtain

$$\frac{\partial J}{\partial \mathbf{m}^{V_p}} = \frac{\partial \|S(G_\theta(\mathbf{z}_t)) - \mathbf{d}_{\text{seis}}\|_2^2}{\partial G_\theta(\mathbf{z}_t)} = \left\langle \lambda, \frac{\partial F(\mathbf{u}, \mathbf{m}^{V_p})}{\partial \mathbf{m}^{V_p}} \right\rangle \tag{12}$$

which is the gradient required to perform MALA sampling of the posterior distribution of latent variables, Eq. 6.

We perform a numerical solution of the acoustic wave equation and the respective adjoint computation using the domain-specific symbolic language Devito (Kukreja et al. 2016; Louboutin et al. 2017). The numerical solution is performed using a fourth-order finite-difference scheme in space and second-order in time.

4 Generative Model

We use a generative model to sample realizations of spatially varying model parameters $m(\mathbf{x}) \sim G_\theta(\mathbf{z})$. These realizations are obtained by sampling a number of latent variable vectors \mathbf{z} . The associated model representations represent the a priori knowledge about the spatially varying properties of the geological structures in the subsurface.

We model the prior distribution of the spatially varying model parameters $m(\mathbf{x})$ (Sect. 3.1) by a generative adversarial network (GAN) (Goodfellow et al. 2014). GANs represent a generative model where the underlying probability density function is implicitly defined by a set of training examples. To train GANs, two functions are required: a generator $G_\theta(\mathbf{z})$ and a discriminator $D_\omega(\mathbf{m})$. The role of the generator is to create random samples of an implicitly defined probability distribution that are statistically indistinguishable from a set of training examples. The discriminator’s role is to distinguish real samples from those created by the generator. Both functions are trained in a competitive two-player min-max game where the overall loss is defined by

$$\min_{\theta} \max_{\omega} \{ \mathbb{E}_{\mathbf{m} \sim p_{\mathbf{m}}} [\log D_\omega(\mathbf{m})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}} [\log (1 - D_\omega(G_\theta(\mathbf{z})))] \} \tag{13}$$

Because of the opposing nature of the objective functions, training GANs is inherently unstable, and finding stable training methods remains an open research problem. Nev-

ertheless, a number of training methods have been proposed that allow more stable training of GANs. In this work we use a so-called Wasserstein-GAN (Arjovsky et al. 2017; Gulrajani et al. 2017; Chan and Elsheikh 2017), that seeks to minimize the Wasserstein distance between the generated and real probability distribution. We use a Lipschitz penalty term proposed by Petzka et al. (2017) to stabilize training of the Wasserstein-GAN. For the discriminator, keeping the parameters θ of the generator fixed, we minimize

$$\min_{\omega} \{ \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}} [D_{\omega}(G_{\theta}(\mathbf{z}))] - \mathbb{E}_{\mathbf{m} \sim p_{\mathbf{m}}} [D_{\omega}(\mathbf{m})] + \lambda_{LP} \mathbb{E}_{\hat{\mathbf{m}} \sim p_{\hat{\mathbf{m}}}} [(\max\{0, \|\nabla D_{\omega}(\hat{\mathbf{m}})\| - 1\})^2] \} \quad (14)$$

where $\hat{\mathbf{m}}$ is linear combination between a real and generated sample controlled by a random variable τ (Petzka et al. 2017). For the generator, keeping the parameters of the discriminator ω constant, we minimize

$$-\mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}} [D_{\omega}(G_{\theta}(\mathbf{z}))] \quad (15)$$

In our work we set $\lambda_{LP} = 200$ to train the generative model. We represent both the generator and discriminator² function by deep convolutional neural networks (see Appendix Table 1). The generator uses a number of convolutional layers followed by so-called pixel-shuffle transformations to create output models (Shi et al. 2016).

The latent vector is parameterized as a multivariate standardized normal distribution

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})^{50 \times 1 \times 2} \quad (16a)$$

$$G_{\theta} : \mathbf{z} \rightarrow \mathbb{R}^{3 \times 64 \times 128} \quad (16b)$$

Because of the geological properties represented in our dataset, namely, geological facies indicators $\mathbf{m}^{\text{facies}}$, acoustic P-wave velocity \mathbf{m}^{V_p} and density \mathbf{m}^{ρ} , the generator must output three data channels. We represent the geological facies as the probability of a spatial location belonging to a sandstone facies. To facilitate numerical stability of the GAN training process, we apply a hyperbolic tangent activation function and convert to a probability $\mathbf{m}^{\text{facies}}$ for subsequent computation (Eq. 7). We apply a hyperbolic tangent activation function to model the output distribution of the P-wave model parameters \mathbf{m}^{V_p} . For rock density \mathbf{m}^{ρ} , a soft-plus activation function is used to ensure positive values (Appendix A.1). In this study, only the facies indicator $\mathbf{m}^{\text{facies}}$ and acoustic P-wave velocity \mathbf{m}^{V_p} are used in the inversion process.

The generator-discriminator pairing is trained on the set of training images described in Sect. 5. GAN training required approximately 8 hours on eight NVIDIA K80 graphics processing units. A set of samples obtained from the GAN prior are presented in Appendix Fig. 9. After training, the generator $G_{\theta}(\mathbf{z})$ and the forward modeling operator $S(\mathbf{m})$ are arranged in a fully differentiable computational graph. To accommodate the sources and receivers of the acoustic forward modeling process described in Sect. 3 and Fig. 1, we pad the output of the generator by a domain of constant P-wave velocity.

² In the Wasserstein-GAN literature, the discriminator is also termed a “critic”.

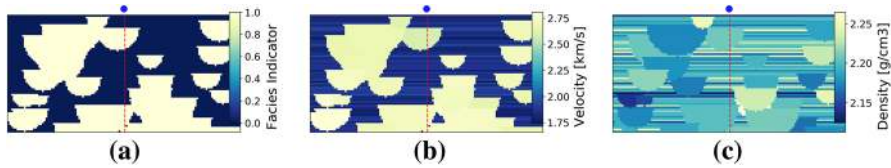


Fig. 3 Overview of the object-based model realization used as a reference model for evaluating the inversion procedure. Geological facies **a** distinguish between river channel bodies (light) and shale (dark). **b** Acoustic P-wave velocity V_p and **c** rock density ρ are constant within river channels and vary by layer within shale

5 Dataset

To demonstrate the proposed inversion method, we will use a model of a fluvial-dominated system consisting of highly porous sandstones embedded in a fine-grained shaly material. Object-based models are commonly used to model such geological systems (Deutsch and Wang 1996). They represent the fluvial environment as a set of randomly located geometric objects following various size, shape, and property distributions. We train a set of GANs on a dataset of 10,000 realizations of two-dimensional cross-sections of fluvial object-based models.

The individual cross-sections are created with an object-based model, where half-circle sand-bodies follow a uniform width distribution. P-wave velocity and density are constant within each channel-body, and their values are sampled independently from a Gaussian distribution for each individual channel-body. The locations of the channel-bodies are determined by a uniform distribution in spatial location. The fine-grained material surrounding the river systems comprises layers of single-pixel thickness, where each layer has a constant value of acoustic P-wave velocity and density which varies randomly from one layer to another and is sampled from a Gaussian distribution. We use a binary indicator variable to distinguish the two facies regions, river channel versus shale matrix. The ratio of how much of a given cross-section is filled with river channels compared with the overall area of the geological domain is a key property in understanding the geological nature of these structures. This ratio follows a uniform distribution from 30 to 60% in our dataset, and river channels are placed at random until a cross-section meets the randomly sampled ratio.

A total of 10,000 training images were created as a training set for the GAN. A further 4000 images were retained as a test set to evaluate the inversion technique. While training the generative model outlined in Sect. 4, we monitor image quality and output distribution for each of the modeled properties. The reference realization (Fig. 3) used to evaluate the Bayesian inversion approach was chosen randomly from the test set of object-based models. Figure 3 shows a comparison of the distribution of the three modeled properties: geological facies indicator, acoustic P-wave velocity, and rock density for the reference model.

6 Results

We evaluate the proposed method of inversion by sampling a set of latent variables \mathbf{z} determining the output of the generative model $G_\theta(\mathbf{z})$ (Sect. 5, Fig. 3). First, we

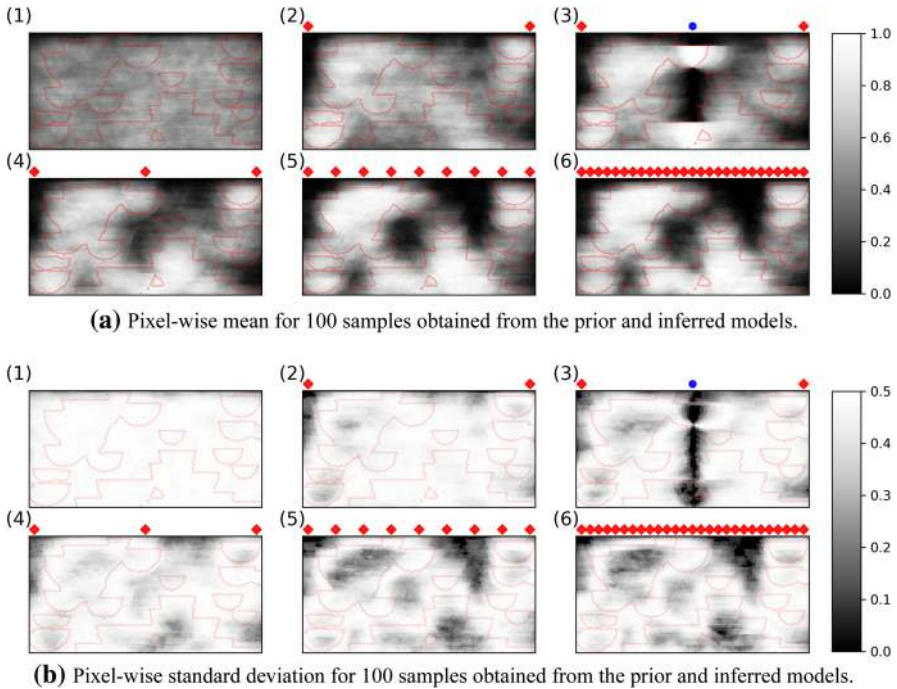


Fig. 4 **a** Pixel-wise mean and **b** standard deviation of the ensemble of 100 models sampled unconditionally from the prior (1) represented by the generator function $G_\theta(\mathbf{z})$. Posterior ensemble of geological indicator variables matched to the seismic representation of the reference model shown in Fig. 3 for (2) two sources, (3) two sources and a single borehole, (4) three sources, (5) nine sources, (6) 27 sources. Source locations are indicated by red diamonds and the borehole location by a blue circle. The reference model is indicated by red contours

evaluate the generative model as a prior for representing possible earth models and generating $N = 100$ unconditional samples (Fig. 4-1, Appendix Fig. 9).

Two cases of inversion are considered: inversion for the acoustic P-wave velocity V_p and combined inversion of acoustic velocity and of geological facies along a borehole. In all the cases presented, we assume that density is a constant. For all tests, we perform inversion using the approximate MALA scheme. For the additional borehole constraint, we require accuracy of geological facies of above 95% to be accepted as a valid inverted sample. While lower errors in seismic mismatch and borehole accuracy can be achieved, evaluating the forward problem and adjoint of the partial differential equation comes at a high computational cost, and therefore a cost-effectiveness trade-off was necessary.

For the first case of seismic inversion without borehole constraints, we perform simulations where the number of acoustic sources are increased. Fewer acoustic sources means that less of the domain is properly imaged, leading to high uncertainty in areas where no incoming waves have been reflected and recorded by the receivers on the surface. The acoustic sources and 128 receivers are equally spaced across the top edge of the domain.

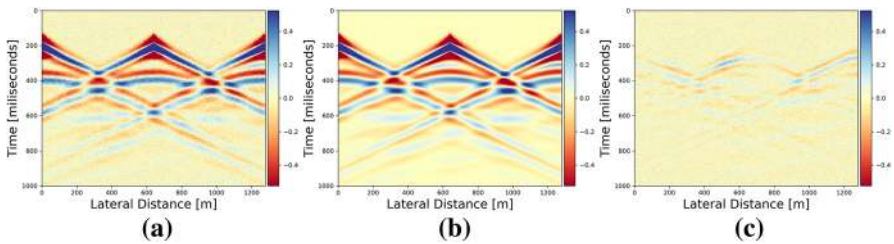


Fig. 5 Comparison of **a** the seismic waveform based on the reference model acoustic velocity with **b** the waveform of an inferred model with three seismic sources. The difference **c** in amplitude of the two waveforms. Color maps are scaled based on one standard deviation in amplitudes of the reference model waveform (**a**)

In Fig. 4 we show the pixel-wise mean (Fig. 4a) and standard deviation (Fig. 4b) of 100 inferred models for an increasing number of acoustic sources (from 2 to 27 sources). As the total number of acoustic sources increases, we obtain a lower standard deviation for the resulting model ensembles. In the case of two acoustic sources (Fig. 4b-2), we find that close to the sources, there is a small variation among the inferred models (dark shades), whereas the central area where no acoustic source has been placed shows a very high degree of variation. This is confirmed by the three-source case where a central acoustic source has been placed in addition to the sources on the borders of the domain. Lower variability in the inverted ensemble can be observed. This correlates well with the Bayesian interpretation of the inverse problem: where acoustic sources allow the subsurface to be imaged, we arrive at a low standard deviation in the posterior ensemble of geological models, whereas within regions that are only sparsely sampled by the acoustic sources, we expect the prior—the unconditional generative model—to be more prevalent, leading to a higher variability in geological features. As expected, when we increase the number of sources, we find overall smaller variability in the resulting ensemble of inverted earth models. We observe only marginal reduction in variability between the cases with 9 and 27 sources (Fig. 4b, 5, 6). For all inversion scenarios considered, we present samples from the posterior in the Appendix (Figs. 10, 11, 12, 13, 14).

In the case where lower-dimensional information such as a borehole was included as an additional objective function constraining the generative model (Fig. 4b-3), we find a lower standard deviation around this borehole. The standard deviation along the borehole is close to zero due to the per-realization 95% accuracy constraint. Furthermore, there is a region of influence where the borehole constrains lateral features such as channel bodies. This is shown by channel-shaped features of low standard deviation at the top and bottom of the domain. Comparison with the reference model (Fig. 3-a) shows that two channel bodies can be found along the one-dimensional feature.

For each generated realization we have recorded the ratio of the squared error norm (Eq. 6) and the squared norm of the noise in the seismic data (Fig. 6) at each MALA sampling iteration. The global minimum of the data mismatch in the presence of Gaussian noise is reached when the objective function value is equal to the squared norm of the noise in the data, i.e., at a ratio equal to 1 (Fig. 6). In practice, we find that performing 200 MALA iterations leads to a sufficient reduction in the mismatch

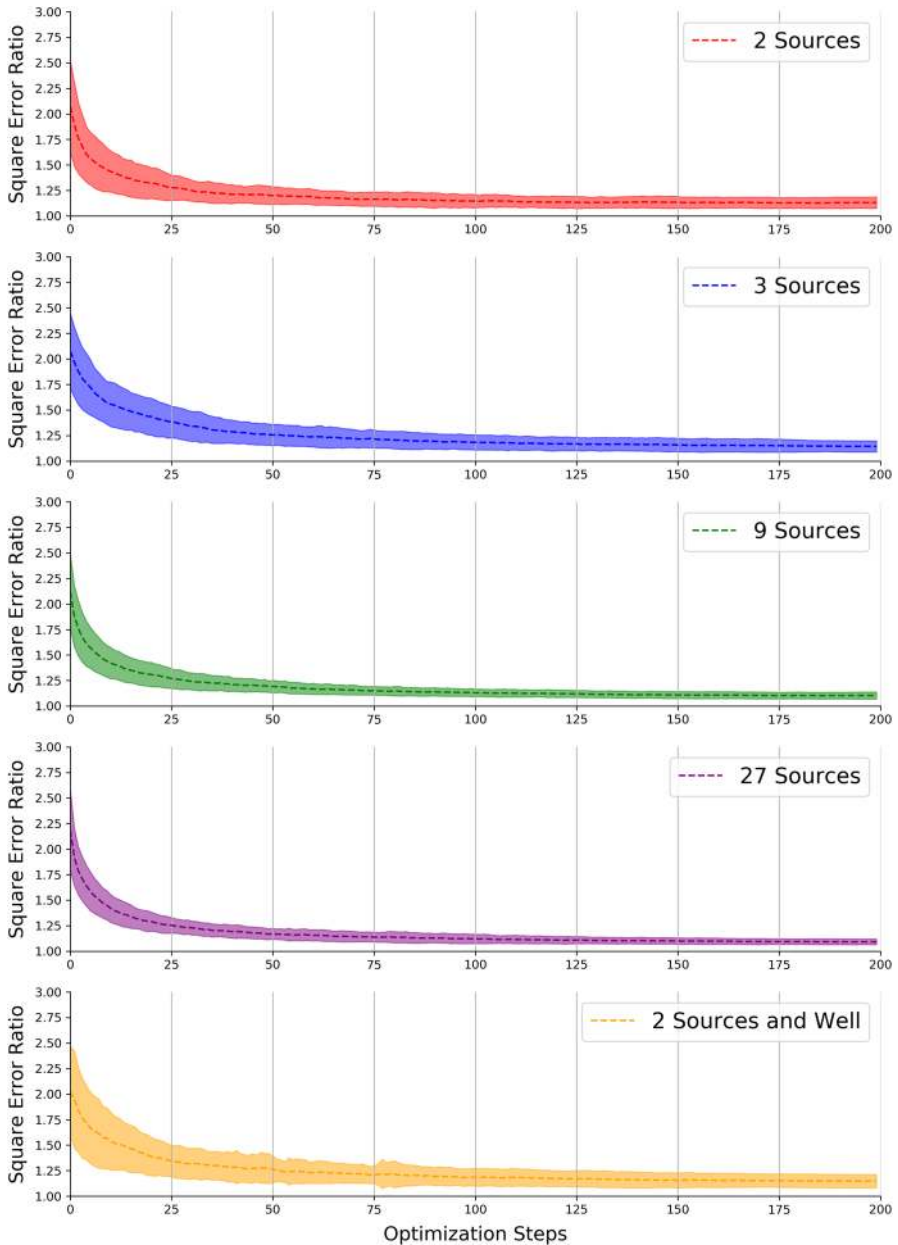


Fig. 6 Comparison of the ratio of the squared error norm and the squared norm of the Gaussian noise. The global minimum is reached at values of 1. Shaded regions indicate $\pm \sigma$ of the squared error ratio. We perform 200 approximate MALA iterations to obtain samples of the posterior given seismic observations only, as well as where borehole information and seismic observations are included. The step size was annealed to very small values, leading to a stabilization of the squared error norm at the end of the sampling procedure

of the seismic data, and as required by the approximate MALA algorithm, the error stabilizes as the step size is reduced.

Because modern FWI methods come at very high computational cost for two- and possibly three-dimensional inversion, a small number of required iterations is imperative. In further tests, reducing the number of iterations of the MALA approximation or simply optimizing by gradient descent, as performed by Richardson (2018), enables convergence to small errors, but this approach has been shown to lead to reduced sample diversity (Nguyen et al. 2016).

7 Discussion

We have shown that it is possible to obtain posterior realizations inferred from the latent space of a GAN generator that honor seismic and well-bore data by using an approximate Bayesian sampling method. A number of open questions remain concerning the generative model and the posterior distribution of models that are obtained.

A common challenge with GANs specifically is their so-called mode-collapse behavior, where the distribution represented by the generative model has collapsed to one or a few modes of the distribution implicitly represented by the set of training images. GANs do not represent the density explicitly, and therefore it is not possible to evaluate the ability of a GAN to represent the distribution by, for example, evaluating the likelihood of a set of test images given the model parameters $p(\mathbf{m}|\theta)$. Theis et al. (2015) have shown that evaluating sample quality and diversity of generative adversarial networks is difficult. Nevertheless, a number of heuristic approaches have been proposed, such as the inception score (IS) (Salimans et al. 2016) or the Fréchet Inception Distance (FID) (Heusel et al. 2017), and while these methods are popular for evaluating GANs trained on natural images, they may not be representative measures for comparing GANs, as shown by Barratt and Sharma (2018). Arora and Zhang (2017) propose a method to empirically evaluate the support of the distribution represented by a GAN.

Another common failure case of GANs occurs when the generator only memorizes the images of the training set and does not learn a representation of the entire distribution. In this case, it should only be possible to infer models which are part of the training set and which match the well and seismic data associated with the reference model. In the following, we investigate whether the ensemble of models obtained by solving the inverse problem represent new stochastic realizations of the underlying distribution implicitly represented by the training images.

We have evaluated the mean-squared-error (MSE) and the structural similarity index (SSIM) (Wang et al. 2004) between pairs of binary facies models. A perfect agreement between two models is reached for an MSE of zero and an SSIM of one. The MSE, while being a common measure for comparing pairs of data, is very sensitive to small translations of the models that are compared. The structural similarity index attempts to capture perceptual similarity and is less sensitive to pixel-wise differences in the two compared models (Wang and Bovik 2009).

In Fig. 7 we show kernel density estimates for the distributions of the two image similarity measures. First, we compare the distribution of the MSE and SSIM between

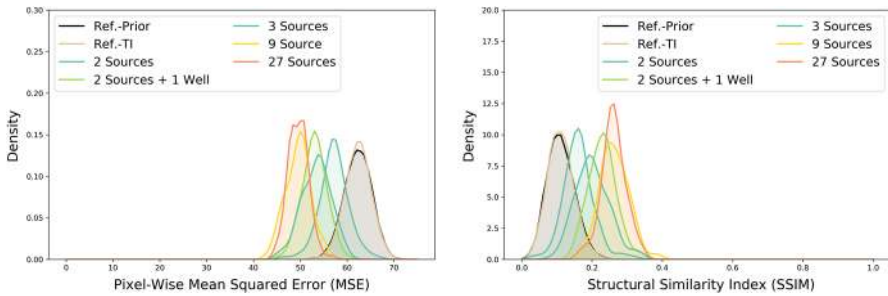


Fig. 7 Kernel density estimates of the distributions of the (left) mean-squared-error and (right) structural similarity index (SSIM) with respect to the reference model for models sampled from the GAN prior, and inferred models obtained by Bayesian inversion

the reference model and the 10^4 models in the training set (Ref.-TI) with that between the reference model and 10^5 models sampled from the GAN prior (Ref.-Prior). We find that the two distributions match closely. This confirms that images drawn from the GAN prior and from the training set are statistically similar and that none of the images from the training set and prior are likely to be identical to the reference model. This finding is a good indication that the GAN does not seem to have collapsed to a few modes, but it does not exclude the possibility of our generative model having memorized the training set, as in this case we would expect the distributions between Ref.-Prior and Ref.-TI to match.

In a second step we now compare the reference model to the models inferred by our Bayesian inversion approach using the GAN as a prior. We find that the distributions are all consistently shifted to regions of higher similarity to the reference model, i.e., lower MSE and higher SSIM for models inferred when considering the seismic data as well as seismic and well data. This shows that our inversion, when the number of data is increased, tends towards models that are increasingly similar to the reference model. When 9 and 27 acoustic sources are used, we find that inversion leads to models that on average have a SSIM that has very low probability under the Ref.-TI. and Ref.-Prior distributions showing that our GAN is able to create images outside the set of training images. If the generator had only memorized the training set, we should not be able to infer models with higher similarity as the number of data increased.

In Fig. 8a we show models from the training set, samples from the GAN prior, and models inferred with the highest SSIM when compared with the reference case (Fig. 3). In Fig. 8b we show models that have an SSIM close to the mode of the SSIM distributions and find that the model from the posterior inferred by inversion using 27 acoustic sources is visually more similar to the reference case (Fig. 3) than the samples obtained from the prior and from the training set.

It is important to note that the evaluation of the inferred models with respect to a reference model is only possible in the case of synthetic data. In subsurface applications, it is not possible to obtain the entire reference model. Furthermore, models that are structurally very different can be valid solutions of the ill-posed inverse problem. These models, which represent possible solutions of the inverse problem, may be associated with different modes of the prior distribution. In the case of GANs, the gen-

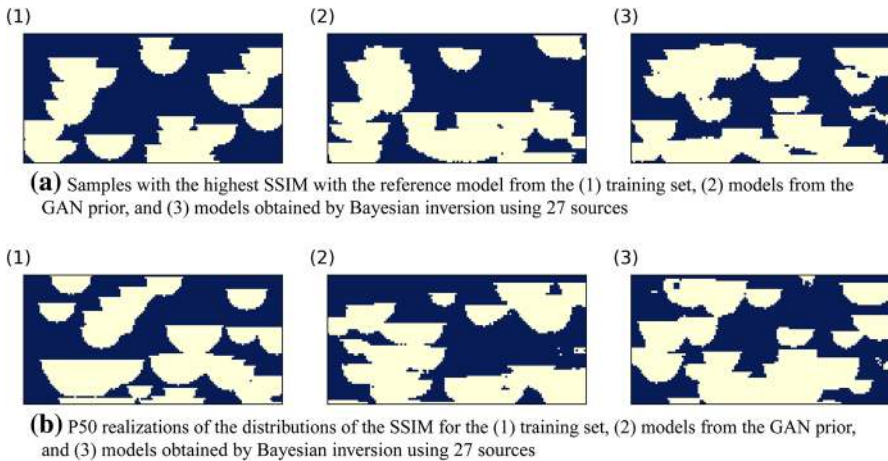


Fig. 8 Overview of models from the training set, GAN prior, and inferred models using the MALA approach that **a** show the highest similarity to the reference model (Fig. 3) measured by the SSIM metric and **b** represent realizations close to the mode of the distribution of the SSIM

erator may be able to represent all of these modes or only a subset (mode-collapse). If mode-collapse has occurred, the posterior ensemble only represents solutions obtained from the modes represented by the generator. Therefore, checking for the occurrence of mode-collapse is key for practical applications, as mode-collapse may significantly affect the ensemble of obtained solutions and possibly lead to underestimated uncertainty.

For future work, evaluating other deep generative models based on explicit density representations (Kingma and Welling 2013; Dinh et al. 2016; van den Oord et al. 2016; Kingma and Dhariwal 2018), which can calculate the likelihood of a set of test images, may help to improve the representation of the prior distribution and mitigate the effect of mode-collapse on inversion.

8 Conclusions

Inversion of subsurface geological heterogeneities from acoustic reflection seismic data is a classical method designed to aid the understanding of the earth's interior. The inference of model parameters from measured acoustic properties is often performed in the very high-dimensional space of model properties, leading to very CPU-intensive optimization (Akcelik et al. 2003).

We apply a method that combines a generative model of geological heterogeneities efficiently parameterized by a lower-dimensional set of latent variables, with a numerical solution of the acoustic inverse problem for seismic inversion using the adjoint method. Leveraging the adjoint of the studied partial differential equation, we deduce gradients that are subsequently used to sample from the posterior over the latent variables given the mismatch of the observed seismic data by following an approximate MALA scheme (Nguyen et al. 2016).

While the proposed application was illustrated on a simple geophysical inversion, this method may find use in other domains where spatial property models control the evolution of physical systems, such as in fluid flow in porous media or materials science. The combination of a deep generative model parameterized by a lower-dimensional set of latent variables and gradients obtained by the adjoint method may lead to new efficient techniques for solving high-dimensional inverse problems.

Acknowledgements O. Dubrule would like to thank Total S.A. for seconding him as visiting professor at Imperial College London.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

A Appendix

A.1 Generative Model Network Architectures

See Table 1 and Figs. 9, 10, 11, 12, 13 and 14.

Table 1 Generator and discriminator network architectures used to create synthetic geological structures. Geological facies are represented by a binary indicator variable and P-wave velocities are represented by a bimodal Gaussian distribution. A hyperbolic tangent activation function is used to represent the two families of properties. Rock density shows a Gaussian distribution. A soft-plus activation function ($f(x) = \frac{1}{\beta} \log(1 + \exp(\beta x))$, $\beta = 1$) is used to ensure positive values of density. Notation for convolutional layers: LayerType (Number of filters), k = kernel size, s = stride, p = padding, BN = BatchNorm, PS = PixelShuffle

<i>(a) Multi-channel generator</i>
Latent variables $z \in \mathbf{R}^{50 \times 1 \times 2}$
Conv2D(512)k3s1p1, BN, ReLU, PSx2
Conv2D(256)k3s1p1, BN, ReLU, PSx2
Conv2D(128)k3s1p1, BN, ReLU, PSx2
Conv2D(64)k3s1p1, BN, ReLU, PSx2
Conv2D(64)k3s1p1, BN, ReLU, PSx2
Conv2D(64)k3s1p1, BN, ReLU, PSx2
Conv2D(3)k3s1p1
Tanh (0,1)—Softplus (2)
<i>(b) Discriminator for multi-channel GAN</i>
Geological properties $m \in \mathbf{R}^{3 \times 64 \times 128}$
Conv2D(64)k5s2p2, ReLU
Conv2D(64)k5s2p1, ReLU
Conv2D(128)k3s2p1, ReLU
Conv2D(256)k3s2p1, ReLU
Conv2D(512)k3s2p1, ReLU
Conv2D(512)k3s2p1, ReLU
Conv2D(1)k3s1p1, ReLU

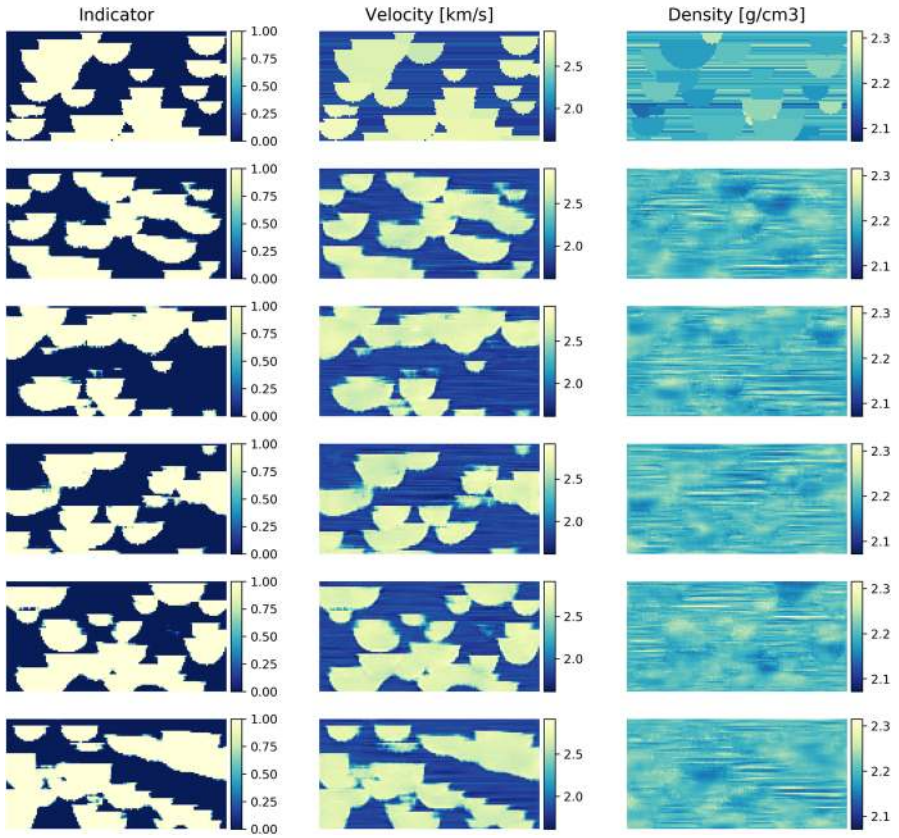


Fig. 9 Samples from the prior distribution of models obtained from the GAN with the reference model (Fig. 3) shown in the first row

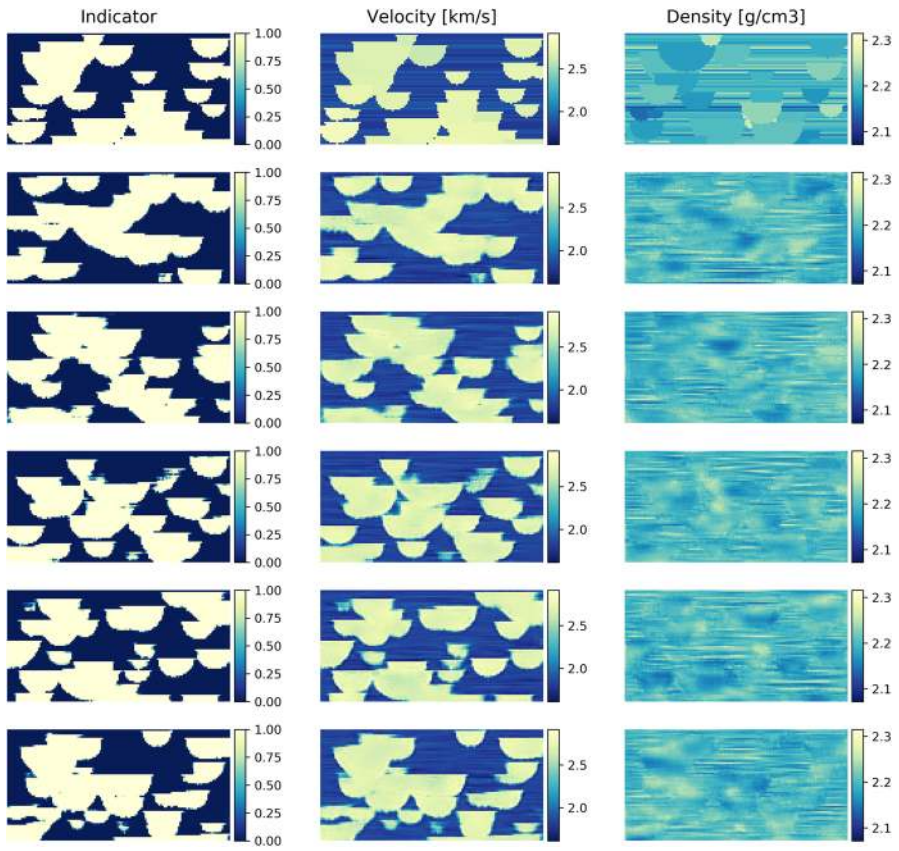


Fig. 10 Samples obtained from latent space optimization with two acoustic sources with the reference model (Fig. 3) shown in the first row

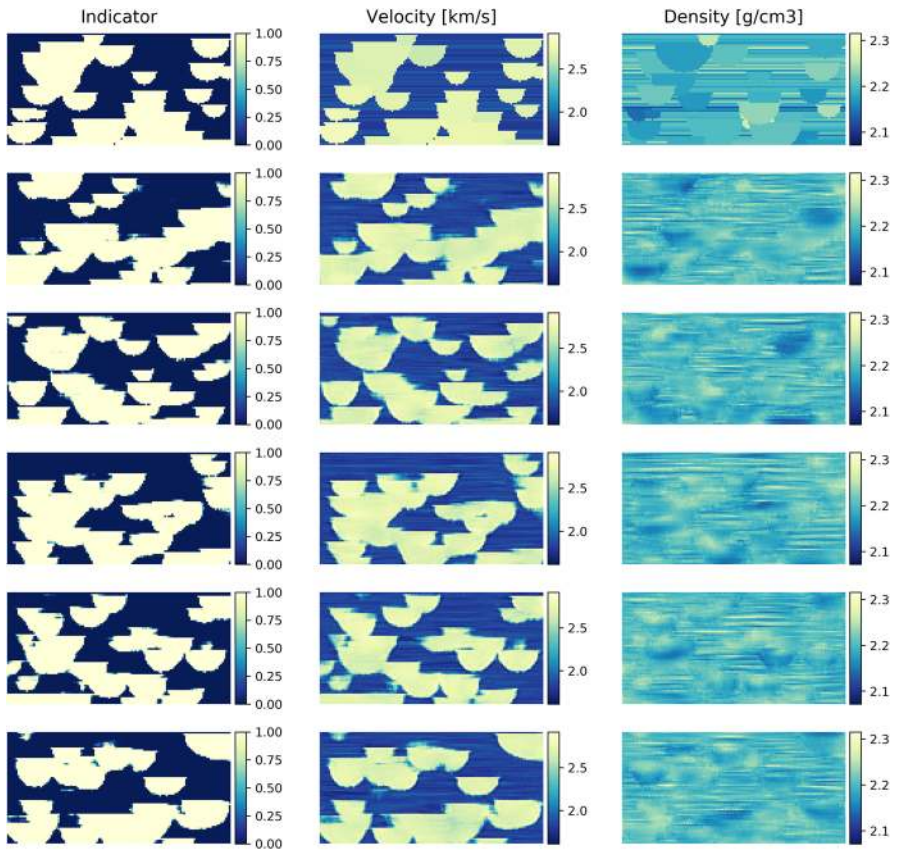


Fig. 11 Samples obtained from latent space optimization with three acoustic sources with the reference model (Fig. 3) shown in the first row

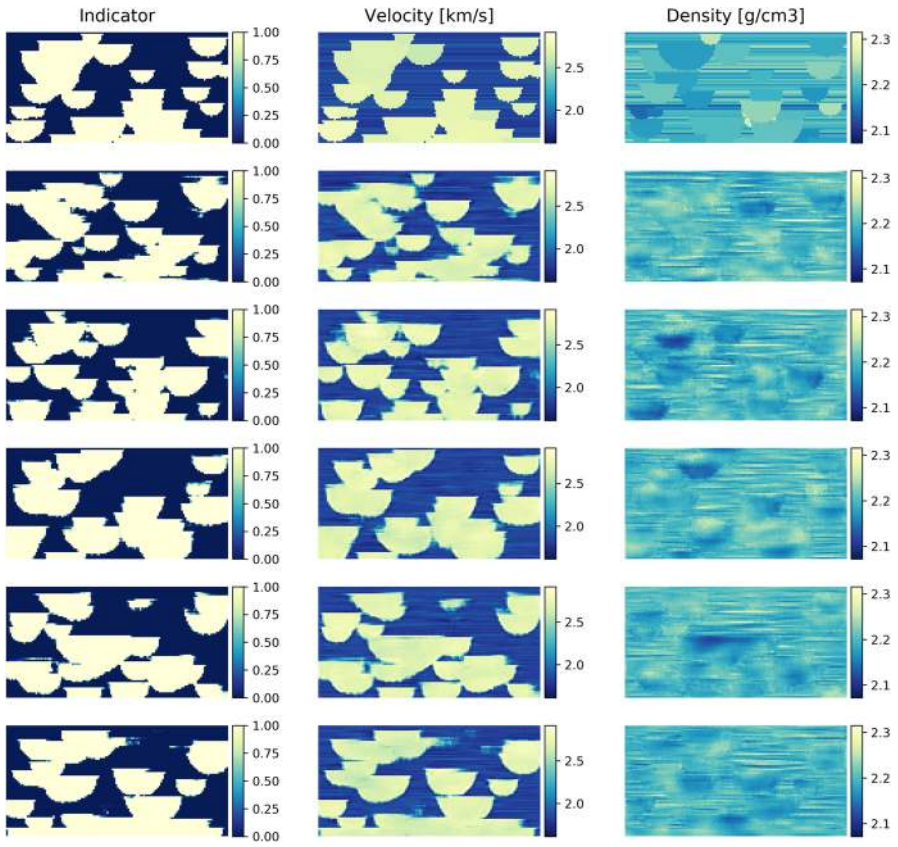


Fig. 12 Samples obtained from latent space optimization with nine acoustic sources with the reference model (Fig. 3) shown in the first row

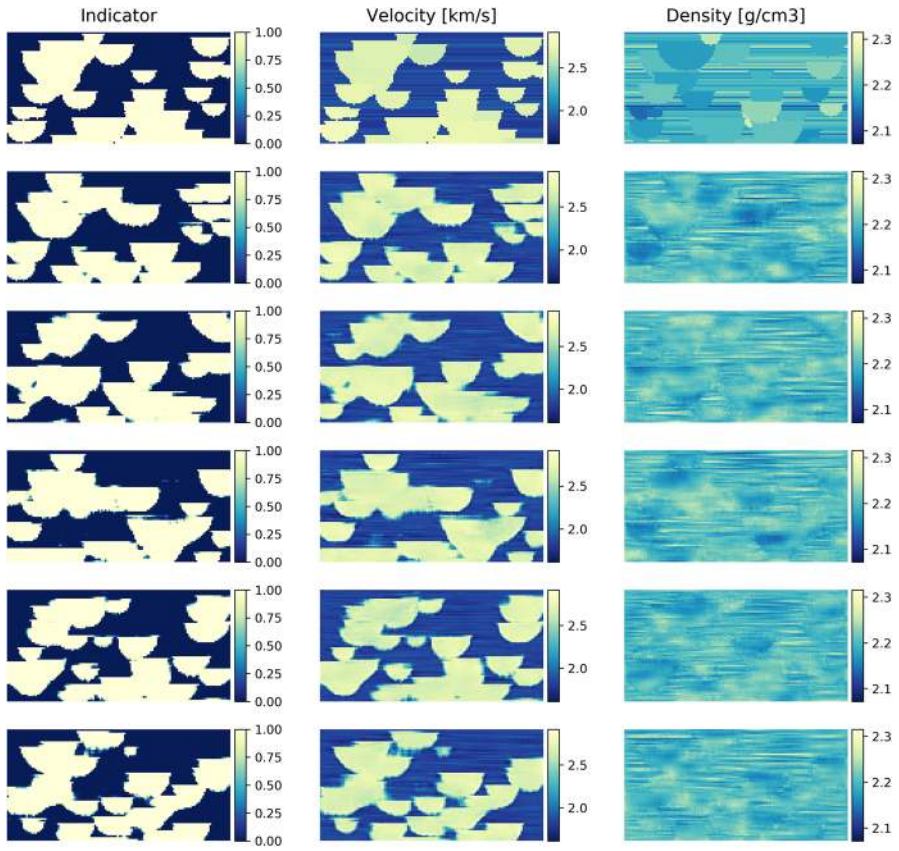


Fig. 13 Samples obtained from latent space optimization with 27 acoustic sources with the reference model (Fig. 3) shown in the first row

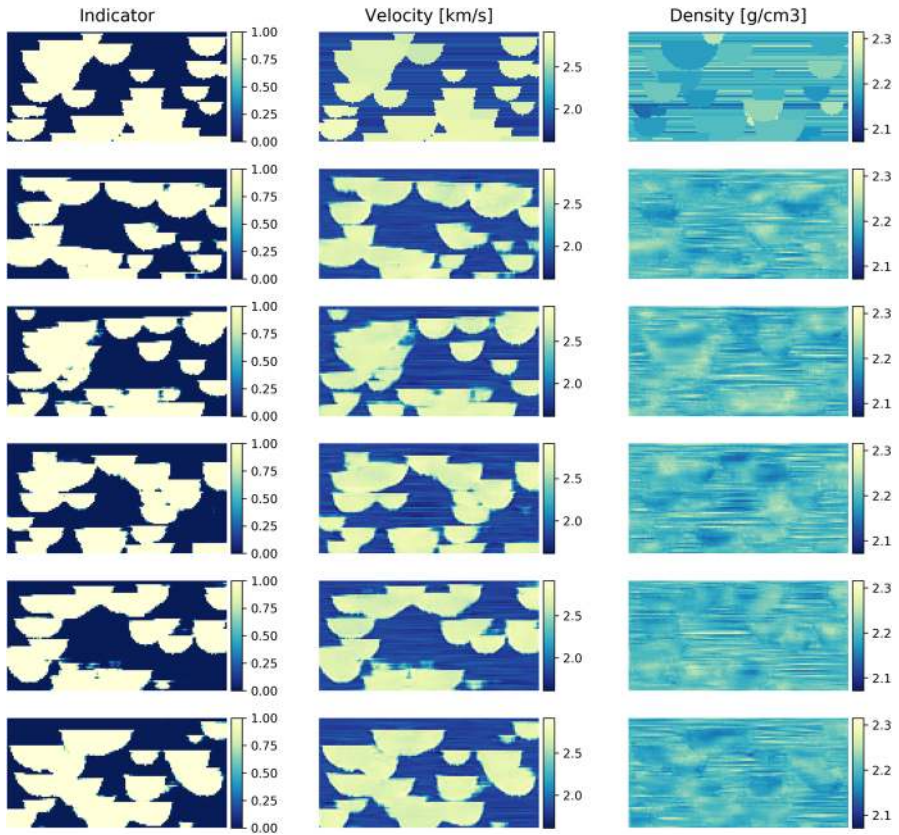


Fig. 14 Samples obtained from latent space optimization with two acoustic sources and one borehole with the reference model (Fig. 3) shown in the first row

References

- Akcelik V, Biros G, Ghattas O (2002) Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation. In: ACM/IEEE 2002 conference on supercomputing. IEEE, pp 41–41
- Akcelik V, Bielik J, Biros G, Epanomeritakis I, Fernandez A, Ghattas O, Kim EJ, Lopez J, O’Hallaron D, Tu T (2003) High resolution forward and inverse earthquake modeling on terascale computers. In: 2003 ACM/IEEE conference on supercomputing. IEEE, pp 52–52
- Araya-Polo M, Jennings J, Adler A, Dahlke T (2018) Deep-learning tomography. *Lead Edge* 37(1):58–66
- Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN. arXiv preprint [arXiv:1701.07875](https://arxiv.org/abs/1701.07875)
- Arora S, Zhang Y (2017) Do GANs actually learn the distribution? An empirical study. arXiv preprint [arXiv:1706.08224](https://arxiv.org/abs/1706.08224)
- Asim M, Shamshad F, Ahmed A (2018) Blind image deconvolution using deep generative priors. arXiv preprint [arXiv:1802.04073](https://arxiv.org/abs/1802.04073)
- Azevedo L, Soares A (2017) Geostatistical methods for reservoir geophysics. Springer, Berlin
- Azevedo L, Grana D, Amaro C (2018) Geostatistical rock physics AVA inversion. *Geophys J Int* 216(3):1728–1739 ISSN 0956-540X
- Barratt S, Sharma R (2018) A note on the inception score. arXiv preprint [arXiv:1801.01973](https://arxiv.org/abs/1801.01973)
- Berg J, Nyström K (2017) Neural network augmented inverse problems for PDEs. arXiv preprint [arXiv:1712.09685](https://arxiv.org/abs/1712.09685)

- Bora A, Jalal A, Price E, Dimakis AG (2017) Compressed sensing using generative models. arXiv preprint [arXiv:1703.03208](https://arxiv.org/abs/1703.03208)
- Bortoli LJ, Alabert F, Haas A, Journel A (1993) Constraining stochastic images to seismic data. In: Soares A (ed) *Geostatistics Tróia'92*. Springer, Dordrecht, pp 325–337
- Bosch M, Carvajal C, Rodrigues J, Torres A, Aldana M, Sierra J (2009) Petrophysical seismic inversion conditioned to well-log data: methods and application to a gas reservoir. *Geophysics* 74(2):1–15
- Bosch M, Mukerji T, Gonzalez EF (2010) Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: a review. *Geophysics* 75(5):75A165–75A176
- Buland A, Omre H (2003) Bayesian linearized AVO inversion. *Geophysics* 68(1):185–198
- Chan S, Elsheikh AH (2017) Parametrization and generation of geological models with generative adversarial networks. arXiv preprint [arXiv:1708.01810](https://arxiv.org/abs/1708.01810)
- Chan S, Elsheikh AH (2018) Parametric generation of conditional geological realizations using generative neural networks. arXiv preprint [arXiv:1807.05207](https://arxiv.org/abs/1807.05207)
- Chang JHR, Li CL, Poczos B, Vijaya Kumar BVK, Sankaranarayanan AC (2017) One network to solve them all—solving linear inverse problems using deep projection models. arXiv preprint [arXiv:1703.09912](https://arxiv.org/abs/1703.09912)
- Creswell A, Bharath AA (2018) Inverting the generator of a generative adversarial network. *IEEE Trans Neural Netw Learn Syst* 30:1967–1974
- Deutsch CV, Wang L (1996) Hierarchical object-based stochastic modeling of fluvial reservoirs. *Math Geol* 28(7):857–880
- Dinh L, Sohl-Dickstein J, Bengio S (2016) Density estimation using real NVP. ArXiv e-prints
- Doyen P (2007) *Seismic reservoir characterization: an earth modelling perspective*, vol 2. EAGE, Houten
- Dubrule O (2003) *Geostatistics for seismic data integration in earth models*. European Association of Geoscientists and Engineers, Houten
- Dupont E, Zhang T, Tilke P, Liang L, Bailey W (2018) Generating realistic geology conditioned on physical measurements with generative adversarial networks. arXiv preprint [arXiv:1802.03065](https://arxiv.org/abs/1802.03065)
- González EF, Mukerji T, Mavko G (2007) Seismic inversion combining rock physics and multiple-point geostatistics. *Geophysics* 73(1):R11–R21
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp 2672–2680
- Grana D, Della Rossa E (2010) Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion. *Geophysics* 75(3):O21–O37
- Grana D, Fjeldstad T, Omre H (2017) Bayesian Gaussian mixture linear inversion for geophysical inverse problems. *Math Geosci* 49(4):493–515
- Guardiano FB, Srivastava RM (1993) Multivariate geostatistics: beyond bivariate moments. In: Soares A (ed) *Geostatistics Troia'92*. Springer, Dordrecht, pp 133–144
- Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A (2017) Improved training of Wasserstein GANs. arXiv preprint [arXiv:1704.00028](https://arxiv.org/abs/1704.00028)
- Haas A, Dubrule O (1994) Geostatistical inversion—a sequential method of stochastic reservoir modelling constrained by seismic data. *First Break* 12(11):561–569
- Hansen TM, Cordua KS (2017) Efficient Monte-Carlo sampling of inverse problems using a neural network-based forward-applied to GPR crosshole traveltime inversion. *Geophys J Int* 211(3):1524–1533
- Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S (2017) GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: *Advances in Neural Information Processing Systems*, pp 6626–6637
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Netw* 2(5):359–366
- Kabanikhin SI (2008) Definitions and examples of inverse and ill-posed problems. *J Inverse Ill Posed Probl* 16(4):317–357
- Kadu A, Van Leeuwen T, Mulder W (2016) A parametric level-set approach for seismic full-waveform inversion. In: *SEG technical program expanded abstracts 2016*. Society of Exploration Geophysicists, pp 1146–1150
- Kingma DP, Dhariwal P (2018) Glow: generative flow with invertible 1×1 convolutions. In: *Advances in neural information processing systems*, pp 10236–10245
- Kingma DP, Welling M (2013) Auto-encoding variational Bayes. arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114)
- Kukreja N, Louboutin M, Vieira F, Luporini F, Lange M, Gorman G (2016) Devito: automated fast finite difference computation. arXiv preprint [arXiv:1608.08658](https://arxiv.org/abs/1608.08658)

- Laloy E, Vrugt JA (2012) High-dimensional posterior exploration of hydrologic models using multiple-try DREAM_(ZS) and high-performance computing. *Water Resour Res* 48(1):W01526. <https://doi.org/10.1029/2011WR010608>
- Laloy E, Héroult R, Jacques D, Linde N (2017) Efficient training-image based geostatistical simulation and inversion using a spatial generative adversarial neural network. arXiv preprint [arXiv:1708.04975](https://arxiv.org/abs/1708.04975)
- LeCun Y, Touresky D, Hinton G, Sejnowski T (1988) A theoretical framework for back-propagation. In: *Proceedings of the 1988 connectionist models summer school*, CMU, Pittsburgh, PA. Morgan Kaufmann, Los Altos, pp 21–28
- Louboutin M, Witte P, Lange M, Kukreja N, Luporini F, Gorman G, Herrmann FJ (2017) Full-waveform inversion, part 1: forward modeling. *Lead Edge* 36(12):1033–1036
- Modrak R, Tromp J (2015) Computational efficiency of full waveform inversion algorithms. *Society of Exploration Geophysicists*, pp 4838–4842
- Mosegaard K (1998) Resolution analysis of general inverse problems through inverse Monte Carlo sampling. *Inverse Probl* 14(3):405
- Mosegaard K, Tarantola A (1995) Monte Carlo sampling of solutions to inverse problems. *J Geophys Res Solid Earth* 100(B7):12431–12447
- Mosser L, Dubrulle O, Blunt MJ (2017) Reconstruction of three-dimensional porous media using generative adversarial neural networks. *Phys Rev E* 96(4):043309
- Mosser L, Dubrulle O, Blunt MJ (2018a) Conditioning of generative adversarial networks for pore and reservoir scale models. In: *80th EAGE conference and exhibition 2018*
- Mosser L, Dubrulle O, Blunt MJ (2018b) Stochastic reconstruction of an oolitic limestone by generative adversarial networks. *Transp Porous Media* 125(1):81–103
- Mosser L, Kimman W, Dramsch J, Purves S, De la Fuente A, Ganssle G (2018c) Rapid seismic domain transfer: seismic velocity inversion and modeling using deep generative neural networks. arXiv preprint [arXiv:1805.08826](https://arxiv.org/abs/1805.08826)
- Nguyen A, Clune J, Bengio Y, Dosovitskiy A, Yosinski J (2016) Plug and play generative networks: conditional iterative generation of images in latent space. arXiv preprint [arXiv:1612.00005](https://arxiv.org/abs/1612.00005)
- Petzka H, Fischer A, Lukovnicov D (2017) On the regularization of Wasserstein GANs. arXiv preprint [arXiv:1709.08894](https://arxiv.org/abs/1709.08894)
- Plessix RE (2006) A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophys J Int* 167(2):495–503
- Richardson A (2018) Generative adversarial networks for model order reduction in seismic full-waveform inversion. arXiv preprint [arXiv:1806.00828](https://arxiv.org/abs/1806.00828)
- Roberts GO, Rosenthal JS (1998) Optimal scaling of discrete approximations to Langevin diffusions. *J R Stat Soc Ser B (Stat Methodol)* 60(1):255–268
- Roberts GO, Tweedie RL (1996) Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* 2(4):341–363
- Röth G, Tarantola A (1994) Neural networks and inversion of seismic data. *J Geophys Res Solid Earth* 99(B4):6753–6768
- Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X (2016) Improved techniques for training GANs. In: *Advances in neural information processing systems*, pp 2226–2234
- Sen MK, Stoffa PL (1996) Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. *Geophys Prospect* 44(2):313–350
- Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. arXiv preprint [arXiv:1609.05158](https://arxiv.org/abs/1609.05158)
- Tarantola A (2005) *Inverse problem theory and methods for model parameter estimation*, vol 89. SIAM, Philadelphia
- Theis L, van den Oord A, Bethge M (2015) A note on the evaluation of generative models. arXiv preprint [arXiv:1511.01844](https://arxiv.org/abs/1511.01844)
- van den Oord A, Kalchbrenner N, Espeholt L, Vinyals O, Graves A, Others (2016) Conditional image generation with PixelCNN decoders. In: *Advances in neural information processing systems*, pp 4790–4798
- Wang Z, Bovik AC (2009) Mean squared error: love it or leave it? a new look at signal fidelity measures. *IEEE Signal Process Mag* 26(1):98–117
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP et al (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612

- Xifara T, Sherlock C, Livingstone S, Byrne S, Girolami M (2013) Langevin diffusions and the Metropolis-adjusted Langevin algorithm. arXiv preprint [arXiv:1309.2983](https://arxiv.org/abs/1309.2983)
- Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint [arXiv:1703.10593](https://arxiv.org/abs/1703.10593)