Geophysical Journal International

Geophys. J. Int. (2023) **235,** 2598–2613
Advance Access publication 2023 September 22
GJI Applied and Marine Geophysics

https://doi.org/10.1093/gji/ggad362

Multimodal variational autoencoder for inverse problems in geophysics: application to a 1-D magnetotelluric problem

Oscar Rodriguez [1,2] Jamie M. Taylor[1,3] and David Pardo[1,2,4]

[1] *Basque Center for Applied Mathematics (BCAM), Alameda Mazarredo* 14, *E*-48009 *Bilbao, Spain. E-mail: orodriguez053@ikasle.ehu.eus*
[2] *Department of Mathematics, University of the Basque Country (UPV/EHU), Barrio Sarriena, E*-48940 *Leioa, Spain*
[3] *Department of Quantitative Methods, CUNEF Universidad, C/ de los Pirineos* 55, 28040 *Madrid, Spain*
[4] *Basque Foundation for Science (Ikerbasque), Plaza Euskadi* 5, *E*-48009 *Bilbao, Spain*

Accepted 2023 September 9. Received 2023 September 5; in original form 2022 November 24

SUMMARY

Estimating subsurface properties from geophysical measurements is a common inverse problem. Several Bayesian methods currently aim to find the solution to a geophysical inverse problem and quantify its uncertainty. However, most geophysical applications exhibit more than one plausible solution. Here, we propose a multimodal variational autoencoder model that employs a mixture of truncated Gaussian densities to provide multiple solutions, along with their probability of occurrence and a quantification of their uncertainty. This autoencoder is assembled with an encoder and a decoder, where the first one provides a mixture of truncated Gaussian densities from a neural network, and the second is the numerical solution of the forward problem given by the geophysical approach. The proposed method is illustrated with a 1-D magnetotelluric inverse problem and recovers multiple plausible solutions with different uncertainty quantification maps and probabilities that are in agreement with known physical observations.

Key words: Magnetotellurics; Inverse theory; Numerical modelling; Probabilistic forecasting; Statistical methods; Variational autoencoder; Multimodal Models.

1 INTRODUCTION

Advances in geophysics allow us to estimate subsurface properties from geophysical measurements. With this, it is possible to extract oil, gas and water more efficiently and solve other geophysical problems: CO_2 and hydrogen storage, geothermal energy production, etc. Geophysical observations usually result from (approximately) known physical processes that can be excited naturally or artificially. Different measurements systems include tomography (Arridge & Schotland 2009), magnetotellurics (MT, Pace *et al.* 2019), ground-penetrating radar (Travassos *et al.* 2020) and seismic measurements (Zhang & Lin 2020). Estimating the internal subsurface properties from these measurements is often performed by solving an inverse problem governed by known physical laws (e.g. Maxwell's equations, Kolesnikov & Fedin 2018; Berliner 2003). The corresponding inverse solutions are often non-unique (Michel *et al.* 2020), and measurements are usually contaminated with noise.

Mathematically, given the subsurface properties, denoted by x, we may represent the recorded measurements y as the image of a function $\mathcal{F}(x)$ representing the physical laws governing the process. In the inverse problem, we estimate the subsurface properties x from the measurements y (Aster *et al.* 2018; Vogel 2002; Kaipio & Somersalo 2006). Different deterministic techniques exist to solve inverse problems (Agranovich & Marchenko 2020; Lai & Lin 2022; Sidky *et al.* 2020); they find a solution to the inverse problem but often fall short of providing a realistic uncertainty quantification. Uncertainty quantification maps are of utmost importance in geophysical problems, since distant solutions can provide identical measurements up to noise precision. In particular, it is often the case that rocks estimated to have a high content of hydrocarbons are indeed dry. These misestimations provoke huge operational costs.

Probabilistic techniques produce solutions to inverse problems and also quantify the uncertainty (Fichtner *et al.* 2019; Mandolesi *et al.* 2018; Kaipio & Somersalo 2006; Rammay *et al.* 2022a; Jahani *et al.* 2022). In particular, Bayesian inference provides solutions using expert knowledge and sometimes decreases the computational time compared to classical probabilistic models (Calvetti & Somersalo 2018; Yan & Zhou 2019; Xia & Zabaras 2022; Capistrán *et al.* 2021). This offers advantages when solving inverse problems, as the large amount of data necessary to train an inverse operator can incur large computational costs. However, classical Bayesian methods solve the inverse problem for a given set of measurements instead of building an inverse operator, which solves the inverse problem for all possibilities of random variables. It often prevents its practical use in real-time inversion when the measurements are not associated with the estimated parameter.

Neural networks (NNs, Wang *et al.* 2020; Raissi *et al.* 2019) have been increasingly employed in different fields (Shahriari *et al.*

2020; Pilozzi *et al.* 2018; Häggström *et al.* 2019), including geophysical applications (Van der Baan & Jutten 2000; Rammay *et al.* 2022b). Given a well-defined non-injective forward operator, the inverse operator is, in general, non-unique (or not well-defined). One key advantage of NNs for solving inverse problems is that they often construct a full inverse operator, where we refer to one of the inverse branches as 'a full inverse operator', which is what the NN approximates during the training phase. The training set is given by measurements generated by the application of the forward problem using different geophysics properties. After that, NNs can rapidly evaluate the inverse solution of unseen measurements in a fraction of a second (Livingstone *et al.* 1997; Shepherd 2012; Chen & Manning 2014). Some NN schemes that can be used to quantify the uncertainty of the inverse solutions from noisy measurements (Olierook *et al.* 2021; Hermans *et al.* 2018; Malinverno & Briggs 2004) are generative adversarial networks (GANs, Alyaev *et al.* 2021; Fossum *et al.* 2022) and variational autoencoders (VAEs, Liu *et al.* 2022; Goh *et al.* 2021). While NNs for deterministic methods typically output a single, uniquely defined solution, with GANs and VAEs, we take the output of an NN to parametrize a probability density function (PDF). GANs can be used to artificially produce new data from a well-trained probability density (generator) to mitigate the cost of producing real data via a classification function (discriminator) (Creswell *et al.* 2018). VAEs have an encoder–decoder scheme, and they can be used as probabilistic models to find a probability density using variational inference (Tomczak & Welling 2018; Guo *et al.* 2020, 2018). For example, Goh *et al.* (2021) employ VAEs, where the solution to the inverse problem is a Gaussian density (unimodal model). The mean of this density is a point estimate and can be interpreted as the most plausible solution to the inverse problem; the variance represents the quantification of the uncertainty around the estimated solution (the mean).

In geophysics, we often have more than one solution. Thus, an unimodal Gaussian density cannot describe the set of all possible solutions to the inverse problem. Therefore, a model that quantifies uncertainty and returns multiple solutions with an acceptance probability is necessary. We can use multimodal models implemented in geophysical applications (Grana *et al.* 2017; Astic & Oldenburg 2019; de Figueiredo *et al.* 2019) and estimate a multimodal distribution with multiple modes (peaks) and a variance around each mode. Moreover, we can combine multimodal models (e.g. truncated Gaussian mixture densities) and NNs to estimate multiple solutions associated with an inverse problem. For example, Bishop (1994) combines Gaussian mixture densities with NNs and denotes the resulting method as mixture density network (MDN). However, an MDN (Alyaev & Elsheikh 2022; Zhang & Curtis 2021; Earp *et al.* 2020; Meier *et al.* 2007) estimates the inverse problem without considering the physical laws that govern the forward problem. Alternatively, VAEs and GANs can consider the inverse and forward problems in the model scheme, and they have been applied in combination with MDNs in Śmieja *et al.* (2020) and Oikarinen *et al.* (2021). These methods have been implemented chiefly for clustering and classification. Consequently, the computations of internal functions such as encoder, decoder, prior and posterior, among others, are different from those required in a geophysical application.

Inverse problems are ill-posed and often have multiple solutions. Thus, it is desirable to obtain several of them. However, a question arising from this problem is how we obtain more than one solution and which solutions are correct if they differ. To overcome this problem, authors like Alyaev & Elsheikh (2022) and Meier *et al.* (2007) employ an MDN introduced by Bishop (1994) as a multiple prediction model, which shows an improvement in the classical architecture of MDNs, and adopt a multimodal trajectory predictions (MTP) loss function (Cui *et al.* 2019) to solve the mode collapse present in some application when we use MDNs. Other investigations, such as Zhang & Curtis (2021) and Earp *et al.* (2020), combine MDNs with Bayesian statistics to improve the estimations. However, Earp *et al.* (2020) conclude that the prior distribution considerably affects these estimations: an incorrect selection of the prior distribution deteriorates the multimodality of the MDN.

The main contribution of this work is to design a multimodal NN using variational inference, which provides multiple solutions for geophysical inverse problems together with probability distributions and uncertainty quantification for each solution. To do this, we proposed a multimodal variational autoencoder (MVAE) that consists of two components: an encoder and a decoder. The encoder uses an NN to return the parameters of a PDF. This PDF provides a set of multiple solutions and also quantifies the associated uncertainties. The decoder solves the forward problem by taking the encoder's solutions and producing the corresponding numerical solutions to the geophysical formulation.

The proposed MVAE combines an MDN with the VAE framework to achieve multimodal capability. The MDN allows for modelling the probability distribution of the solutions, capturing multiple modes and their associated probabilities. The VAE framework provides a structured way to train and generate samples from the multimodal model. In contrast to previous works (Park *et al.* 2018; Xu *et al.* 2021), we assume a known prior distribution and approximate the posterior distribution with a mixture of truncated Gaussian densities via variational inference. To illustrate the resulting method, we first apply it to a simple inverse problem with two solution branches. Then, we apply it to a 1-D magnetotelluric (MT) inverse problem (Grandis *et al.* 1999; Medin 2008). We employ synthetic data generated by computer algorithms to produce a sufficiently large data set that ensures the model is well trained and assumes known measurement noise and prior. The algorithms are executed in Python using the TensorFlow, TensorFlow Probability and NumPy libraries. Extending the above method to 2-D and 3-D problems is straightforward, although this leads to a significantly increased computational cost of the corresponding forward problem (Alvarez-Aramberri & Pardo 2017; Spichak & Popova 2000).

The organization of this paper is as follows. Section 2 introduces some preliminary concepts and definitions. Section 3 presents the motivation for using a mixture of densities by considering a simple inverse problem. Section 4 proposes the MVAE that provides multiple solutions to an inverse problem. Section 5 presents an application of MVAE to a 1-D MT inverse problem. We summarize our conclusions in Section 6.

2 PRELIMINARY DEFINITIONS AND ASSUMPTIONS

This section introduces some preliminary concepts about NNs and probability needed to explain the BVAEs models and the proposed loss functions $\mathscr{L}_\theta$.

2.1 Neural networks

In classical supervised learning, given physical measurements y and properties x, an NN is a model that approximates a function from y to x through a composition of several mappings called layers (Hornik *et al.* 1989; Samek *et al.* 2021) Let s be the number of

layers considered in the model,

$$\mathcal{I}_\theta(y) = \left(\mathcal{L}^{(s)} \circ \mathcal{L}^{(s-1)} \circ \cdots \circ \mathcal{L}^{(1)}\right)(y), \tag{1}$$

where $\mathcal{L}^{(i)}(y) = f_i\left(\mathcal{W}^{(i)} \cdot y + b^{(i)}\right)$, $\mathcal{W}^{(i)}$ is a matrix, $b^{(i)}$ is a vector and f_i is the ith layer activation function for $i = 1,\ldots, s$. Each layer $\mathcal{L}^{(i)}(y)$ contains m_i nodes $\{z_0^{(i)}, ..., z_{m_i}^{(i)}\}$ defined as $z_j^{(i)}(y) = f_i(w_j^{(i)} y + b^{(i)})$. Fig. 1 shows the fully connected NN scheme with s layers. The loss function $\mathscr{L}_\theta(x, y)$ is a function that minimizes the error between the predicted output and the *ground truth* x in some norm.

However, the approximation generated by conventional deep learning methods is sometimes inadequate to solve inverse geophysical problems where the solution is non-unique. In these cases, it is necessary to implement strategies that provide adequate solutions to the inverse problem (see Shahriari *et al.* 2020).

2.2 Preliminary statistical concepts

Let X be a random variable with a PDF $p(x)$. We write the *expected value* of $g(X)$ as (see Gut 2013):

$$\mathbb{E}_p\left[g(X)\right] = \int g(x)p(x)\mathrm{d}x. \tag{2}$$

Eq. (2) can be approximated by:

$$\mathbb{E}_p\left[g(X)\right] \approx \frac{1}{H}\sum_{i=1}^{H} g(x^i), \tag{3}$$

where x^i are samples distributed according to $p(x)$ for $i = 1,\ldots, H$. We will consider truncated Gaussian PDFs characterized by their mean μ and variance σ^2 and domain $(a, b) \subset \mathbb{R}$ with $a < b$. We denote these PDFs by $p(x; \varphi, a, b)$, where $\varphi = (\mu, \sigma)$.

2.2.1 Truncated Gaussian mixture density

Let X be a random variable with PDF $p(x; \varphi)$ given by:

$$p(x; \varphi, a, b) = \sum_{m=1}^{M} \pi_m \cdot p^m(x; \varphi_m, a, b), \tag{4}$$

where $\pi_i \geq 0$, $\sum_{i=1}^{M} \pi_i = 1$, and $p^i(x; \varphi_i, a, b)$ with $\varphi_i = (\mu_i, \sigma_i)$ is a truncated Gaussian density with mean μ_i, variance σ_i for $m = 1,\ldots, M$. The PDF $p(x; \varphi, a, b)$ is called truncated Gaussian mixture density with M densities and parameters $\varphi = (\pi_1, \mu_1, \sigma_1,\ldots, \pi_M, \mu_M, \sigma_M)$. We take

$$p^m(x; \mu_m, \sigma_m, a, b) = \frac{1}{Z_m}\exp\left(-\frac{(x-\mu_m)^2}{2\sigma_m^2}\right), \tag{5}$$

when $x \in (a, b)$ and $p^m = 0$ otherwise, where Z_m is an appropriate normalizing constant to ensure that p^m integrates to 1.

2.2.2 Bayes' theorem in inverse problems

Bayes' theorem acts as a bridge between the probabilistic forward and inverse problem (Kaipio & Somersalo 2006; Gut 2013). Given X, Y and ε random variables, we assume observations of geophysical measurements from a mathematical forward problem $\mathcal{F}$ with an additive noise such that $Y = \mathcal{F}(X) + \varepsilon$, that is, the probability occurrence of physical measurements Y given subsurface properties X is described by a conditional PDF $p(y|x)$ known as *likelihood*, and contains information about the uncertainty present in observations associated with the forward problem. Similarly, the uncertainty associated with the inverse problem $\mathcal{I}(Y)$ is given by a PDF $p(x|y)$ known as *posterior* corresponding to the inverse problem. By Bayes' theorem, we obtain:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \tag{6}$$

where $p(x)$ is a *prior* knowledge (Kaipio & Somersalo 2006; Goh *et al.* 2021). In the context of our work, this corresponds to the probability distribution describing the ground truth in the absence of further information. The choice of the prior is a modelling assumption that, according to eq. (6), has a clear effect on the posterior. The prior may be informed by expert knowledge and data, and when limited information is available, one may turn to simple distributions such as Gaussian or lognormal distributions, where we assume known qualitative properties of the ground truth. The PDF $p(y)$ corresponds to all possible measurements, given by:

$$p(y) = \int p(y|x)p(x)\mathrm{d}x. \tag{7}$$

2.3 Conditional PDF parameter estimation by NNs

Given X and Y random variables, we approximate the PDF $p(x|y)$ by the PDF $p(x; \varphi_\theta(y))$, where the objective is to estimate the parameters $\varphi_\theta(y)$ by an NN. The function $\varphi_\theta(y)$ depends upon the trainable parameters $\theta = (\theta^{(1)}, \theta^{(2)},\ldots, \theta^{(s)})$, where $\theta^{(i)} = (\mathcal{W}^{(i)}, b^{(i)})$ for $i = 1,\ldots, s$ (Wang 1994; Neuneier *et al.* 1994; Mao *et al.* 2000). Fig. 2 shows the scheme to estimate a PDF by an NN, where the loss function $\mathscr{L}_\theta(x, y)$ is typically defined from the PDF $p(x; \varphi_\theta(y))$ and expressed as:

$$\mathscr{L}_\theta(x, y) = -\log p(x; \varphi_\theta(y)). \tag{8}$$

The NNs in this section are commonly used for predictions or classification like linear or logistic regression. However, in inverse problems, there may be occasions when a more elaborate model is needed. For example, a model in which we have information about the origin of the data. In this way, we present the autoencoder NNs in the next section.

3 VARIATIONAL AUTOENCODER FROM GAUSSIAN NOISE

Autoencoders have been used for applications such as dimensionality reduction or image reconstruction (Goh *et al.* 2021; Liu *et al.* 2022; Guo *et al.* 2020; Häggström *et al.* 2019; Kingma & Welling 2013). Given $D = (x, y)$, we have an encoder part (input model, which is our inverse problem) with an input y, and we obtain an estimation for x. In the decoder (output model, which is our forward problem), given an input x estimated by the encoder, we obtain an estimation for y.

Currently, it is possible to use probabilistic and deterministic versions, see Appendix A), as well as a probabilistic one. Herein, we focus on the solution of inverse problems with uncertainty quantification, and therefore, we describe the construction of an autoencoder scheme that allows us to quantify the uncertainty of the inverse problem (Kingma & Welling 2013).

We use an NN to estimate the probabilistic inverse problem solution determined by a posterior distribution of the random variable X given the random variable Y. First, we construct the likelihood $p(y|x)$ from the distribution of additive noise ε. We consider $Y = \mathcal{F}(X) + \varepsilon$, where ε follows a Gaussian density $p(\varepsilon; \varphi_\varepsilon)$ with

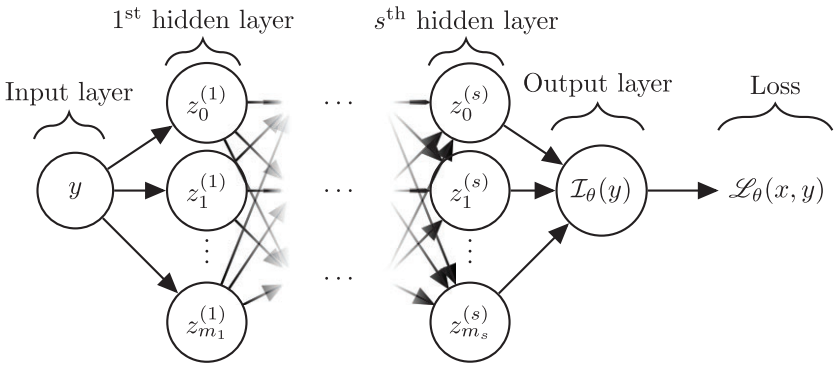


Figure 1. Fully connected NN scheme.

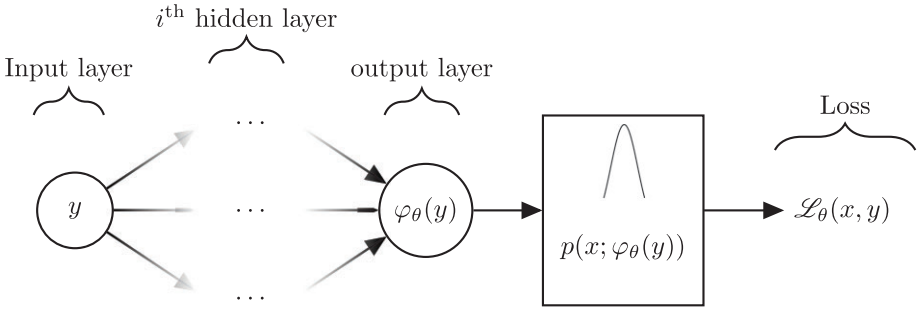


Figure 2. Density NN scheme.

mean vector $\mathbf{0}_k$ and diagonal covariance matrix Γ_k, where $\varphi_\varepsilon = (\mathbf{0}_k, \Gamma_k)$. The likelihood $p(y|x;\varphi_\varepsilon)$ is expressed by (Goh *et al.* 2021):

$$p(y|x;\varphi_\varepsilon) = (2\pi)^{-\frac{k}{2}}|\Gamma_k|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(y-\mathcal{F}(x))^t\,\Gamma_k^{-1}\,(y-\mathcal{F}(x))\right). \tag{9}$$

Our interest in using NNs with probabilities is to quantify uncertainty by using a posterior PDF $p(x|y)$. Using Monte Carlo methods to obtain samples from $p(x|y)$ can be computationally expensive in Bayesian inverse problems. Furthermore, whilst Bayes' theorem gives us an explicit expression for the posterior $p(x|y)$, the modes of the distribution are not easily identifiable, and thus, we cannot make point estimations for solutions to the inverse problem. To solve this, we approximate the posterior distribution $p(x|y)$ (usually unknown and non-parametric PDF) with a more tractable parametric previously selected PDF $q(x;\varphi_\theta(y))$, where the parameters $\varphi_\theta(y)$ of the distribution are given by an NN with weights θ. The parameters $\varphi_\theta(y)$ of the posterior are then simple to interpret in the context of the inverse problem. To estimate θ, we use the evidence lower bound (ELBO) loss function (Blundell *et al.* 2015) given by:

$$\theta\mathrm{KL} = \underset{\theta}{\operatorname{argmin}} \int p(y)\cdot\mathscr{L}_\theta^{\mathrm{KL}}(y)\mathrm{d}y, \tag{10}$$

$$\approx \frac{1}{H\cdot N}\cdot\underset{\theta}{\operatorname{argmin}}\sum_{n=1}^{N}\sum_{i=1}^{H}\left[\log q(x_n^i;\varphi_\theta(y_n)) - \left(\log p(x_n^i) + \log p(y_n|x_n^i)\right)\right]. \tag{11}$$

The goal is to measure the discrepancy between the posterior distribution $p(x|y)\propto p(x)p(y|x)$ and the proposed distribution $q(x|\varphi_\theta(y))$. Minimizing this discrepancy helps us shape and scale the proposed distribution with known parameters $\varphi_\theta(y)$ towards the posterior distribution. To do this, we use a Monte Carlo approximation, and the complete derivation is in Appendix B. As previously commented in Section 2.2, the choice of prior affects the posterior, which is reflected in eq. (11) via the dependence on $p(x)$.

Fig. 3 shows how to obtain $q(x;\varphi_\theta(y))$ and construct the loss function.

4 MULTIMODAL VARIATIONAL AUTOENCODER

In this work, we consider $q(x;\varphi_\theta(y))$ in eq. (11) (see Appendix B) as an MDN, where parameters $\varphi_\theta(y)$ are given by the M means $\mu_\theta^m(y)$—point estimates to the solution of the inverse problem—the M standard deviations $\sigma_\theta^m(y)$—uncertainty quantification of each solution—and the M probabilities $\pi_\theta^m(y)$ of occurrence associated with each solution for $m = 1,\ldots,M$. Fig. 4 shows a general form of the MVAE scheme. Unlike the works cited in the previous paragraph, in which they estimate the parameters of the MDN by using maximum likelihood, we propose using the Kullback–Leibler divergence to approximate the posterior distribution of an MDN.

Eq. (11) describes the loss function used for MVAE, where q (a Gaussian mixture density) is a more tractable density than the posterior distribution, represented by the sum of the likelihood and the prior. The samples obtained from the latent space q are evaluated in the posterior distribution. Therefore, the modes with the highest peak will have higher values in the evaluation of the posterior distribution, and q will approximate that same shape, giving a higher weight π^m to those modes.

Note that the negative log-likelihood (NLL) loss function quantifies the negative log-probability of the observed data given a PDF, maximizing it where the data have a higher probability of occurrence. On the other hand, the ELBO loss function is often used in variational inference and seeks to approximate the posterior distribution. It measures the discrepancy between the posterior distribution with a more tractable PDF. Maximizing the ELBO corresponds to minimizing the divergence between the two distributions.

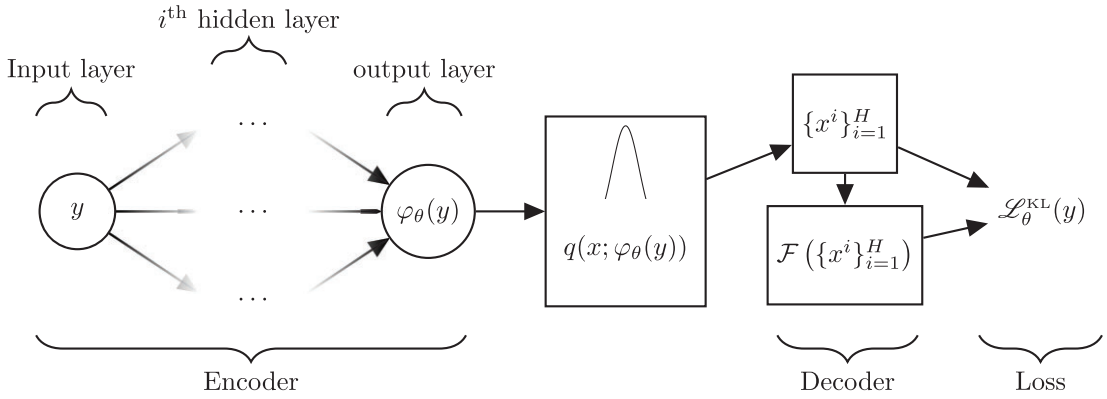


Figure 3. VAE scheme.

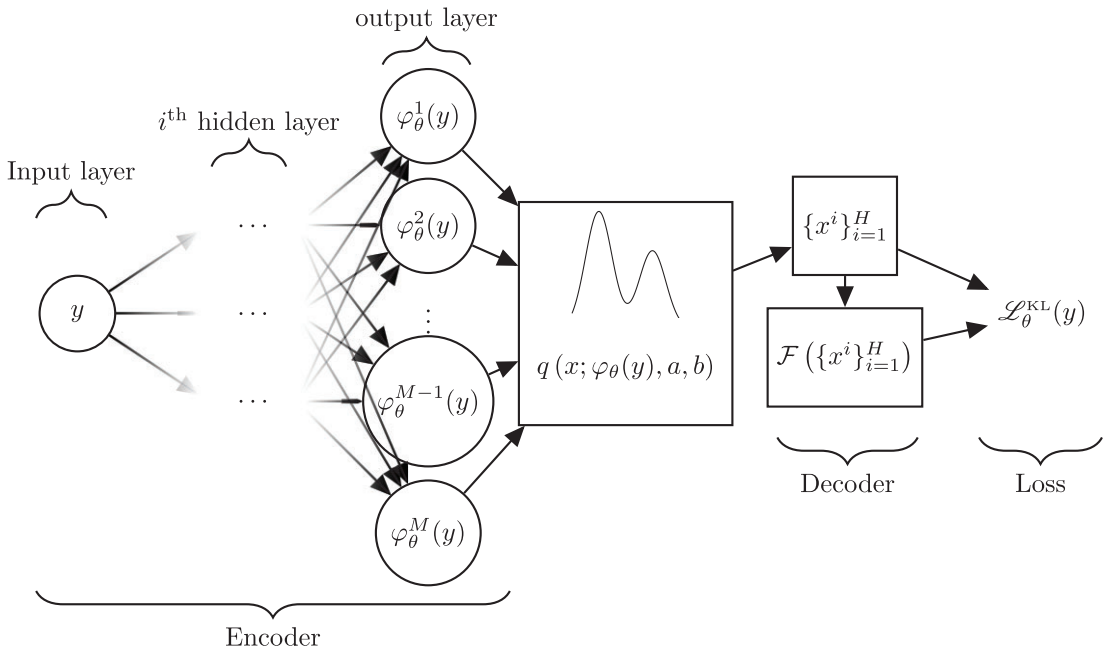


Figure 4. MVAE scheme of M truncated Gaussian densities $q^m(x;\varphi_\theta^m(y),a,b)$ that define $q(x;\varphi_\theta(y),a,b)=\sum_{m=1}^M \pi_\theta^m(y)q^m(x;\varphi_\theta^m(y),a,b)$, where $\sum_{j=1}^M \pi_\theta^m(y)=1$.

The NLL loss function is often employed in supervised learning (Schapire 2003; Farnia & Tse 2016), while the ELBO loss function is more popular in generative models, such as VAEs (Tomczak & Welling 2018; Hoffman & Johnson 2016), where the goal is to learn latent representations and generate new samples.

5 APPLICATION TO THE MAGNETOTELLURIC INVERSE PROBLEM

The MT method is a passive electromagnetic exploration technique governed by Maxwell's equations. The source is naturally generated within the ionosphere, and it is modelled in the form of a plane wave solution that arrives at the Earth. Based on the impedance measured at different locations along the surface, we can recover a resistivity image of the Earth's subsurface by solving the corresponding inverse problem. In this paper, we employ the proposed MVAE scheme to solve a 1-D MT inverse problem. Fig. 5 shows an example of 1-D MT exploration for a fixed set of subsurface layers

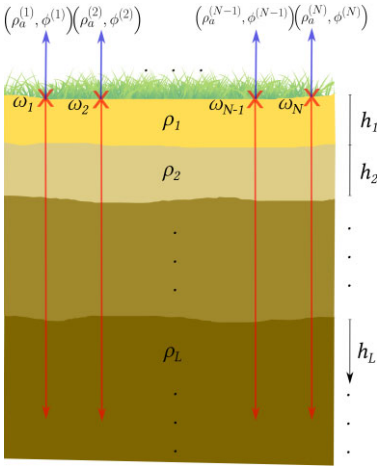
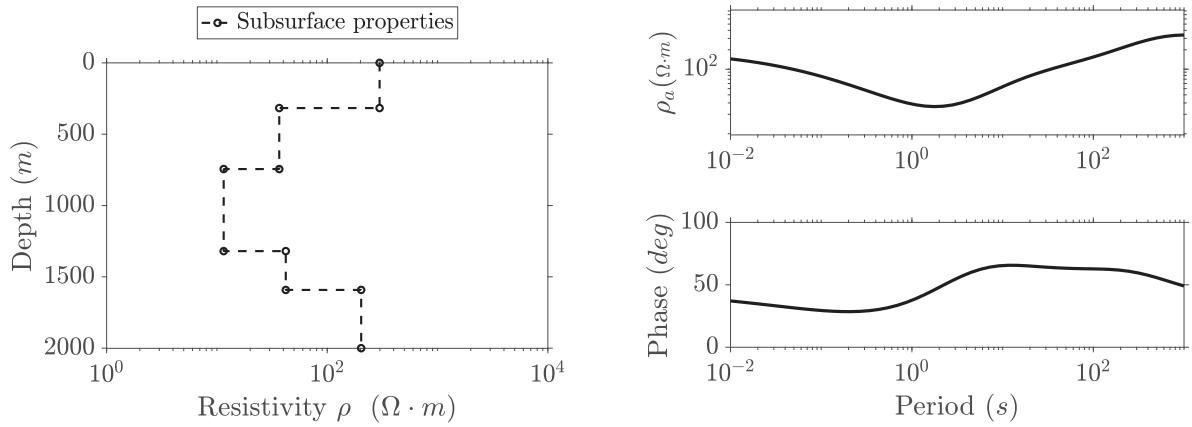


Figure 5. 1-D MT exploration.



(a) Subsurface layers with associated resistivities (x_{MT}) (b) Apparent resistivities and phases for each period (y_{MT})

Figure 6. Example of an MT forward problem input and output.

Table 1. Architecture for MVAE, where the input is the apparent resistivity and phase $[\rho_a, \phi]$, and output is a mixture PDF of the resistivity ρ. Constants N, M and L are the number of frequencies, mixtures and layers, respectively.

	MVAE	
Layer (type)	Output shape	Activation function
Input non-trainable layer (ρ_a, ϕ)	(None,$2 \cdot N$)	*linear*
Hidden dense layer	(None,300)	*tanh*
Hidden dense layer	(None,300)	*softplus*
Output dense layer	(None,$M \cdot (2 \cdot L + 1)$)	*linear*
Reshape non-trainable layer	(None,M,$2 \cdot L + 1$)	*linear, softplus, softmax*

with associated thicknesses and resistivities. We position one electrode on the surface operating at N different frequencies and obtain as response N impedances from which we calculate their apparent resistivities and phases.

To find the solution to the MT inverse problem, we briefly explain the forward problem's formulation, define the MT inverse problem, and apply the MVAE model.

5.1 1-D MT forward problem

The analytical solution for the forward problem is:

$$\mathcal{Z}(\rho, \omega) = \frac{E(\rho, \omega)}{H(\rho, \omega)}, \tag{12}$$

$$\rho_a = \frac{1}{\mu_0 \omega} |\mathcal{Z}|^2, \tag{13}$$

$$\phi = \arctan\left(\frac{\mathrm{Re}(\mathcal{Z})}{\mathrm{Im}(\mathcal{Z})}\right). \tag{14}$$

$\mathcal{Z}$ is the impedance obtained from the electric and magnetic fields E and H, respectively, generated by the interaction between the subsurface resistivity ρ and the applied plane wave source with frequency ω. From the obtained impedance, we generate the output of the forward problem given by the apparent resistivity ρ_a and the phase difference ϕ. If we further assume that ρ as a piecewise-constant function in a domain $(0, h_1) \cup (h_1, h_2) \cup \cdots \cup (h_{L-1}, \infty)$ with $h_i < h_{i+1}$ for $i = 1, \ldots, L-1$, the MT forward problem has a simple analytical solution, see Mandolesi *et al.* (2018), Medin (2008) and Parker (1983). For a set of N fixed frequencies $\omega = (\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(N)})$ and L layers, the input and output of our MT forward problem are given by:

(i) Input:

–$L-1$ subsurface thickness layers $h = (h_1, h_2, \ldots, h_{L-1})$,
–L resistivities $\rho = (\rho^{(1)}, \rho^{(2)}, \ldots, \rho^{(L)})$.

(ii) Output:
(iii) N apparent resistivities $\rho_a = \left[\rho_a^{(1)}, \rho_a^{(2)}, \ldots, \rho_a^{(N)}\right]$.
(iv) N phases $\phi = [\phi^{(1)}, \phi^{(2)}, \ldots, \phi^{(N)}]$.

We denote the MT forward problem $\mathcal{F}^{\mathrm{MT}}$ as:

$$\mathcal{F}^{\mathrm{MT}}(x_{\mathrm{MT}}) = y_{\mathrm{MT}}, \tag{15}$$

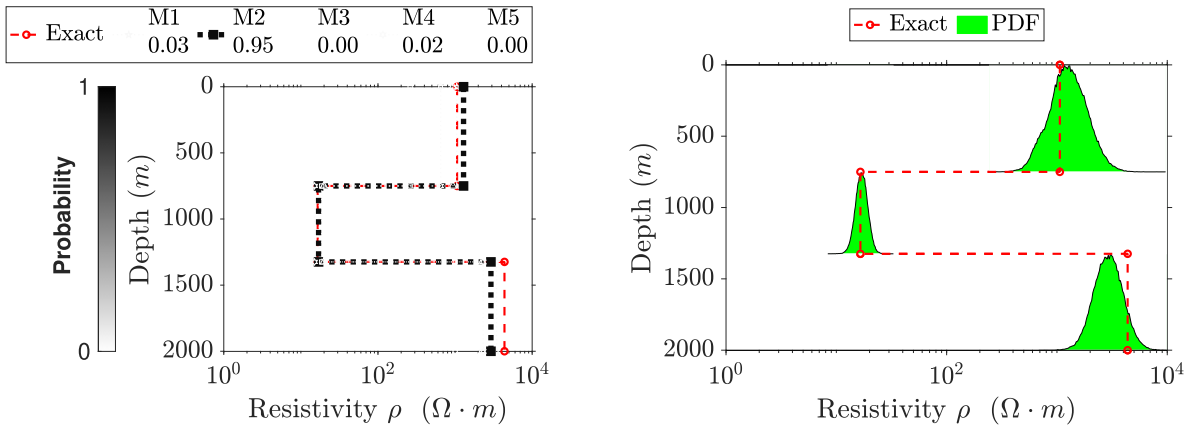
where $x_{\mathrm{MT}} = [\rho, h]$ is the set of subsurface properties, and the MT forward solution $y_{\mathrm{MT}} = [\rho_a, \phi]$ is the set of physical measurements on the surface.

Fig. 6 shows a graphic representation of an MT forward problem input and output. The input x_{MT} is displayed in Fig. 6(a). The output displayed in Fig. 6(b) represents the logarithm of the apparent resistivity and the phase as a function of the period.

5.1.1 1-D inverse problem

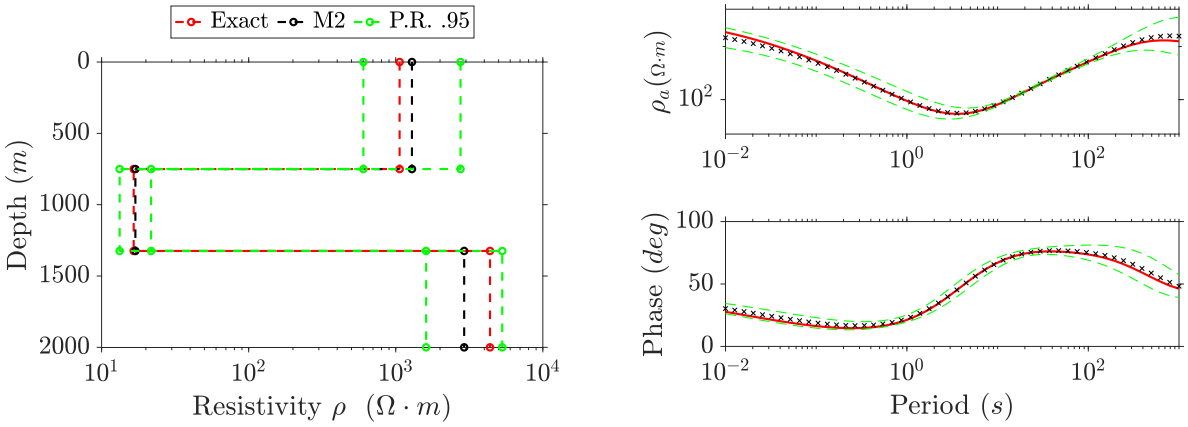
Estimating the thickness of the layers is sometimes part of the inverse problem. However, in here we assume the bed boundaries have been previously determined using seismic measurements (Xu *et al.* 2012, this is also a common situation in practice). Thus, in this work, the associated inverse problem to eq. (15) estimates the resistivity ρ for a set of N apparent resistivities $\rho_a = \left[\rho_a^{(1)}, \rho_a^{(2)}, \ldots, \rho_a^{(N)}\right]$, N phases $\phi = [\phi_1, \phi_2, \ldots, \phi_N]$ and N frequencies $\omega = (\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(N)})$ in $L-1$ subsurface fixed thickness layers $h = (h_1, h_2, \ldots, h_{L-1})$. We denote the inverse problem as:

$$\mathcal{I}^{\mathrm{MT}}(y_{\mathrm{MT}}) = \rho. \tag{16}$$



(a) x_{MT} point estimations (b) x_{MT} densities estimations

Figure 7. (a) and (b) show the estimations obtained for one sample of the validation set by MVAE with five densities. The mean M2 has the highest probability of occurrence.



(a) x_{MT} - M2 point estimation (b) y_{MT} - M2 point estimation

Figure 8. Estimations obtained for one sample of the validation set by MVAE with five densities.

In order to provide a solution to eq. (16), we use the MVAE scheme over an artificial set of physical measurements with known additive noise.

5.1.2 Data generation

The database is composed of training and validation sets created under the same assumptions. In particular, we assume that the number and thickness of subsurface layers are known, and the employed frequencies are also fixed. We generate a database with 50 000 synthetic samples, using 80 per cent for training and 20 per cent for validation. The synthetic samples are constructed in four steps:

(i) Initialize data. We fix the number of layers L and the number of frequencies N. Then, we select frequencies $\omega^{(j)}$ obtained from $\omega^{(j)} = 10^{(-2+(j-1)\Delta)}$ for $j = 1,\ldots, N$ and $\Delta = 5/(N-1)$.

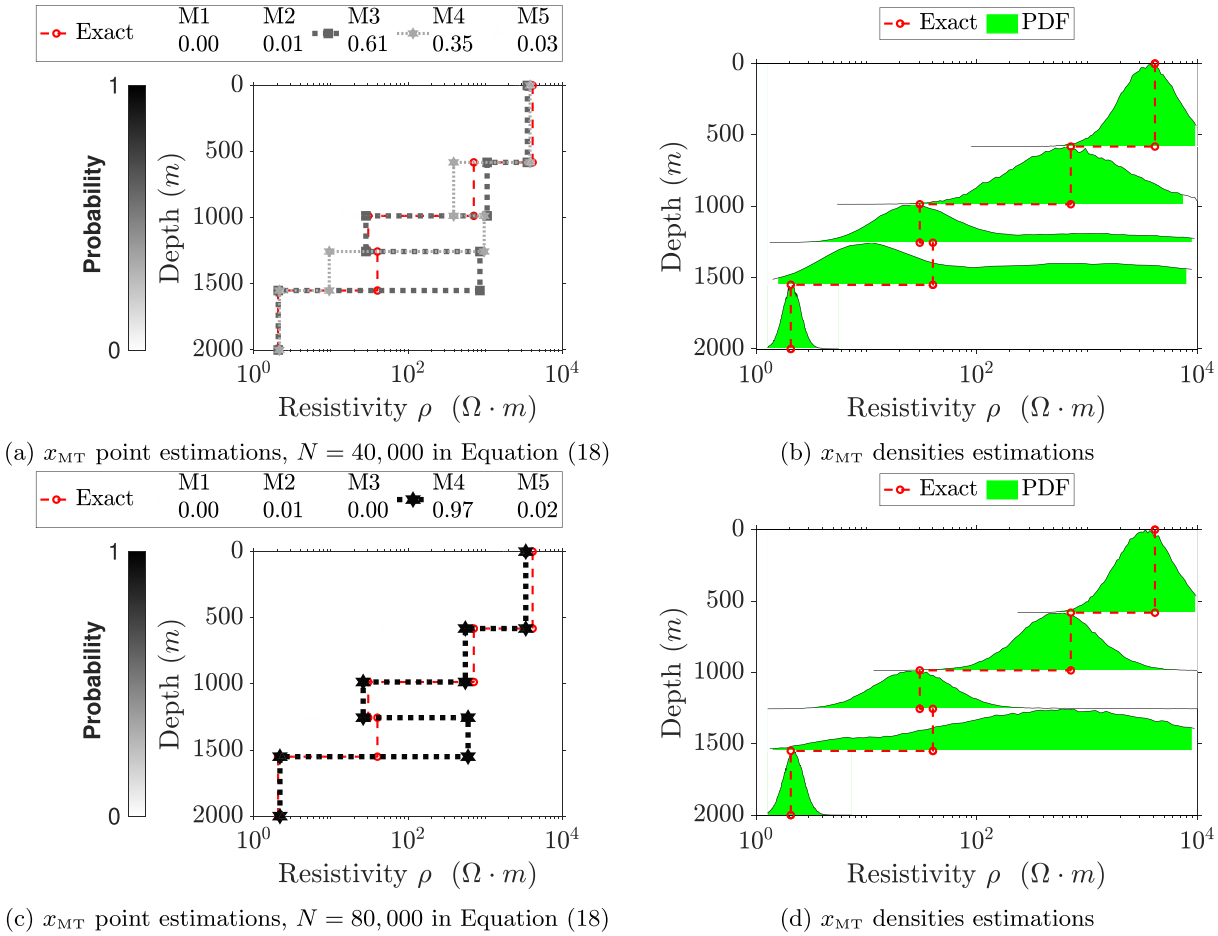
(ii) Generate subsurface properties x_{MT}. Subsurface thickness h_j^i are generated from $h_1^i = 2,000 \cdot \exp(v_1^i)/\sum_{j=1}^{L}\exp(v_j^i)$ and $h_{j+1}^i = 2,000 \cdot \exp(v_j^i)/\sum_{j=1}^{L}\exp(v_j^i) + h_j^i$ in metres, where each v_j^i is obtained from a randomly uniform distribution in the domain (0,4) for $j = 1,\ldots, L$ and $i = 1,\ldots, 50\,000$; resistivities $\rho_i^{(j)}$ are generated from $\rho_i^{(j)} = 10^{u_i^j}$, where u_i^j are obtained randomly from a uniform distribution in the domain (0,4) for $j = 1,\ldots, L$ and $i = 1,\ldots, 50\,000$.

(iii) Generate apparent resistivities and phases y_{MT}. From the above x_{MT}, we solve eqs (13) and (14) to obtain y_{MT}.

(iv) Incorporate additive noise. In Xiang *et al.* (2018), Ghaedrahmati *et al.* (2022) and Siripunvaraporn *et al.* (2005), they consider the admissible noise between 5 and 10 per cent for each measurement in the inverse MT problem. Therefore, for each y_{MT}, we consider a Gaussian additive noise with a vector of mean zero and the diagonal Γ_k^D of the diagonal covariance matrix Γ_k is computed using $\Gamma_k^D = (0.03 y_{\mathrm{MT}})^2$, where the square is applied elementwise.

5.2 Estimations with MVAE

We use the MVAE scheme in Fig. 4 with five truncated Gaussian densities in the domain (0,4), such that the estimations x_{MT} are in the domain of $\mathcal{F}$. Table 1 shows the NN architecture implemented in the application of the MT inverse problem. We consider a uniform prior distribution in the domain (0,4) and the likelihood as in eq. (9) with an unknown covariance matrix. However, we consider that the admissible noise is associated with the prediction. Then, the



(a) x_{MT} point estimations, $N = 40,000$ in Equation (18)

(b) x_{MT} densities estimations

(c) x_{MT} point estimations, $N = 80,000$ in Equation (18)

(d) x_{MT} densities estimations

Figure 9. (a) and (b) show the estimations obtained with 40 000 samples in the training set. And (c) and (d) show the estimations obtained with 80 000 samples in the training set. Bought results are for one sample of the validation set by MVAE with five densities.

likelihood is given by:

$$p(y_{\mathrm{MT}}|x_{\mathrm{MT}}; \mathbf{0}_k, \Gamma_k^D) = (2\pi)^{-\frac{k}{2}} |\Gamma_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(y_{\mathrm{MT}} - \mathcal{F}^{\mathrm{MT}}(x_{\mathrm{MT}})\right)^t \Gamma_k^{-1}\left(y_{\mathrm{MT}} - \mathcal{F}^{\mathrm{MT}}(x_{\mathrm{MT}})\right)\right), \tag{17}$$

where Γ_k^D is the diagonal of the diagonal covariance matrix Γ_k, and $\Gamma_k^D = (\beta \mathcal{F}^{\mathrm{MT}}(x_{\mathrm{MT}}))^2$ with $\beta = 0.03$ (considering noise between 0 and 6 percent). Then, the estimation of θ_{KL} is given by:

$$\theta_{\mathrm{KL}} \approx \frac{1}{H \cdot N} \underset{\theta}{\operatorname{argmin}} \sum_{n=1}^{N} \sum_{i=1}^{H} \left[\log q(x^i_{\mathrm{MT}_n}; \varphi_\theta(y_{\mathrm{MT}_n}), 0, 4) - \left(\log \frac{1}{4} + \log p(y_{\mathrm{MT}_n}|x^i_{\mathrm{MT}_n}; \mathbf{0}_k, \Gamma_k^D)\right)\right]. \tag{18}$$

We generate the database with 50 frequencies using 3, 5 and 10 subsurface layers. Codes for solving forward and inverse problems were implemented in TensorFlow, and the mixture of Gaussian densities was implemented in TensorFlow Probability with automatic differentiation. We employed the optimizer Adam with a learning rate of 10^{-5}. The following results were obtained using only MVAE instead of VAE because the proposed model can obtain unimodal distributions if a unique solution exists, as shown in Fig. 7(b).

5.2.1 MT model problem with three geological layers

Fig. 7(a) (red curve) shows our model example selected from the validation data set, composed of three rather resistive layers (all of them over 10 Ω·m). The remaining curves in Fig. 7(a) show all possible solutions encountered by our inversion algorithm, along with the probability of each solution. A full probability distribution of the solution of each layer is displayed in Fig. 7(b). We obtain an unimodal Gaussian distribution in each layer and observe that the last layer exhibits a lower variance than the first one. This occurs because the last layer is infinitely thick, which can be more easily determined based on low-frequency measurements. The middle layer shows the lowest variance because electromagnetic methods have a larger sensitivity to conductive layers than resistive ones.

Out of the five considered Gaussian distributions, Fig. 8 shows the two with the highest probability of occurrence. The mean M2 is the most probable solution, which we show together with 95 per cent of the probability regions calculated over the validation data set. Fig. 8(b) results from applying the forward problem to the estimation obtained in Fig. 8(a), thus comparing the level of proximity of the estimation with the trained observation. These figures show that the estimated phase and apparent resistivities are close to the exact values of the observation in the validation set.

5.2.2 MT model problem with five geological layers

Fig. 9(a) (red curve) shows our model example selected from the validation data set, composed of five layers. The remaining curves in Fig. 9(a) show all possible solutions encountered by our inversion algorithm, along with the probability of each solution. A full probability distribution of the solution of each layer is displayed in Fig. 9(b). We obtain unimodal density in the first and last layers. This is expected because the first layer is near the surface—and it can be uniquely determined from high-frequency measurements, which have a rather low depth of investigation—while the last layer is the thickest one (extending to $-\infty$), thus it can also be (almost) uniquely determined from low-frequency data. However, most mid-layer resistivities are much more challenging to determine, and different combinations can provide almost identical results (with negligible differences within the noise level). As a result, we obtain a heavy-tailed distribution in the third and fourth layers.

Out of the five considered densities, Fig. 10 shows the two with the highest probability of occurrence. Figs 10(b) and (d) result from applying the forward problem to the estimation obtained in Figs 10(a) and (c), thus comparing the level of proximity of the estimation with the trained observation. We infer that high variance in one of the middle layers does not significantly affect the phase and apparent resistivity estimations.

Fig. 9 shows a training set of 40 000 and 80 000 samples, with a low-uncertainty estimation for the first and last layers. The fourth layer, in both cases, has high uncertainty. Although for 80 000 samples the results tend to be unimodal, it is observed that the distribution obtained in the fourth layer has a wider variance. In both cases, the point estimations solve the 1-D MT inverse problem, and 40 000 samples are sufficient to obtain an accurate estimation, as shown in Fig. 10(b).

5.2.3 MT model problem with 10 geological layers

Fig. 11(a) (red curve) shows a 10-layer model example selected from the validation data set. The remaining curves in Fig. 11(a) show all possible solutions encountered by our inversion algorithm, along with the probability of each solution. A full probability distribution of the solution of each layer is displayed in Fig. 11(b). As before, the first and last layer resistivities can be uniquely determined from high- and low-frequency measurements, respectively. The remaining layers show higher uncertainty as we go deeper into the subsurface, as physically expected. We infer that our MVAE would require more densities to encounter a larger number of solutions to this inverse problem.

Out of the five considered densities, Fig. 12 shows the highest probability of occurrence. Figs 12(a) and (d) result from applying the forward problem to the estimation obtained in Figs 12(a) and (c), respectively; thus comparing the level of proximity of the estimation with the trained observation. We infer that high variance in the mid-layers does not significantly affect phase and apparent resistivity estimations. Although the exact solution is not included in the probability region (see Figs 12a and c), solutions from that interval satisfy the physical measurements up the noise precision (see Figs 12a and d). We can thus conclude that the failure of our method to identify the exact solution is not a failure of the method itself. Rather, the ill-conditioned nature of the inverse problem has produced measurements too noisy, relative to the complexity of the ground truth, to be able to obtain an accurate solution.

5.3 Results of MT model problem with five geological layers using classical MHMC

We replicate the example of five geological layers with the same subsurface configuration by using a basic Monte-Carlo approach calculated from the posterior distribution. Note that the posterior distribution p is proportional to a Gaussian additive noise because we use a uniform distribution as prior, thus $p(x^r_{\mathrm{MT}}|y^r_{\mathrm{MT}}) \propto p(y^r_{\mathrm{MT}}|x^r_{\mathrm{MT}})$ in eq. (17), where r is the rth sample of the validation set. We use the Metropolis–Hastings Monte Carlo (MHMC) scheme described in Algorithm 1 to generate 500 000 samples $\{k_i\}$ of $p(x^r_{\mathrm{MT}}|y^r_{\mathrm{MT}})$. Subsequently, we use a systematic sampling, where we select the samples $\{k_{10s}\}$ for $s = 1,\ldots, 50\,000$, reducing the number of samples to 50 000. To replicate the application of MT model problem with five geological layers, we use as initial values of Algorithm 1 $\eta = 1000$ and $\sigma_0 = 0.5$.

Algorithm 1 Random walk Metropolis-Hastings with adaptive variance

Data:
1: Domain Ω of x^r_{MT}
2: Posterior $p(x^r_{\mathrm{MT}}|y^r_{\mathrm{MT}})$
3: Number of samples n
4: Initial state $k1$
5: Size for the adaptive variance η
6: Initial value of the variance σ_0^2

Result: n samples $\{k_i\}_{i=1}^n$ from $p(x^r_{\mathrm{MT}}|y^r_{\mathrm{MT}})$
$i \leftarrow 1$; $k_1 \leftarrow k1$; $\sigma \leftarrow \sigma_0$; $z_0 \leftarrow 0$; $count \leftarrow 0$;
while $i \leq n$ **do**
  Generate k' from $\mathcal{TN}(k_i, \sigma^2; \Omega)$; ▷ Truncated Normal in Ω with mean k_i and variance σ^2
  $\alpha \leftarrow \min\left(0, \log p(k_i|y^r_{\mathrm{MT}}) - \log p(k'|y^r_{\mathrm{MT}})\right)$;
  Generate u from $\mathcal{U}(0, 1)$;
  if $\log u < \alpha$ **then**
    $k_{i+1} \leftarrow k'$;
    $count \leftarrow count + 1$;
  else
    $k_{i+1} \leftarrow k_i$;
  end
  if $\mod(i, \eta) = 0$ **then**
    $z_1 \leftarrow count$;
    $\sigma \leftarrow \max\left(|\frac{z1-z0}{\eta}|, 0.01\right)$; ▷ $\sigma \in [0.01, 1]$
    $z_0 \leftarrow z_1$;
  end
  $i \leftarrow i + 1$;
end

Figs 13, 14 and 15 show a representation of different modes using classical Monte Carlo. These figures provide a quantification of the uncertainty and point estimation of the solution to the inverse problem. We have a set of apparent resistivities and phases from 50 000 combinations of subsurface properties (divided in 80 per cent for training and 20 per cent for validation).

Using a Monte Carlo for each combination may be computationally expensive. Using MVAE presents an advantage in computational cost vs the classical Monte Carlo since the MVAE provides a parametric distribution for each element in the data set and M different point estimations—as shown in Figs 9 and 10—with a lower computational cost. In Fig. 15, we observe a very similar behaviour to the results obtained in Figs 10(a) and (b), which corresponds to the most plausible result in the

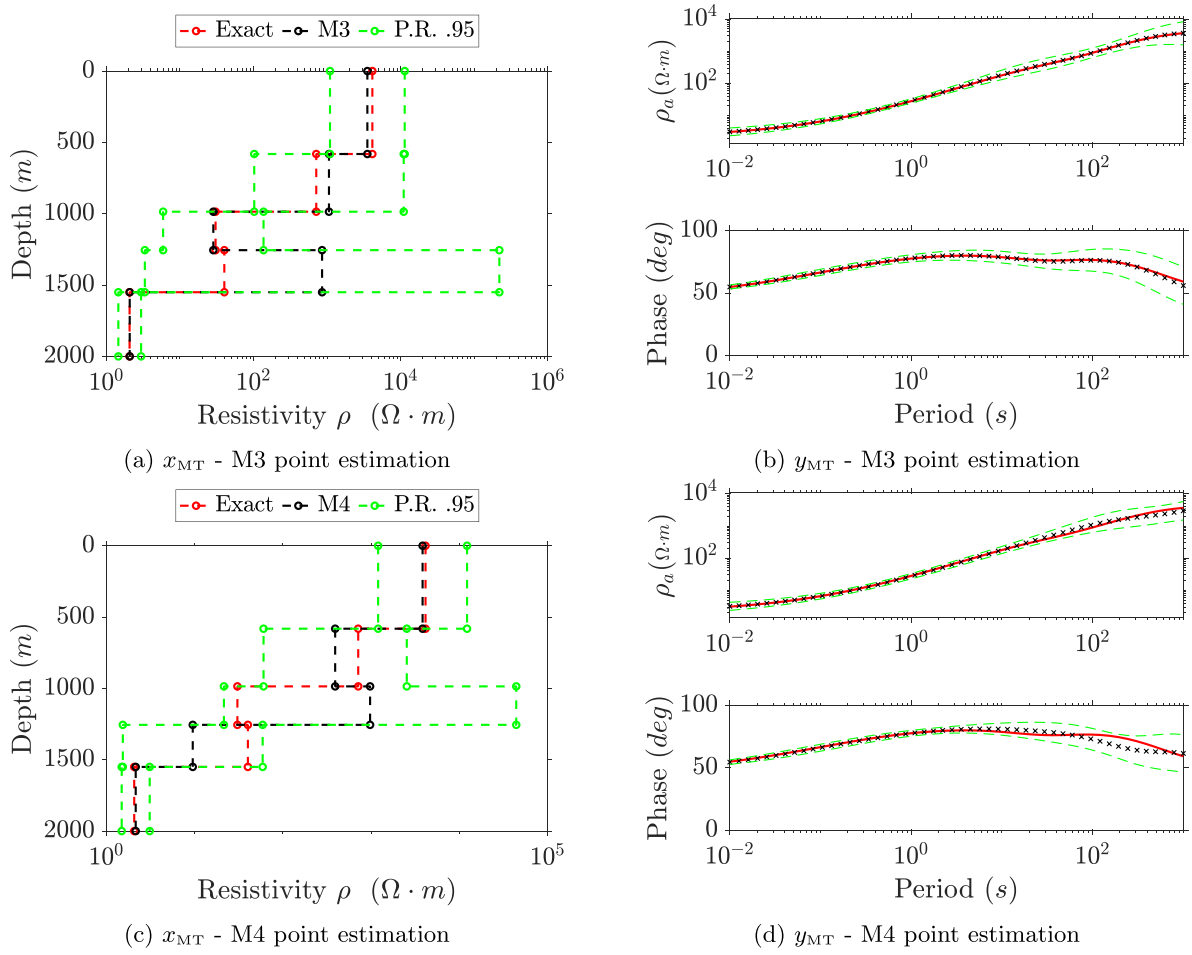


Figure 10. Estimations obtained for one sample of the validation set by MVAE with five densities.

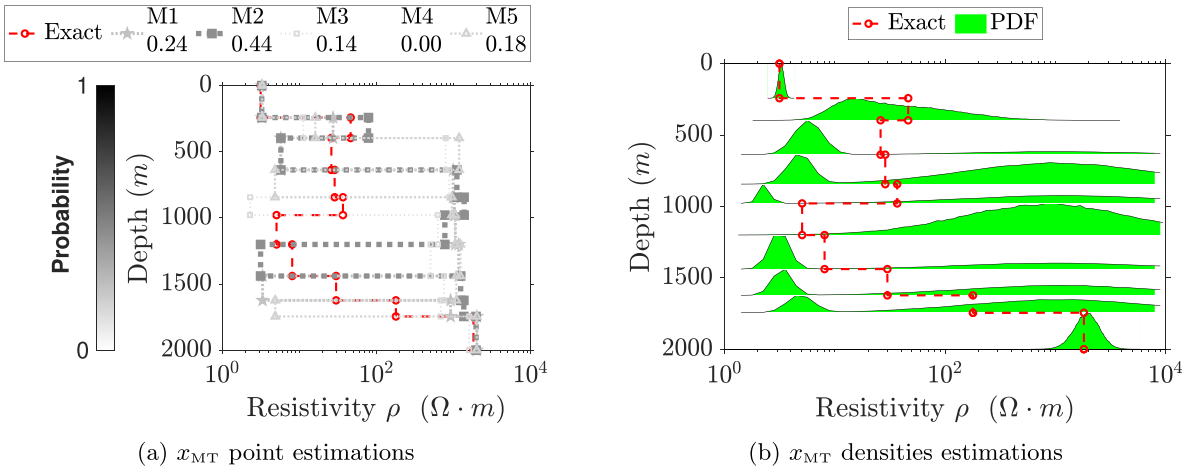
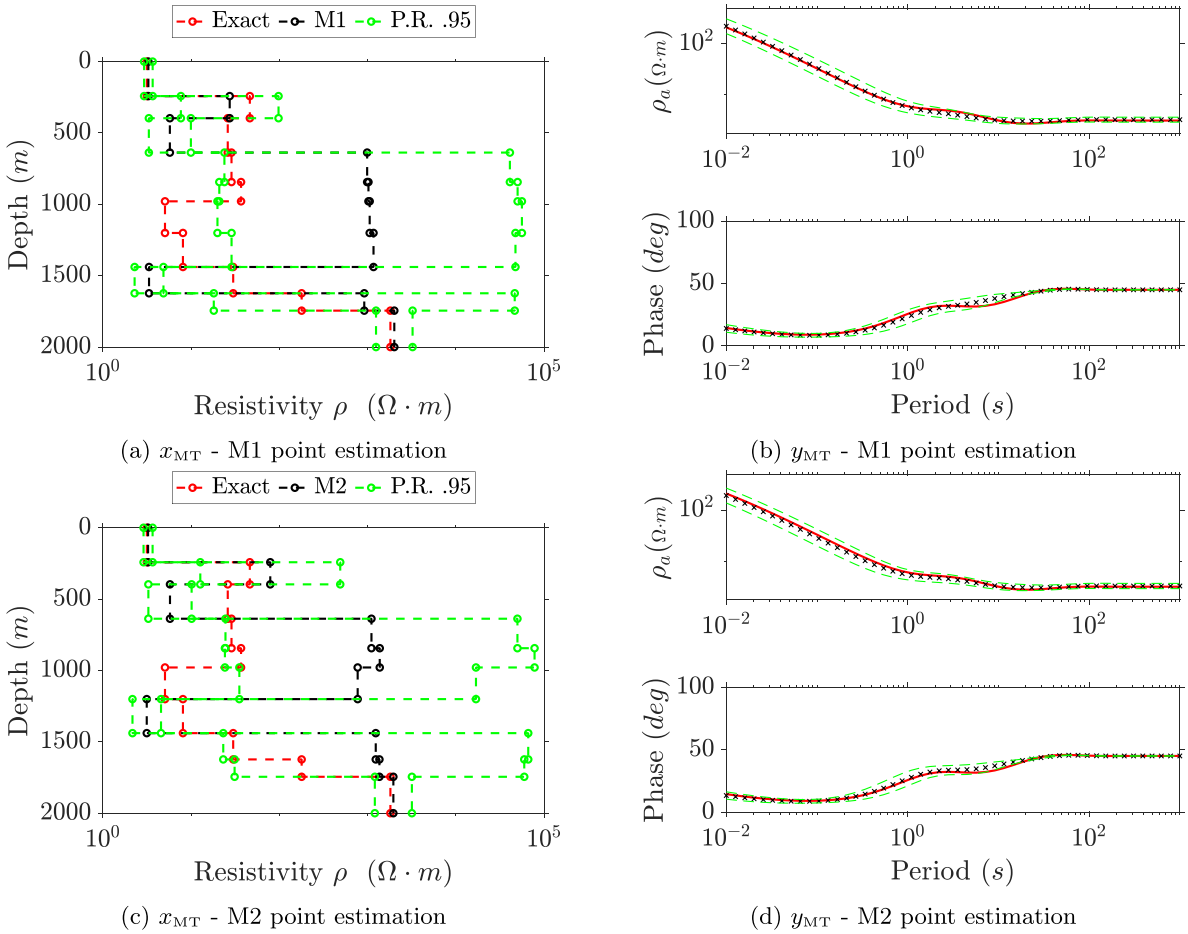


Figure 11. (a) and (b) show the estimations obtained for one sample of the validation set by MVAE with five densities. The mean M2 has the highest probability of occurrence.

MT problem with five geophysical layers solved by MVAE. This result confirms that the estimates provided by MVAE are reliable. In addition, with MVAE, we can predict the results of elements that are not part of the training set, such as in this case, where the provided example is in the validation set. This is not possible with a classical Monte Carlo approach, as each solution must be obtained independently. Furthermore, whilst MHMC is able to accurately approximate the posterior distribution, it is generally difficult to interpret the obtained results, as they amount to a large sample from the posterior distribution. The MVAE represents the posterior via a parametrized PDF, and thus the parameters of the distribution provide a low-dimensional and easily explainable interpretation of the obtained results.



(a) x_{MT} - M1 point estimation (b) y_{MT} - M1 point estimation

(c) x_{MT} - M2 point estimation (d) y_{MT} - M2 point estimation

Figure 12. Estimations obtained for one sample of the validation set by MVAE with five densities.

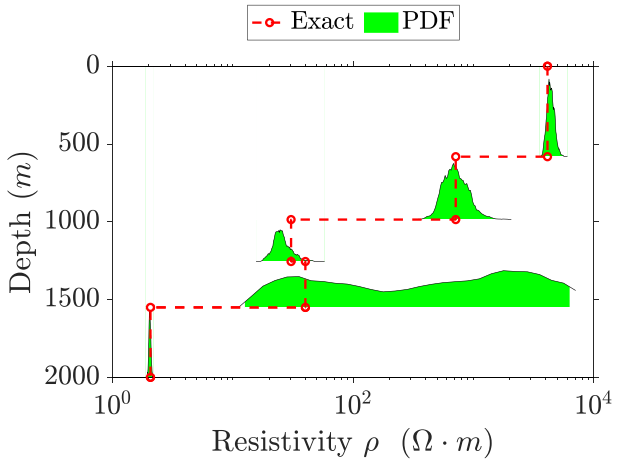


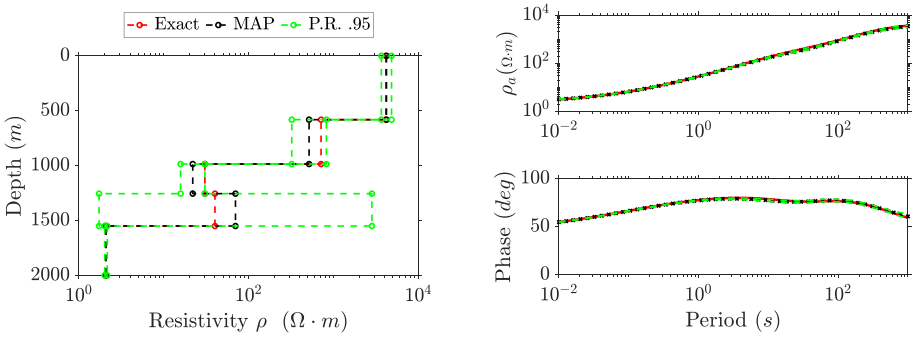
Figure 13. x_{MT} densities estimations.

6 CONCLUSIONS

Implementing a VAE implies approximating the posterior distribution by a more tractable one. Usually, selecting an unimodal distribution is the standard option. However, different configurations of the subsurface can generate the same geophysical observations. With this, we propose the variational autoencoder MVAE, a multimodal autoencoder model that provides multiple solutions to a geophysical inverse problem associated with a known forward problem. This model has three advantages: (1) it provides multiple point estimates for each observation; (2) each point estimate has a probability of occurrence, which allows inferring the most plausible inverse solution, and the variance represents the confidence level of the obtained estimations and (3) we obtain a distribution associated with each solution parameter. Applying MVAE in geophysical problems that exhibit multiple inverse solutions is useful since MVAE allows us to obtain different modes in the solution with indicated probabilities.

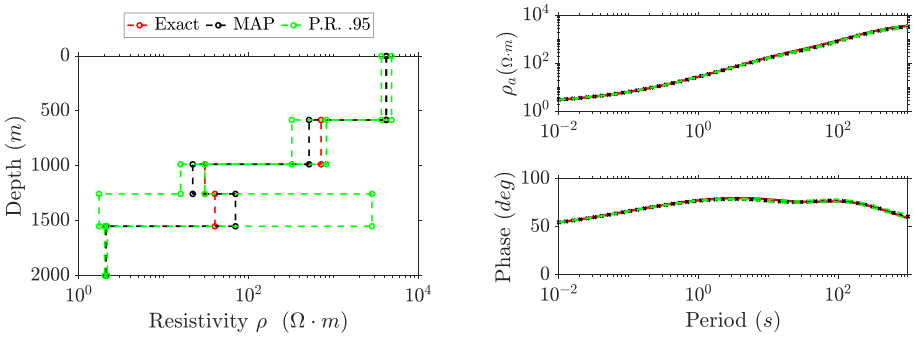
MVAE has been applied to the MT1-D inverse problem. In the considered numerical examples, we obtain a unique solution to the inverse problem with unimodal distributions in the first and last layers. This is physically expected since the first layer is near the surface, and high-frequency measurements are sufficient to determine it uniquely. The last layer is infinitely thick, and low-frequency measurements can uniquely determine its resistivity. In addition, when the number of subsurface layers increases, the estimated solutions in the middle layers show a high dispersion, obtaining multimodal estimations in some cases. In others, we obtain estimations of high dispersion without a distinctive mode, for example, in the case of 10 subsurface layers. We validate the obtained estimates of subsurface properties x_{MT} by applying the forward problem and comparing these solutions with the synthetic physical measurements y_{MT}.

Understanding and predicting uncertainty have tremendous value in subsurface exploration. The proposed method contributes to this understanding since we can infer whether unimodal, multimodal, or other distributions are obtained from the solution distributions. As in other multimodal models, a limitation of MVAE is that it does



(a) x_{MT} - Maximum a posterior estimation (MAP) (b) y_{MT} - Maximum a posterior estimation (MAP)

Figure 14. Estimations obtained for one sample of the validation set by Monte Carlo approach.



(a) x_{MT} - Maximum a posterior estimation (MAP) (b) y_{MT} - Maximum a posterior estimation (MAP)

Figure 15. Autoencoder NN scheme.

not provide the exact number of solutions to the inverse problem. It provides, at most, as many solutions as the number of mixtures we choose.

Whilst we have shown that our proposed method works effectively on samples chosen from our selected prior (a log-uniform distribution), we are essentially using synthetic data, and thus, the model is trained according to a distribution ground truth that may not reflect reality adequately, leading to large errors when applied to non-synthetic problems. In order to be applicable to real-world systems, one would need to employ a prior that more accurately reflects the ground truth, which would need to be informed by expert knowledge and real-world data. However, if the data are unavailable, there are generic distributions to solve inverse problems, such as Gaussian random Markov fields Bardsley & Kaipio (2013) or Gaussian mixtures Grana *et al.* (2017). In future work, we will consider the thickness of each layer as an unknown of the inverse problem. In addition, the forward problem can be scaled to two or three dimensions, where the autoencoder supports the structure. However, obtaining the forward problem solution needed for training is computationally more expensive. We also consider applying the proposed autoencoder in applications with multiphysics data to characterize in which cases the solution could be unique, inferring this from the estimation of a posterior distribution as an unimodal or multimodal model.

ACKNOWLEDGMENTS

DP has received funding from: the Spanish Ministry of Science and Innovation projects with references TED2021-132783B-I00, PID2019-108111RB-I00 (FEDER/AEI) and PDC2021-121093-I00 (MCIN/AEI/10.13039/501100011033/Next Generation EU), the 'BCAM Severo Ochoa' accreditation of excellence CEX2021-001142-S/MICIN/AEI/10.13039/501100011033; the Spanish Ministry of Economic and Digital Transformation with Misiones Project IA4TES (MIA.2021.M04.008/NextGenerationEU PRTR); and the Basque Government through the BERC 2022-2025 program, the Elkartek project SIGZE (KK-2021/00095), and the Consolidated Research Group MATHMODE (IT1456-22) given by the Department of Education.

JMT is supported by the Basque Government through the BERC 2018-2021 program and by the Spanish State Research Agency through BCAM Severo Ochoa excellence accreditation SEV-2017-0718 and through project PID2020-114189RB-I00 funded by Agencia Estatal de Investigación (PID2020-114189RB-I00/AEI/10.13039/501100011033).

DATA AVAILABILITY

The 1-D MT inverse problem code via MVAE used in this work is available at https://zenodo.org/badge/latestdoi/620889753

REFERENCES

Agranovich, Z. & Marchenko, V.A., 2020. *The Inverse Problem of Scattering Theory,* Courier Dover Publications.

Alvarez-Aramberri, J. & Pardo, D., 2017. Dimensionally adaptive hp-finite element simulation and inversion of 2D magnetotelluric measurements, *J. Comput. Sci.,* **18,** 95–105.

Alyaev, S. & Elsheikh, A.H., 2022. Direct multi-modal inversion of geophysical logs using deep learning, *Earth Space Sci.,* **9**(9), e2021EA002186.

Alyaev, S., Tveranger, J., Fossum, K. & Elsheikh, A.H., 2021. Probabilistic forecasting for geosteering in fluvial successions using a generative adversarial network, *First Break,* **39**(7), 45–50.

Arridge, S.R. & Schotland, J.C., 2009. Optical tomography: forward and inverse problems, *Inverse Prob.,* **25**(12), 123010.

Aster, R.C., Borchers, B. & Thurber, C.H., 2018. *Parameter Estimation and Inverse Problems,* Elsevier.

Astic, T. & Oldenburg, D.W., 2019. A framework for petrophysically and geologically guided geophysical inversion using a dynamic Gaussian mixture model prior, *Geophys. J. Int.,* **219**(3), 1989–2012.

Bardsley, J.M. & Kaipio, J., 2013. Gaussian Markov random field priors for inverse problems., *Inverse Prob. Imag.,* **7**(2). https://doi.org/10.3934/ipi.2013.7.397.

Berliner, L.M., 2003. Physical-statistical modeling in geophysics, *J. geophys. Res.: Atmos.,* **108**(D24). https://doi.org/10.1029/2002JD002865.

Bishop, C.M., 1994. Mixture density networks, Tech. Rep., Aston University.

Blundell, C., Cornebise, J., Kavukcuoglu, K. & Wierstra, D., 2015. Weight uncertainty in neural network, in *International Conference on Machine Learning,* pp. 1613–1622, PMLR.

Calvetti, D. & Somersalo, E., 2018. Inverse problems: From regularization to Bayesian inference, *Wiley Interdiscip. Rev.: Comput. Stat.,* **10**(3), e1427.

Capistrán, M.A., Christen, J.A., Daza-Torres, M.L., Flores-Arguedas, H. & Montesinos-López, J.C., 2021. Error control of the numerical posterior with Bayes factors in Bayesian uncertainty quantification, *Bayesian Anal.,* **1**(1), 1–23.

Chen, D. & Manning, C.D., 2014. A fast and accurate dependency parser using neural networks, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),* Association for Computational Linguistics Moschitti, A., Pang, B. & Daelemans, W.. pp. 740–750. https://aclanthology.org/volumes/D14-1/.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. & Bharath, A.A., 2018. Generative adversarial networks: an overview, *IEEE Signal Process. Mag.,* **35**(1), 53–65.

Cui, H., Radosavljevic, V., Chou, F.-C., Lin, T.-H., Nguyen, T., Huang, T.-K., Schneider, J. & Djuric, N., 2019. Multimodal trajectory predictions for autonomous driving using deep convolutional networks, in *2019 International Conference on Robotics and Automation (ICRA),* IEEE. pp. 2090–2096. https://doi.org/10.1109/ICRA39644.2019.

de Figueiredo, L.P., Grana, D., Roisenberg, M. & Rodrigues, B.B., 2019. Gaussian mixture Markov chain Monte Carlo method for linear seismic inversion, *Geophysics,* **84**(3), R463–R476.

Earp, S., Curtis, A., Zhang, X. & Hansteen, F., 2020. Probabilistic neural network tomography across grane field (north sea) from surface wave dispersion data, *Geophys. J. Int.,* **223**(3), 1741–1757.

Farnia, F. & Tse, D., 2016. A minimax approach to supervised learning, Lee, D., Sugiyama, M., Luxburg, U., Guyon, I. & Garnett, R., eds. *Adv. Neural Inform. Process. Syst.,* **29**. Garnett ISBN: 9781510838819.

Fichtner, A., Zunino, A. & Gebraad, L., 2019. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.,* **216**(2), 1344–1363.

Fossum, K., Alyaev, S., Tveranger, J. & Elsheikh, A.H., 2022. Verification of a real-time ensemble-based method for updating earth model based on gan, *J. Comput. Sci.,* **65,** 101876.

Ghaedrahmati, R., Moradzadeh, A. & Moradpouri, F., 2022. An effective estimate for selecting the regularization parameter in the 3D inversion of magnetotelluric data, *Acta Geophys.,* **70**(2), 609–621.

Goh, H., Sheriffdeen, S., Wittmer, J. & Bui-Thanh, T., 2021. Solving Bayesian inverse problems via variational autoencoders, *Proc. Mach. Learn. Res.,* **145,** 1–40.

Grana, D., Fjeldstad, T. & Omre, H., 2017. Bayesian Gaussian mixture linear inversion for geophysical inverse problems, *Math. Geosci.,* **49**(4), 493–515.

Grandis, H., Menvielle, M. & Roussignol, M., 1999. Bayesian inversion with Markov chains–i. the magnetotelluric one-dimensional case, *Geophys. J. Int.,* **138**(3), 757–768.

Guo, C., Zhou, J., Chen, H., Ying, N., Zhang, J. & Zhou, D., 2020. Variational autoencoder with optimizing Gaussian mixture model priors, *IEEE Access,* **8,** 43992–44005.

Guo, Y., Liao, W., Wang, Q., Yu, L., Ji, T. & Li, P., 2018. Multidimensional time series anomaly detection: a gru-based Gaussian mixture variational autoencoder approach, in *Asian Conference on Machine Learning,* pp. 97–112, PMLR.

Gut, A., 2013. *Probability: A Graduate Course,* Vol. **75,** Springer Science & Business Media.

Häggström, I., Schmidtlein, C.R., Campanella, G. & Fuchs, T.J., 2019. Deeppet: a deep encoder–decoder network for directly solving the PET image reconstruction inverse problem, *Med. Image Anal.,* **54,** 253–262.

Hermans, T., Nguyen, F., Klepikova, M., Dassargues, A. & Caers, J., 2018. Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using Bayesian evidential learning, *Water Resour. Res.,* **54**(4), 2931–2948.

Hornik, K., Stinchcombe, M. & White, H., 1989. Multilayer feedforward networks are universal approximators, *Neural Netw.,* **2**(5), 359–366.

Hoffman, M.D. & Johnson, M.J., 2016. ELBO surgery: yet another way to carve up the variational evidence lower bound, in *Workshop in Advances in Approximate Bayesian Inference, NIPS,* Vol. **1,** 2. http://approximateinference.org/2016/.

Jahani, N., Ambia Garrido, J., Alyaev, S., Fossum, K., Suter, E. & Torres-Verdín, C., 2022. Ensemble-based well-log interpretation and uncertainty quantification for well geosteering, *Geophysics,* **87**(3), IM57–IM66.

Kaipio, J. & Somersalo, E., 2006. *Statistical and Computational Inverse Problems,* Vol. **160,** Springer Science & Business Media.

Kingma, D.P. & Welling, M., 2013. Auto-encoding variational Bayes, in *The 2nd International Conference on Learning Representations,* ICLR. https://doi.org/10.48550/arXiv.1312.6114.

Kolesnikov, Y.I. & Fedin, K., 2018. Detecting underground cavities using microtremor data: physical modelling and field experiment, *Geophys. Prospect.,* **66**(2), 342–353.

Kullback, S. & Leibler, R.A., 1951. On information and sufficiency, *Ann. Math. Stat.,* **22**(1), 79–86.

Lai, R.-Y. & Lin, Y.-H., 2022. Inverse problems for fractional semilinear elliptic equations, *Nonlinear Anal.,* **216,** 112699. https://doi.org/10.1016/j.na.2021.112699.

Liu, M., Grana, D. & de Figueiredo, L.P., 2022. Uncertainty quantification in stochastic inversion with dimensionality reduction using variational autoencoder, *Geophysics,* **87**(2), M43–M58.

Livingstone, D.J., Manallack, D.T. & Tetko, I.V., 1997. Data modelling with neural networks: advantages and limitations, *J. Comput.-Aided Mol. Des.,* **11**(2), 135–142.

Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: hierarchical Bayes and empirical Bayes, *Geophysics,* **69**(4), 1005–1016.

Mandolesi, E., Ogaya, X., Campanyà, J. & Agostinetti, N.P., 2018. A reversible-jump Markov chain Monte Carlo algorithm for 1D inversion of magnetotelluric data, *Comput. Geosci.,* **113,** 94–105.

Mao, K.Z., Tan, K.-C. & Ser, W., 2000. Probabilistic neural-network structure determination for pattern classification, *IEEE Trans. Neural Netw.,* **11**(4), 1009–1016.

Medin, A.E., 2008. *The Magnetotelluric Inverse Problem,* PhD thesis, UC San Diego.

Meier, U., Curtis, A. & Trampert, J., 2007. Global crustal thickness from neural network inversion of surface wave data, *Geophys. J. Int.,* **169**(2), 706–722.

Michel, H., Nguyen, F., Kremer, T., Elen, A. & Hermans, T., 2020. 1D geological imaging of the subsurface from geophysical data with Bayesian evidential learning, *Comput. Geosci.,* **138,** 104456. https://doi.org/10.1016/j.cageo.2020.104456.

Neuneier, R., Hergert, F., Finnoff, W. & Ormoneit, D., 1994. Estimation of conditional densities: A comparison of neural network approaches, in *International Conference on Artificial Neural Networks,* pp. 689–692, Springer.

Oikarinen, T., Hannah, D. & Kazerounian, S., 2021. Graphmdn: Leveraging graph structure and deep learning to solve inverse problems, in *2021 International Joint Conference on Neural Networks (IJCNN),* pp. 1–9, IEEE.

Olierook, H.K., Scalzo, R., Kohn, D., Chandra, R., Farahbakhsh, E., Clark, C., Reddy, S.M. & Müller, R.D., 2021. Bayesian geological and geophysical data fusion for the construction and uncertainty quantification of 3D geological models, *Geosci. Front.,* **12**(1), 479–493.

Pace, F., Santilano, A. & Godio, A., 2019. Particle swarm optimization of 2D magnetotelluric data, *Geophysics,* **84**(3), E125–E141.

Park, D., Hoshi, Y. & Kemp, C.C., 2018. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder, *IEEE Robot. Automat. Lett.,* **3**(3), 1544–1551.

Parker, R.L., 1983. The magnetotelluric inverse problem, *Geophys. Surv.,* **6**(1), 5–25.

Pilozzi, L., Farrelly, F.A., Marcucci, G. & Conti, C., 2018. Machine learning inverse problem for topological photonics, *Commun. Phys.,* **1**(1), 1–7.

Raissi, M., Perdikaris, P. & Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.,* **378,** 686–707.

Rammay, M.H., Alyaev, S. & Elsheikh, A.H., 2022a. Probabilistic model-error assessment of deep learning proxies: an application to real-time inversion of borehole electromagnetic measurements, *Geophys. J. Int.*. https://doi.org/10.1093/gji/ggac147.

Rammay, M.H., Alyaev, S., Larsen, D.S., Bratvold, R.B. & Saint, C., 2022b. Strategic geosteeering workflow with uncertainty quantification and deep learning: A case study on the goliat field, preprint, arXiv:2210.15548.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C.J. & Müller, K.-R., 2021. Explaining deep neural networks and beyond: a review of methods and applications, *Proc. IEEE,* **109**(3), 247–278.

Schapire, R.E., 2003. The boosting approach to machine learning: an overview, in *Nonlinear Estimation and Classification,* pp. Springer, 149–171.

Shahriari, M., Pardo, D., Picón, A., Galdran, A., Del Ser, J. & Torres-Verdín, C., 2020. A deep learning approach to the inversion of borehole resistivity measurements, *Comput. Geosci.,* **24**(3), 971–994.

Shahriari, M., Pardo, D., Rivera, J.A., Torres-Verdín, C., Picon, A., Del Ser, J., Ossandón, S. & Calo, V.M., 2021. Error control and loss functions for the deep learning inversion of borehole resistivity measurements, *Int. J. Numer. Methods Eng.,* **122**(6), 1629–1657.

Shepherd, A.J., 2012. *Second-Order Methods for Neural Networks: Fast and Reliable Training Methods for Multi-Layer Perceptrons,* Springer Science & Business Media.

Sidky, E.Y., Lorente, I., Brankov, J.G. & Pan, X., 2020. Do cnns solve the ct inverse problem?, *IEEE Trans. Biomed. Eng.,* **68**(6), 1799–1810.

Siripunvaraporn, W., Egbert, G., Lenbury, Y. & Uyeshima, M., 2005. Three-dimensional magnetotelluric inversion: data-space method, *Phys. Earth planet. Inter.,* **150**(1–3), 3–14.

Śmieja, M., Wołczyk, M., Tabor, J. & Geiger, B.C., 2020. Segma: Semi-supervised Gaussian mixture autoencoder, *IEEE Trans. Neural Netw. Learn. Syst.,* **32**(9), 3930–3941.

Spichak, V. & Popova, I., 2000. Artificial neural network inversion of magnetotelluric data in terms of three-dimensional earth macroparameters, *Geophys. J. Int.,* **142**(1), 15–26.

Tomczak, J. & Welling, M., 2018. VAE with a VampPrior, in *International Conference on Artificial Intelligence and Statistics,* pp. 1214–1223, PMLR.

Travassos, X.L., Avila, S.L. & Ida, N., 2020. Artificial neural networks and machine learning techniques applied to ground penetrating radar: a review, *Appl. Comput. Inform.***17**(2), 296–308.

Van der Baan, M. & Jutten, C., 2000. Neural networks in geophysical applications, *Geophysics,* **65**(4), 1032–1047.

Vogel, C.R., 2002. *Computational Methods for Inverse Problems,* SIAM.

Wang, F., Eljarrat, A., Müller, J., Henninen, T.R., Erni, R. & Koch, C.T., 2020. Multi-resolution convolutional neural networks for inverse problems, *Sci. Rep.,* **10**(1), 1–11.

Wang, S., 1994. A neural network method of density estimation for univariate unimodal data, *Neural Comput. Appl.,* **2**(3), 160–167.

Xia, Y. & Zabaras, N., 2022. Bayesian multiscale deep generative model for the solution of high-dimensional inverse problems, *J. Comput. Phys.,* 111008. https://doi.org/10.1016/j.jcp.2022.111008.

Xiang, E., Guo, R., Dosso, S.E., Liu, J., Dong, H. & Ren, Z., 2018. Efficient hierarchical trans-dimensional Bayesian inversion of magnetotelluric data, *Geophys. J. Int.,* **213**(3), 1751–1767.

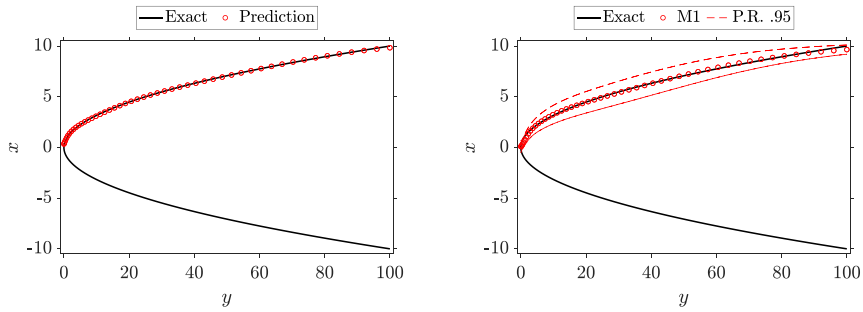
Xu, Y., Liu, X., Pan, L., Mao, X., Liang, H., Wang, G. & Chen, T., 2021. Explainable dynamic multimodal variational autoencoder for the prediction of patients with suspected central precocious puberty, *IEEE J. Biomed. Health Inform.,* **26**(3), 1362–1373.

Xu, Z., Juhlin, C., Gudmundsson, O., Zhang, F., Yang, C., Kashubin, A. & Lüth, S., 2012. Reconstruction of subsurface structure from ambient seismic noise: an example from ketzin, germany, *Geophys. J. Int.,* **189**(2), 1085–1102.

Yan, L. & Zhou, T., 2019. Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems, *J. Comput. Phys.,* **381,** 110–128.

Zhang, X. & Curtis, A., 2021. Bayesian geophysical inversion using invertible neural networks, *J. geophys. Res.: Solid Earth,* **126**(7), e2021JB022320.

Zhang, Z. & Lin, Y., 2020. Data-driven seismic waveform inversion: a study on the robustness and generalization, *IEEE Trans. Geosci. Remote Sens.,* **58**(10), 6900–6913.



(a) Estimation of the inverse problem by scheme in Figure 16

(b) Estimation of the inverse problem by the VAE scheme in Figure 3

Figure A1. Comparison between deterministic autoencoder estimates and VAE, where the obtained estimation can be different depending on the initial values of the NN weights.

APPENDIX A: DETERMINISTIC AUTOENCODER

In the deterministic case, we approximate the inverse problem $\mathcal{I}(y)$ by $\mathcal{I}_\theta(y)$ (Shahriari *et al.* 2021) without considering noise. Fig. 15 shows a basic encoder–decoder scheme that we will refer to as autoeconder, where na NN represents the encoder, and the numerical solution of the forward problem represents the decoder.

We consider the encoder as an NN with trainable parameters θ. The prediction $\mathcal{I}_\theta(y)$ must lie in the domain of $\mathcal{F}$. To enforce this, we select an adequate activation function. The decoder is a numerical solution of the forward problem $\mathcal{F}$, which is known and non-trainable. A typical loss function for autoencoder schemes is given by:

$$\mathscr{L}_\theta(y) = \|y - \mathcal{F} \circ \mathcal{I}_\theta(y)\|_2^2. \tag{19}$$

APPENDIX B: EVIDENCE LOWER BOUND LOSS FUNCTION

The Kullback–Leibler divergence (KL) can be seen as a statistical measure (asymmetric) of distance between PDFs. KL attains its global minimum of zero when the PDFs $p(x|y)$ and $q(x;\varphi_\theta(y))$ are equal, and thus it is possible to use it as a loss function and minimize the distance between PDFs. KL is defined in Kullback & Leibler (1951) by:

$$\mathrm{KL}\left[q(x;\varphi_\theta(y))||p(x|y)\right] = \int q(x;\varphi_\theta(y)) \log \frac{q(x;\varphi_\theta(y))}{p(x|y)} \mathrm{d}x, \tag{20}$$

where $p(x|y)$ is the posterior distribution of X and $q(x;\varphi_\theta(y))$ is a more tractable distribution of X, and parameters $\varphi_\theta(y)$ that characterize the PDF are obtained from an NN. From KL, we define the loss function for an autoencoder scheme as:

$$\mathscr{L}_\theta^{\mathrm{KL}}(y) = \mathrm{KL}\left[q(x;\varphi_\theta(y))||p(x|y)\right], \tag{21}$$

where X is a latent random variable. Therefore, the estimated weights $\theta\mathrm{KL}$ are defined by (we make the following calculations based on the loss function used in Blundell *et al.* 2015):

$$\theta\mathrm{KL} = \underset{\theta}{\operatorname{argmin}} \int p(y) \cdot \mathscr{L}_\theta^{\mathrm{KL}}(y) \mathrm{d}y \tag{22}$$

$$= \underset{\theta}{\operatorname{argmin}} \int p(y) \cdot \mathrm{KL}\left[q(x;\varphi_\theta(y))||p(x|y)\right] \mathrm{d}y \tag{23}$$

$$= \underset{\theta}{\operatorname{argmin}} \int\int p(y) q(x;\varphi_\theta(y)) \log \frac{q(x;\varphi_\theta(y))}{p(x|y)} \mathrm{d}x\,\mathrm{d}y \tag{24}$$

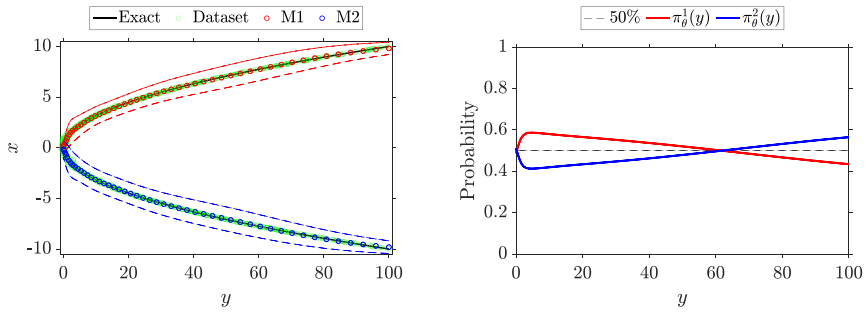
$$= \underset{\theta}{\operatorname{argmin}} \int\int p(y) \left[q(x;\varphi_\theta(y)) \log \left(\frac{q(x;\varphi_\theta(y))}{\frac{p(x)p(y|x;\varphi_\varepsilon)}{p(y)}} \right) \right] \mathrm{d}x\,\mathrm{d}y \tag{25}$$

$$= \underset{\theta}{\operatorname{argmin}} \int\int p(y) q(x;\varphi_\theta(y)) \left[\log \frac{q(x;\varphi_\theta(y))}{p(x)} - (\log p(y|x;\varphi_\varepsilon) - \log p(y)) \right] \mathrm{d}x\,\mathrm{d}y \tag{26}$$

$$\underbrace{=}\ \underset{\theta}{\operatorname{argmin}} \int p(y) \left[\mathbb{E}_q \left[\log \frac{q(x;\varphi_\theta(y))}{p(x)} \right] - \mathbb{E}_q \left[\log p(y|x) \right] - \log p(y)) \right] \mathrm{d}y \tag{27}$$

$$\underbrace{\approx}\ \frac{1}{H \cdot N} \cdot \underset{\theta}{\operatorname{argmin}} \sum_{n=1}^{N} \sum_{i=1}^{H} \left[\log q(x_n^i;\varphi_\theta(y_n)) - \left(\log p(x_n^i) + \log p(y_n|x_n^i) \right) \right], \tag{28}$$

where $\{x_n^i\}_{i=1}^H$ are samples obtained from $q(x_n;\varphi_\theta(y_n))$, and y_n are the given physical measurements. The expression in eq. (27) is known as ELBO, and it is the objective loss function that we will employ in our numerical scheme.



(a) Estimation of the inverse problem by MVAE with two means (b) Probabilities $\pi_\theta(y)$ for each point of the validation dataset

Figure A2. Application of MVAE in an inverse problem with a known solution.

APPENDIX C: ILLUSTRATIVE EXAMPLE OF AN INVERSE PROBLEM WITH TWO SOLUTION BRANCHES

We present a simple benchmark example to illustrate the difference between VAE and MVAE models. We consider a simple inverse problem with a known inverse solution that exhibits two branches. To do this, we select the forward problem $\mathcal{F}(x) = x^2$. Thus, the associated inverse solution is $\mathcal{I}(y) = \pm\sqrt{y}$.

Data generation

We generate 100 samples for training and 100 for validation. Samples are constructed in two steps:

(i) For the training set, we generate 100 samples $\{x_n^{tr}\}_{n=1}^{100}$ randomly from a uniform distribution in the domain $(-10,10)$; for the validation set, we select 100 equally spaced samples $\{x_n^{val}\}_{n=1}^{100}$, where $x_j^{val} = -10 + (j-1)\cdot\Delta$ with $\Delta = 20/99$, and $j = 1,\ldots,100$.

(ii) We generate $\{y_n^{tr}\}_{n=1}^{100}$ by $y_n^{tr} = \mathcal{F}(x_n^{tr}) + \varepsilon_n$, where the additive Gaussian noise $\varepsilon_n \sim \mathcal{N}\left(0, (0.05\mathcal{F}(x_n))^2\right)$ for $n = 1,\ldots,100$.

Results using VAE

Since the data are Gaussian noised, we consider the likelihood $p(y|x)$ in eq. (9) as $\mathcal{N}\left(\mathcal{F}(x), (0.005\mathcal{F}(x))^2\right)$, and the prior $p(x)$ as $\mathcal{U}(-10, 10)$. By taking $q(x;\varphi_\theta(y))$ to be a Gaussian density with $\varphi_\theta(y) = (\mu_\theta(y), \sigma_\theta(y))$, the loss in eq. (28) (see Appendix A) can be simplified as:

$$\theta_{\mathrm{KL}} \approx \underset{\theta}{\operatorname{argmin}} \frac{1}{H\cdot N}\sum_{n=1}^{N}\sum_{i=1}^{H}\left[\log q(x_n^i;\varphi_\theta(y_n)) - \left(\log p(x_n^i) + \log p(y_n|x_n^i)\right)\right] \tag{29}$$

$$= \underset{\theta}{\operatorname{argmin}} \frac{1}{H\cdot N}\sum_{n=1}^{N}\sum_{i=1}^{H}\left[\log\frac{0.05\mathcal{F}(x)}{\sqrt{\sigma_\theta(y_n)}} - \frac{1}{2\sigma_\theta(y_n)}(x_n^i - \mu_\theta(y_n))^2 + \frac{1}{2}\left(\frac{y_n - \mathcal{F}(x_n^i)}{0.05\mathcal{F}(x_n^i)}\right)^2\right]. \tag{30}$$

We consider a fully connected NN and estimate the trainable weights θ. Fig. A1(a) shows the evaluation of the validation set with the inverse problem estimation $\mathcal{I}_\theta(y)$ obtained from the deterministic autoencoder scheme. In this case, we do not quantify the uncertainty because, in a deterministic solution, we obtain only a point estimation. Fig. A1(b) shows the inverse problem solution given by the mean of the PDF obtained by a VAE with loss function in eq. (30) and shows the uncertainty quantification, which we represent with a 95 per cent of the probability region (P.R.) $(\mu_\theta(y) \pm 1.96\cdot\sigma_\theta(y))$.

Using VAEs, we obtain a reasonable estimation of our data with an autoencoder scheme and quantify the uncertainty. However, we obtain only one of the possible branches: the selected branch depends on the initial values given to the NN weights. In order to obtain multiple branches in our estimation and quantify their uncertainty, we propose an autoencoder with *mixture densities*.

Results using MVAE

We consider the PDF $q(x;\varphi_\theta(y), a, b)$ as a truncated Gaussian mixture with two densities in the domain $(-10, 10)$. We obtain θ_{KLT} in eq. (28) (see Appendix A) as:

$$\theta_{\mathrm{KL}} \approx \underset{\theta}{\operatorname{argmin}} \frac{1}{H\cdot N}\sum_{n=1}^{N}\sum_{i=1}^{H}\left[-\log q(x;\varphi_\theta(y), -10, 10) + \frac{1}{2}\left(\frac{y_n - \mathcal{F}(x_n^i)}{0.05\mathcal{F}(x_n^i)}\right)^2 + \log(0.05\mathcal{F}(x))\right]. \tag{31}$$

Fig. A2 shows an NN scheme that provides multiple solutions to the inverse problem with uncertainty quantification. Fig. A2(a) shows the inverse problem solution using MVAE with two truncated Gaussian densities. We obtain two different point estimations given by the means M1=$\mu_\theta^1(y)$, and M2=$\mu_\theta^2(y)$, with a probability region of 95 per cent. We may interpret the weights $\pi_\theta^m(y)$ in Fig. A2(b) as a representation of our confidence in each particular solution of y for $m = 1,\ldots,M$.