Subsurface uncertainty quantification with deep geologic priors: A variational Bayesian framework

Ashutosh Tewari[1] *, *Brent Wheelock*[1], *Jacob Clark*[1], *Dallas Foster*[2], *Matthew Li*[2], *Youssef Marzouk*[2]
[1]*ExxonMobil Research & Engineering Company,* [2] *Massachusetts Institute of Technology*

SUMMARY

Recent years have seen surging interest in the area of geophysical inversion and uncertainty quantification (UQ) using data-driven priors for geologic domains. In particular, deep generative models, e.g., variational autoencoders (VAEs) and generative adversarial networks (GANs), have been employed in a number of studies to either regularize deterministic subsurface inversion, or to act as prior distributions for Bayesian inversion. These models offer a low-dimensional representation of high-dimensional property fields, by capturing intrinsic structural features observed in nature, thus providing a common prior for the concomitant assimilation of data of different modalities such as seismic, fluid flow, electromagnetism, gravity etc. On one hand, this low-dimensionality is an important enabler for Bayesian methods, which are generally deemed impractical for industrial-scale subsurface UQ applications. On the other hand, the computational burden of evaluating expensive forward models and the non-convexity/multi-modality of the resulting inverse problems continue to present a formidable challenge for Bayesian methods. Towards the goal of scalable subsurface Bayesian inversion, we present a framework that considerably mitigates the aforementioned challenges of expensive physics and complex posteriors. The proposed framework exploits the low-dimensionality provided by deep generative priors, and approximates the posterior using deep mixture models via variational inference. The efficacy of our proposal is demonstrated on a joint inverse problem employing gravity and seismic data within a synthetic testbed.

INTRODUCTION

Applications in exploration geophysics often yield ill-posed inverse problems —due to high dimensional parameter fields, information loss in the forward operator, noisy and limited data, etc.— where many different configurations of parameter fields explain the observed geophysical response with equal likelihood. The Bayesian approach to inverse problems provides a principled framework for resolving this ill-posedness, but introduces new complications that make its adoption for large-scale inverse problems challenging. Some of these complications center on the core elements of Bayesian modeling such as the sensitivity to the choice of a prior distribution (Hermans et al., 2017) or difficulty in sampling from the high-dimensional posterior distribution with computationally intensive physical models. Meaningful subsurface priors can be designed by carefully encoding expert knowledge, but the process could be cumbersome and unscalable. Conversely, certain loosely prescriptive priors, e.g. based on Gaussian random fields (Ray and Myer, 2019; Tewari and Trenev, 2019), can be used, but they lack desired expressivity to represent subsurface domains with complex geology. Recent research in subsurface inversion has sought to integrate techniques from computer vision and deep learning to yield more expressive *geologic* priors from data. Deep generative models (DGMs), such as GANs or VAEs have been demonstrated to beneficially regularize deterministic inversion (see e.g., Siahkoohi et al., 2021; Mosser et al., 2020; Liu et al., 2021; Lopez-Alvis et al., 2021). Along the same line, but with the intent of full posterior characterization, we consider a general Bayesian subsurface inversion problem

$$\pi_{\text{post}}\left(z|d_1,\cdots,d_q\right) \propto \pi_{\text{pr}}\left(z\right)\cdot\pi_{\text{lk}}\left(d_1|f_1(z)\right)\cdots\pi_{\text{lk}}\left(d_q|f_q(z)\right), \tag{1}$$

given conditionally independent datasets $d_1, d_2, \ldots, d_q$[†]. We assume access to a generative model that maps a latent space, $z \in \mathbb{R}^m$, to a gridded cell space, $x \in \mathbb{R}^n$ (s.t. $m \ll n$) of physical properties (e.g., ρ, V_p, V_s, etc.) from which different geophysical responses can be modeled. The datasets, d_i, included in Eq. (1) could denote measurements from surface-based geophysics (e.g., gravity, EM, seismic, etc.), measurements from well logs, or even geoscientific interpretations (horizons, faults, salt bodies, etc.). The differentiable mappings, $f_i : \mathbb{R}^m \to \mathbb{R}^{|d_i|}$, allow us to combine the prior information, $\pi_{\text{pr}}(z)$, with the likelihoods, $\pi_{\text{lk}}(d_i|f_i(z))$, of the observed data d_i. Note that f_i is a composite function that first transforms a latent vector, z, to requisite property fields, x, via a DGM, which in turn get mapped to data space via the corresponding forward operator.

A further benefit of using DGM priors is the introduction of a lower-dimensional inversion space, which helps to address the high computational cost of posterior characterization. Nevertheless, standard Bayesian methods may still remain prohibitively expensive for geophysical applications. For instance, Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis-Adjusted Langevin algorithm (MALA) and Hamiltonian Monte Carlo (HMC), have been used in conjunction with DGM priors (see Mosser et al., 2020; Liu et al., 2021; Lopez-Alvis et al., 2021), but these still require prohibitively many forward simulations. As an alternative, there is a growing interest in variational inference (VI) methods (see Zhang and Curtis, 2019, 2021, for reviews) that can be far more efficient. In this work, we present a scalable framework for Bayesian inversion based on DGM priors and VI to address the challenges of (i) realistic prior modeling, (ii) multi-modal posteriors, and (iii) inference with expensive geophysical simulations.

METHODOLOGY

The proposed framework is comprised of three components: 1) a deep generative model for subsurface geology; 2) posterior approximation via VI that leverages flexible mixture models; and 3) multistage assimilation of information from different modalities and computational complexities.

Deep geologic prior

The quality of a data-driven prior for subsurface domains largely

[†] The RHS of Eq. (1) is usually referred to as the *unnormalized*-posterior (denoted as $\tilde{\pi}_{post}$) in the absence of the proportionality constant needed for equality.

DOI:10.1190/image2022-3750132.1

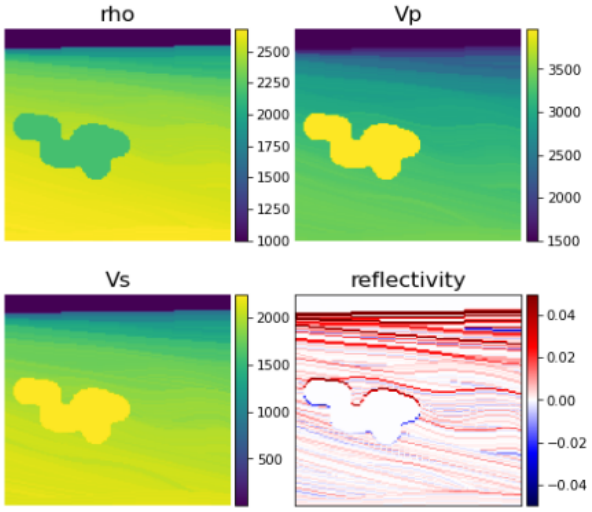

Figure 1: An example of a subsurface model obtained via our synthetic generator, plotted here with 4x vertical exaggeration. The models are comprised of 128×128 cells representing a spatial domain of 25.6 km×6.4 km, in x and z directions, respectively. The inverted properties are density ("rho", kg/m^3), V_p (m/s), and V_s (m/s); Reflectivity (unitless) is shown for illustration purposes. Note, the reflectivity colormap has been artificially capped at ±5% to highlight the subtler interfaces.

depends on the availability of a diverse set of training images representative of the spatial and inter-property correlations as induced by natural geologic processes. In the absence of standard benchmark datasets with desired geologic fidelity, we created a synthetic "geology generator". While our generator does not follow any physical laws, nor do we claim it to be a realistic representation of true subsurface configurations, we do think it generates subsurface models with sufficient geologic resemblance so as to be instructive and useful for the broader research community. Furthermore, if either a better geologic simulation or real subsurface data were substituted, the methods presented would perform as well as they do in the example here. Fig. 1 shows a synthetically generated multi-property subsurface model obtained via our generator. A dataset with ∼16,000 such models are provided in a public repository[‡].

Given these synthetically generated subsurface models, we appeal to DGMs to learn a low-dimensional representation of the property fields. A detailed treatment of this topic is provided by Lopez-Alvis et al. (2021), where the authors experiment with both GANs and VAEs, and suggest that the latter is more suited for the task of inversion. Accordingly, we employ a VAE with a deep convolution architecture; a rough sketch of which is shown in Figure 2. Sparing the details of VAE learning, we just rationalize our choice for the *encoding* and the *decoding* distributions, the two main building blocks of a VAE (Kingma and Welling, 2014). For the former, we proceed with a standard choice of a multivariate Gaussian distribution with diagonal covariance, as shown in the Eq. (2), wherein the dimension is reduced through successive application of strided convolution operations. For the latter, we employ a mixture of Gaussian (MoG) distributions that is independent in the cell space when conditioned on a latent vector as shown in the Eq. (3). For each cell, a k-component MoG is defined using shared but unknown means ($\mu_j \in \mathbb{R}^{k\times 3}$) and fixed variance ($s$).

[‡]https://github.com/BDWheelock/synthetic-salt-geomods/

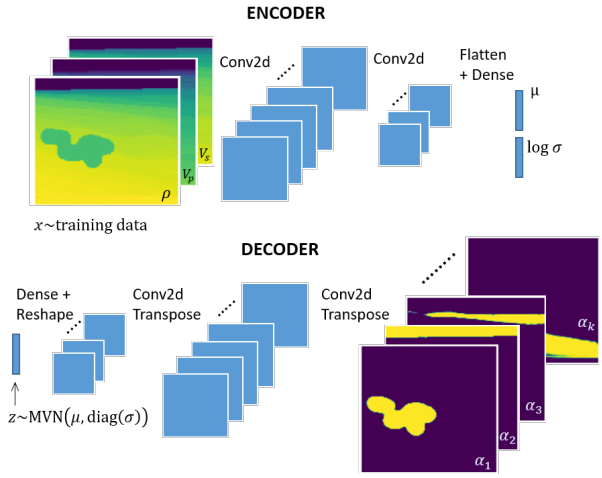

Figure 2: The VAE architecture used in our analysis. The encoder (resp. decoder) employs a series of (2,2) strided convolution (resp. transposed convolution) operations with [64, 32, 16, 8] channels (resp. the reverse of this sequence).

The mixing proportions (α_j) are obtained from the last layer of the decoder given a latent vector z.

$$P_{enc}(z|x) = \text{MVN}\left(z; \mu(x), \sigma(x)\right) \tag{2}$$

$$P_{dec}(x|z) = \prod_{i=1}^{n} \sum_{j=1}^{k} \alpha_j^i(z) \cdot \phi\left(x^i; \mu_j, s\right) \tag{3}$$

This choice of decoding distribution was influenced by the fact that the subsurface fields have natural groups manifested as sedimentary layers with similar properties. Introducing a discrete structure (MoG) helps in explicitly capturing such groups and their spatial patterns. For our case study, we used $k = 15$ components.

Posterior approximation via *mixture of normalizing-flows*

Variational Inference (VI) framework offers a computationally efficient alternative to sampling based methods for posterior characterization; refer to Blei et al., 2017, for a detailed review on this topic. The central idea in VI is to approximate the true posterior distribution shown in Eq. (1) by a tractable (to sample) "variational" distribution, $q(z;\theta)$, parametrized by θ. These parameters are commonly estimated by minimizing the Kullback-Leibler (KL-)divergence between the approximate and the true posterior, which can be computed up to an additive constant. The resulting optimization problem can be equivalently cast as

$$\max\theta \; \mathbb{E}_{q(z;\theta)}[\log \tilde{\pi}_{\text{post}}(()\cdot) - \log q(z;\theta)], \tag{4}$$

where the entity being maximized is referred to as Evidence Lower Bound (ELBO). Note that the ELBO computation relies on the readily accessible unnormalized log-posterior $\tilde{\pi}_{\text{post}}(\cdot)$ (refer to the footnote on page 1). The success of a VI framework largely depends on 1) the expressivity of the variational distribution allowing it to closely approximate the true posterior, and 2) the ability to faithfully compute the expectation in (4) as well as its derivative wrt θ. These two aspects are inherently tied in that the closed-form expectations are possible only for a limited class of variational distributions that make overly restrictive assumptions about the posterior (Blei et al.,

2017, §2.3). Recently, there is an increased interest in expanding the class of variational distributions to include more expressive ones, and in estimating the requisite expectations numerically, e.g., via Monte Carlo. Such attempts, by taking advantage of automatic differentiation and stochastic optimization algorithms, have shown significant uplift in VI performance (Ranganath et al., 2016). Of particular interest to us is a class of distributions referred to as *normalizing-flows*, which uses invertible deep architectures to synthesize highly-flexible distributions (Rezende and Mohamed, 2015). In this work, we further enhance the flexibility of our variational distribution by employing finite *mixtures of normalizing-flows* (MoNF). This enhancement is especially desirable for non-linear inverse problems, such as seismic inversion, that often lead to multimodal posteriors. Let $\sum_i \alpha_i q(z;\theta_i)$ denote a finite mixture variational distribution, wherein each mixing component, $q(z;\theta_i)$ is the RealNVP flow-based distribution proposed by Dinh et al. (2016). With this choice of variation distribution, the VI objective function in (4) can be re-written as

$$\max \sum_i \alpha_i \cdot \mathbb{E}_{q(z;\theta_i)} \left[\log \tilde{\pi}_{\text{post}}(()\cdot) - \log \sum_j \alpha_j q(z;\theta_j) \right], \tag{5}$$

which can then be optimized over component-specific parameters, θ_i. Despite the added flexibility to approximate multimodal posteriors, the use of such mixture distributions has not received commensurate attention until recently (Morningstar et al., 2021), thereby providing a compelling motivation for our work. In the next section, we discuss how MoNF enables a tractable Bayesian inversion in the presence of geophysical models of varying computational complexities.

Multi-Stage Bayesian inversion

Let's rewrite the Bayes' rule in Eq. (1) (without loss of generality) assuming access to two datasets d_c and d_e, i.e.

$$\pi_{\text{post}}(z|d_c,d_e) \propto \pi_{\text{pr}}(z) \cdot \pi_{\text{lk}}(d_c|f_c(z)) \cdot \pi_{\text{lk}}(d_e|f_e(z)). \tag{6}$$

The subscripts signify that the corresponding functional mappings f_c and f_e are computationally "**c**heap" and "**e**xpensive", respectively. For instance, in our case study, f_c involves the gravity operator, while f_e involves the seismic operator. Our ability to characterize the posterior of this form (via VI) will be affected by the computational complexity of f_e and its gradient. To mitigate this problem, we note that the posterior can be updated in stages by assimilating data *sequentially*, provided we can accurately propagate the conditional posteriors between stages. To that end we propose a multi-stage framework with three steps: **1**) sample the posterior exhaustively by assimilating *cheap* information via MCMC, **2**) fit a flexible distribution (e.g. MoNF) on MCMC samples in a post-hoc fashion, and **3**) use the fitted first-stage (cheap) posterior as a prior for the next (expensive) stage and approximate the posterior via VI using another MoNF. Note that in this framework, MoNFs are used for two distinct purposes: (i) to represent the MCMC samples, thus yielding a parametric prior for the next stage, and (ii) to serve as a surrogate distribution to be used in the VI framework in the step 3. Next, we demonstrate the advantage of the proposed multi-stage framework for a joint inverse problem employing gravity and seismic data.

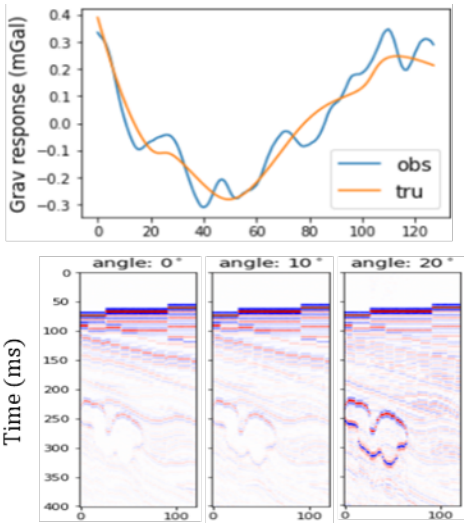

Figure 3: (Top) Observed and True gravitational responses from the ground truth scenario. The former is obtained by adding a laterally correlated Gaussian ($\mu = 0, \sigma = 0.1$) noise to the true response. (Bottom) Observed seismic response using a Ricker source wavelet, with a peak frequency of 10 Hz for all angles. The total record length is 6 seconds with 100 Hz sampling. A Q value of 500, a fixed parameter during inversion, was used to dampen both compressional and shear modes. A zero-mean heteroscedastic Gaussian noise with $\sigma = 10\%$ of the spectral amplitude is added to the true seismic response. Note that while the seismic response is simulated and inverted in the τ-p (or slant-stack) domain, for visualization we perform a τ-p move-out correction to flatten angle gathers.

CASE STUDY

We present a synthetic case study where the goal is to carry out Bayesian inversion with gravity, bathymetry and seismic information, using the framework presented in the previous section. The geologic model shown in Fig. 1 serves as the ground truth, which *wasn't included* in the training dataset for constructing the geologic prior. Particularly, we demonstrate the effectiveness of the proposed multistage Bayesian inversion framework given a strict computational budget on expensive simulations. Geophysical models employed in this study are briefly described next.

Geophysical models

Gravity

The generation of gravity data is based on the method of Talwani et al. (1959) and follows the description in Tewari et al. (2021). The gravitational response for the ground truth model is shown in Fig. 3 (top row), consisting of 128 receivers, one per column of the model domain. The observed gravitational response (to be used for inversion) is synthesized by adding laterally correlated, zero-mean Gaussian noise.

1.5D Seismic

The seismic data are generated using a "trace-by-trace" approximation of the 2D model. That is, for each column of the 3-property image (Fig. 1), we independently calculate a layered-earth response to a compressional plane wave impinging at the top interface with varying angles of incidence (Kennett and Kerry, 1979). Each simulation is non-linear, including interbed multiples, converted waves, transmission effects, and travel-time effects. The seismic response to the ground truth model at different angles of incidence is shown in Fig. 3 (bottom row). Efficient Jacobian calculations, to enable our gradient-based methods, were developed by analytic differen-

tiation of the recursion formula in Kennett and Kerry (1979).

Sequential data assimilation

In the first stage of our framework, we assimilate gravity and bathymetry data. For the gravity data we chose a Gaussian likelihood with pre-specified diagonal covariance, which is a common choice for geophysical data that are often pre-processed through stacking. For the bathymetry data, we used an independent Bernoulli likelihood for each cell since the information is binary. The probability values to specify the corresponding cell-wise Bernoulli probabilities are supplied by the component corresponding to the waterbody from the decoder (see e.g. Fig. 2). Characterizing the posterior via MCMC is tractable in this setting since neither gravity nor bathymetry entails expensive simulations. Fig. 4 shows subsurface realizations (top row) and the corresponding gravity responses of three randomly selected (out of 5000) MCMC samples of the posterior density fields (V_p and V_s fields are omitted for brevity).

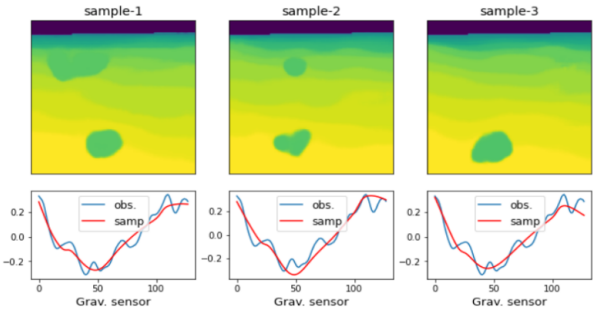

Figure 4: Density fields corresponding to the three random samples obtained via the HMC algorithm that simulates the posterior with gravity and bathymetry information. The samples strictly honor the bathymetry information while loosely fitting the gravity data due to assumed measurement noise.

As often observed in practice, that cheap geophysical surveys are followed by expensive ones, let us assume subsequent seismic data acquisition; the goal thereafter is to update the posterior with this additional information. To highlight the versatility of our framework to fit existing industrial workflows, we further demonstrate that the seismic information may be incorporated in different ways. For instance, one can extract certain subsurface features by interpretation of migrated data, e.g., identifying a reflection event as the "top-of-salt", and use such information to update the posterior. Alternatively, complete seismic information can be assimilated in a full-waveform inversion framework. Therefore, in the second stage, we first account for the top-of-salt information by specifying another set of cell-wise Bernoulli distributions, again parametrized by the corresponding decoder component. To specify the prior, posterior samples from the previous steps are used to learn a MoNF in a density estimation framework (refer to Section *Multi-Stage Bayesian inversion*), and the posterior is again simulated using MCMC. The posterior mean and standard deviation of the density field are shown in Fig. 5 along with posterior predictive gravity response.

In the final (third) stage we assimilate full-waveform information, which involves a relatively expensive seismic wave simulator. Therefore, VI is chosen to approximate the posterior rather than simulating it via MCMC. The prior is again established from the previous stage's MCMC samples in a post-hoc fashion. For VI, we specify a separate MoNF variational distribution with three mixing components (a subjective choice primarily driven by computational constraints). We impose a strict computational budget of 2000 gradient calls and terminate VI optimization when this limit is reached. Note that the required number of gradient calls is equal to the product of the number of mixing components, the number of samples used for MC integration in expression (5), and the number of VI iterations. The results derived from the approximated posterior after this third stage are shown in Fig. 6.

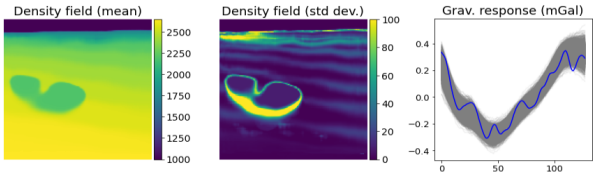

Figure 5: The cell-wise mean (left) and the standard deviation (middle) of the density field obtained from posterior samples from the second stage. Even though the salt body is better localized, the salt-bottom remains uncertain, as does some of the background layering. The posterior predictive gravity response (right plot) still honors the observed gravity response in blue.

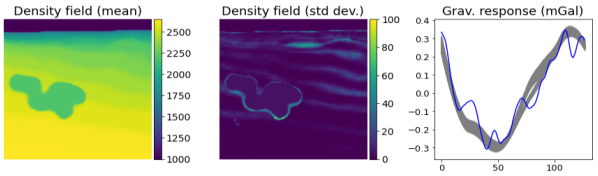

Figure 6: Similar plots as those in Fig. 5, but derived from the samples drawn from the approximated posterior, a 3-component MoNF. With richer information from full seismic waveforms, the salt body is now well reconstructed, while still preserving the previous information (gravity, bathymetry and salt-top).

We conclude the study by demonstrating the evolution of uncertainty of a *quantity of interest* (QoI) through the three stages of data assimilation. A QoI is any quantitative feature of the subsurface that impacts some investment or operational decision. For illustration, we chose salt volume (the physical region of the subsurface constituting a salt body) as our QoI and show the uncertainty in our estimates at each of the aforementioned three stages. Fig. 7 shows the uncertainty estimates of salt volume, which increasingly concentrate as more information is assimilated. We emphasize that if a QoI is not informed by the data from successive geophysical surveys, the illustrated uncertainty reduction over stages will not occur.

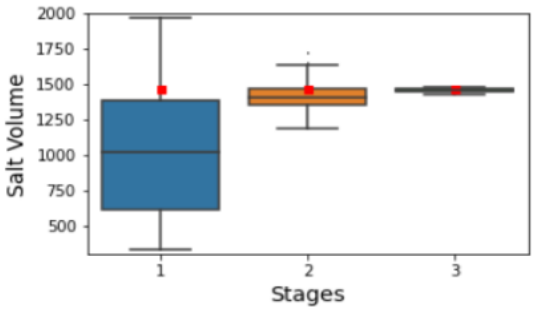

Figure 7: Salt volume distributions (shown as box plots) obtained from posterior samples at each of the three stages of data assimilation. The red marker shows the salt volume for our ground truth scenario. The volume units are number of pixels.

REFERENCES

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

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe, 2017, Variational inference: A review for statisticians: Journal of the American Statistical Association, **112**, 859–877, doi: https://doi.org/10.1080/01621459.2017.1285773.

Dinh, L., J. Sohl-Dickstein, and S. Bengio, 2016, Density estimation using Real NVP: CoRR, abs/1605.08803, http://dblp.uni-trier.de/db/journals/corr/corr1605.html#DinhSB16.

Hermans, T., F. Nguyen, M. Klepikova, A. Dassargues, and J. Caers, 2017, Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using Bayesian evidential learning: Water Resources Research, **54**, 2931–2948, doi: https://doi.org/10.1002/2017WR022135.

Kennett, B. L. N., and N. J. Kerry, 1979, Seismic waves in a stratified halfspace: Geophysical Journal International, **57**, 557–583, doi: https://doi.org/10.1111/j.1365-246X.1979.tb06779.x.

Kingma, D. P., and M. Welling, 2014, Auto-encoding Variational Bayes: 2nd International Conference on Learning Representations, ICLR Conference Track Proceedings, 1–14.

Liu, M., D. Grana, and L. P. de Figueiredo, 2021, Uncertainty quantification in stochastic inversion with dimensionality reduction using variational autoencoder: Geophysics, **87**, no. 2, M43–M58, doi: https://doi.org/10.1190/geo2021-0138.1.

Morningstar, W., S. Vikram, C. Ham, A. Gallagher, and J. Dillon, 2021, Automatic differentiation variational inference with mixtures: Proceedings of Machine Learning Research, **130**, 3250–3258.

Mosser, L., O. Dubrule, and M. J. Blunt, 2018, Stochastic seismic waveform inversion using generative adversarial networks as a geological prior: Mathematical Geosciences, **52**, 53–79, doi: https://doi.org/10.1007/s11004-019-09832-6.

Ranganath, R., D. Tran, and D. M. Blei, 2016, Hierarchical variational models: Proceedings of the 33rd International Conference on International Conference on Machine Learning, **48**, 25682577.

Ray, A., and D. Myer, 2019, Bayesian geophysical inversion with trans-dimensional Gaussian process machine learning: Geophysical Journal International, **217**, 1706–1726, doi: https://doi.org/10.1093/gji/ggz111.

Rezende, D. J., and S. Mohamed, 2015, Variational inference with normalizing flows: Proceedings of the 32nd International Conference on International Conference on Machine Learning, **37**, 15301538.

Siahkoohi, A., G. Rizzuti, and F. J. Herrmann, 2021, Deep Bayesian inference for seismic imaging with tasks: arXiv preprint, arXiv:2110.04825.

Talwani, M., J. L. Worzel, and M. Landisman, 1959, Rapid gravity computations for two-dimensional bodies with application to the Mendocino submarine fracture zone: Journal of Geophysical Research, **64**, 49–59, doi: https://doi.org/10.1029/JZ064i001p00049.

Tewari, A., and D. Trenev, 2019, An ultra low-dimensional inversion of homogeneous geobodies in full waveform inversion: 89th Annual International Meeting, SEG, Expanded Abstracts, 1690–1694, doi: https://doi.org/10.1190/segam2019-3216897.1.

Tewari, A., B. Wheelock, A. Paiva, A. Fathi, and M. S. Cheon, 2021, Towards practical Bayesian inversion of geobodies using geologic priors: First International Meeting for Applied Geoscience & Energy, SEG/AAPG, Expanded Abstracts, 1641–1645, doi: https://doi.org/10.1190/segam2021-3584203.1.

Zhang, X., and A. Curtis, 2019, Seismic tomography using variational inference methods: Journal of Geophysical Research: Solid Earth, **125**, e2019JB018589.

Zhang, X., and A. Curtis, 2021, Bayesian full-waveform inversion with realistic priors: Geophysics, **86**, no. 5, A45–A49, doi: https://doi.org/10.1190/geo2021-0118.1