DOI:10.1190/segam2020-3428223.1

New likelihood functions and level-set prior for Bayesian full-waveform inversion

Matt Dunlop and Yunan Yang, Courant Institute of Mathematical Sciences, New York University

SUMMARY

Seismic full-waveform inversion aims to reconstruct subsurface medium parameters from recorded seismic data. It is solved as a constrained optimization problem in the deterministic approach. Many different objective functions have been proposed to tackle the nonconvexity that originated from the cycle-skipping issues. The analogy between objective functions in the deterministic inversion and likelihood functions in Bayesian inversion motivates us to analyze the noise model each objective function accounts for under the Bayesian inference setting. We also show the existence and well-posedness of their corresponding posterior measures. In particular, the theorem shows that the Wasserstein-type likelihood offers better stability with respect to the noise in the recorded data. Together with an application of the level-set prior, we demonstrate by numerical examples the successful reconstruction from Bayesian full-waveform inversion under the proper choices of the likelihood function and the prior distribution.

INTRODUCTION

Full-waveform inversion (FWI) (Tarantola, 1987) is the state-of-art seismic imaging technique that seeks the optimal parameter by minimizing the discrepancy between the recorded real data and the simulated waveforms produced by the current prediction. The main challenges of FWI include the nonconvexity of the objective function and the existence of the noises that affect the accuracy of the inversion results (Tarantola, 2005). Thus, the resolution analysis and uncertainty quantification of the reconstructions are equally important as solving the optimization problem (Gouveia and Scales, 1998; Fichtner and Trampert, 2011). Under the framework of Bayesian seismic inversion Zhu et al. (2016); Izzatullah et al. (2019), one can utilize the systematic way provided by Bayesian inference to quantify uncertainties in geophysical inverse problems (Zhu et al., 2016; Dashti and Stuart, 2016).

The likelihood function and the prior distribution are the main components of a Bayesian calculation. A realistic noise model is an essential a priori which partially determines the likelihood function in Bayesian inversion. However, quantification of the noise model is nontrivial, and the common additive Gaussian assumption might not be enough to characterize the real uncertainty (Motamed and Appelo, 2019). The shape and curvature of the likelihood surface represent information about the stability of the estimates, whose analogy in the deterministic approach of solving FWI is the objective function. Recently, numerous work on the new class of objective functions from optimal transport (Engquist and Froese, 2014; Engquist et al., 2018) demonstrates the effectiveness in mitigating the cycle-skipping issues in FWI. The subject of optimal transport studies probability measures, so the use of the corresponding Wasserstein metric is natural in statistical inference, particularly in the Bayesian setting (El Moselhy and Marzouk, 2012; Motamed and Appelo, 2019). The connection motivates us to further analyze the Wasserstein metric under the framework of Bayesian inversion (Zhu et al., 2016; Izzatullah et al., 2019).

In this abstract, we first analyze several new likelihood functions and their corresponding noise model. We rigorously define the posterior distributions and study the existence and stability with respect to perturbations to the observed data for the different choices of likelihood. We prove that these posterior measures are stable with respect to perturbations of the observed data measured in different norms. The theoretical analysis demonstrates the advantages of choosing Wasserstein-type likelihood. Besides, we discuss the advantages of using a new type of level-set prior (Kadu et al., 2016; Dunlop et al., 2016), and its high potential for salt body inversion. Numerical results demonstrate the importance of the properly chosen likelihood and priors by combining the mathematical tool and the physics knowledge of the geophysical problem.

BAYESIAN INVERSION

The Bayesian approach combines a probabilistic model for the observed data y, $\mathbb{P}(\mathrm{d}y|u)$, with a probability distribution $\mathbb{P}(\mathrm{d}u)$ representing our prior belief about the unknown u. Bayes theorem then tells us how to construct the posterior distribution $\mathbb{P}(\mathrm{d}u|y)$ of the unknown given the data: formally, if $\mathbb{P}(\mathrm{d}y|u) = \mathbb{P}(y|u)\,\mathrm{d}y$ admits a Lebesgue density,

$$\mathbb{P}(\mathrm{d}u|y) = \frac{\mathbb{P}(y|u)\mathbb{P}(\mathrm{d}u)}{\mathbb{P}(y)}. \tag{1}$$

The probability measure $\mathbb{P}(\mathrm{d}u|y)$ is the solution to the Bayesian inverse problem, rather than a single state as in the deterministic approach. It offers credible bounds on the solution and the uncertainty associated with quantities of interest.

We denote the forward operator (for wave propagation) as $\mathscr{G}$, which maps the model domain X to the data domain Y. The observable data is y. We assume that both the synthetic data $\mathscr{G}(u) \in Y,\ u \in X$ and the observed data $y \in Y$ are functions defined on a spatial domain $D \subseteq \mathbb{R}^d$ equipped with either the counting or Lebesgue measure λ, and temporal domain $T \subseteq \mathbb{R}^s$ equipped with the Lebesgue measure.

Next, we define four potentials/likelihood functions, $\{\Phi_{L^2}, \Phi_{W_2}, \Phi_{\dot{H}^{-1}}, \Phi_M\}\colon X \times Y \to \mathbb{R}^+$, and discuss the corresponding noise models. The choice of norm on Y depends on the choice of potential; here P_σ is an operator that maps functions into probability densities (which is a prerequisite for optimal transport and Wasserstein-based metrics (Engquist and Yang, 2018)).

L^2 misfit and additive Gaussian noise

We first consider the simplest case wherein the loss $\Phi(u;y) = J(\mathscr{G}(u),y)$ is given by the L^2 misfit:

$$\Phi_{L^2}(u;y) = \frac{1}{2}\int_D \|\mathscr{G}(u)(x,\cdot) - y(x,\cdot)\|^2_{L^2(T)}\,\lambda(\mathrm{d}x). \tag{2}$$

This loss arises as a negative log-likelihood by assuming the data is corrupted by additive Gaussian space-time white noise:

$$y = \mathscr{G}(u) + \eta, \quad \eta \sim N(0, I).$$

A general additive Gaussian noise

Another loss function that has been widely used in geophysics is closely related to the integral wavefields misfit (Huang et al., 2014; Liu et al., 2012). Mathematically, the loss function is equivalent to the Sobolev space $\dot{H}^{-1}$ semi-norm in the time domain and L^2 norm in the spatial domain (Yang and Engquist, 2018a). $\|y(x,\cdot)\|_{\dot{H}^{-1}(T)}$ is well-defined only if $\int_T y(x,t)\,\mathrm{d}t = 0$. Since zero-frequency component is often removed from seismic data, the use of $\dot{H}^{-1}$ here is proper.

$$\Phi_{\dot{H}^{-1}}(u;y) = \frac{1}{2}\int_D \|\mathscr{G}(u)(x,\cdot) - y(x,\cdot)\|^2_{\dot{H}^{-1}(T)}\,\lambda(\mathrm{d}x). \quad (3)$$

$\Phi_{\dot{H}^{-1}}$ has a similar data model as Φ_{L^2}, except instead of assuming that the noise is white, temporal correlations are assumed:

$$y = \mathscr{G}(u) + \eta, \quad \eta \sim N(0, \Gamma)$$

where $\Gamma = -\Delta_T$ is the negative Laplacian on the temporal variable. By changing the noise assumption on the temporal correlations, one can obtain a more general class of loss functions that are closely related to (semi-)norms of the Sobolev space.

W_2 loss and the multiplicative noise

In this subsection, we consider a Wasserstein loss function, which can be viewed as an unnormalized state-dependent multiplicative noise loss in the small noise limit.

In the absence of an explicit data model, one typically introduces a scalar parameter $\beta > 0$, often referred to as the inverse temperature, and work with $\beta\Phi(u;y)$ in place of $\Phi(u;y)$. As the calibration, the parameter may either be chosen empirically (Syring and Martin, 2018) or treated as a hyperparameter as part of the inverse problem (Zhu et al., 2016). We do not discuss the choice of β here and assume it to be fixed.

We first introduce the Wasserstein loss function, which comes from optimal transport theory. The core of the subject is on the optimal plan that maps one probability distribution on a domain X into another one on domain Y, intending to minimize the total transport cost of a given cost function. The transport cost function $c(x,y)$ maps pairs $(x,y) \in X \times Y$ to $\mathbb{R} \cup \{+\infty\}$, which denotes the cost of transporting one unit mass from location x to y. If $c(x,y) = |x-y|^2$, the optimal transport cost between probability measure ν_1, ν_2 gives the following *squared* quadratic Wasserstein distance:

$$W_2^2(f,g) = \inf_{T_{\nu_1,\nu_2} \in \mathscr{M}} \int_{\mathbb{R}^n} |x - T_{\nu_1,\nu_2}(x)|^2\, f(x)\,\mathrm{d}x.$$

where $f = d\nu_1$ and $g = d\nu_2$, $\mathscr{M}$ is the set of all measure-preserving maps that rearrange the distribution ν_1 into ν_2. The W_2 distance has been a popular choice of objective function in FWI to mitigate the local minima issues since its first proposal (Engquist and Froese, 2014).

In order to evaluate the Wasserstein distance between the data and output of the forward map, we must first transform these into probability densities with respect to the temporal variable. Hence given a scalar function $\sigma : \mathbb{R} \to \mathbb{R}^+$, we define the normalization operator P_σ on functions $y : D \times T \to \mathbb{R}$ by

$$(P_\sigma y)(x,t) = \frac{1}{Z_\sigma(x)}\sigma(y(x,t)), \quad Z_\sigma(x) = \int_T \sigma(y(x,t'))\,\mathrm{d}t'.$$

Given this operator, we then define the Wasserstein loss by

$$\Phi_{W_2}(u;y) = \frac{1}{2}\int_D W_2\left((P_\sigma\mathscr{G}(u))(x,\cdot), (P_\sigma y)(x,\cdot)\right)^2 \lambda(\mathrm{d}x).$$

Note that Φ_{W_2} is not normalized in the sense that $\exp(-\Phi_{W_2}(u,\cdot))$ does not define a probability density with respect to some measure on Y, and it does not appear to correspond to a particular data model even if it were normalized. However, via linearization of the W_2 distance (Villani, 2003), we can describe an approximate data model that this loss corresponds to in the limit of small observational noise. Assume that there is $\eta : D \times T \to \mathbb{R}$ such that

$$P_\sigma y = (1+\eta)P_\sigma\mathscr{G}(u),$$

$$\int_T \eta(x,t)(P_\sigma\mathscr{G}(u))(x,t)\,\mathrm{d}t = 0 \quad \text{for all} \quad x \in D.$$

By the linearity between the W_2 metric and the weighted $\dot{H}^{-1}$ norm (Villani, 2003), for small $\|\eta(x,\cdot)\|_{\dot{H}^{-1}(P_\sigma\mathscr{G}(u))}$,

$$\begin{aligned}\Phi_{W_2}(u;y) &\approx \frac{1}{2}\int_D \|\eta(x,t)\|^2_{\dot{H}^{-1}(P_\sigma\mathscr{G}(u))}\,\lambda(\mathrm{d}x)\\ &= \frac{1}{2}\int_D \left\| \frac{(P_\sigma\mathscr{G}(u))(x,\cdot) - (P_\sigma y)(x,\cdot)}{(P_\sigma\mathscr{G}(u))(x,\cdot)} \right\|^2_{\dot{H}^{-1}(P_\sigma\mathscr{G}(u))} \lambda(\mathrm{d}x).\end{aligned} \quad (4)$$

This is the negative logarithm of an unnormalized Gaussian density $N(1, \mathscr{L}(u))$, for some operator $\mathscr{L}(u)$ defined below, evaluated at the ratio $P_\sigma y / P_\sigma\mathscr{G}(u)$. This suggests that the Wasserstein loss could be thought of as asymptotically coming from the state-dependent multiplicative noise data model

$$P_\sigma y = \eta \cdot P_\sigma\mathscr{G}(u), \quad \eta | u \sim N(1, \mathscr{L}(u))$$

where $\mathscr{L}(u) : D(\mathscr{L}(u)) \to L^2(D; L^2(T))$ is defined by

$$\mathscr{L}(u)\varphi = -\frac{1}{P_\sigma\mathscr{G}(u)}\nabla_T \cdot (P_\sigma\mathscr{G}(u)\nabla_T\varphi),$$

where $D(\mathscr{L}(u)) = \left\{\varphi \in L^2(D; H^2(T)) \mid \int_T \varphi(P_\sigma\mathscr{G}(u))\,\mathrm{d}t = 0\right\}$ and ∇_T is the gradient with respect to the temporal variable.

A general multiplicative noise

The relation of the Wasserstein loss to a multiplicative noise model is unexpected. The use of a loss function corresponding to a multiplicative noise model was not uncommon in other inverse problems (Isaac et al., 2015; Iglesias et al., 2018):

$$\Phi_M(u;y) = \frac{1}{2}\int_D \left\| \frac{(P_\sigma\mathscr{G}(u))(x,\cdot) - (P_\sigma y)(x,\cdot)}{(P_\sigma y)(x,\cdot)} \right\|^2_{L^2(T)} \lambda(\mathrm{d}x), \quad (5)$$

which can be viewed either as arising from the model

$$P_\sigma y = \eta \cdot P_\sigma\mathscr{G}(u), \quad 1/\eta \sim N(1, I),$$

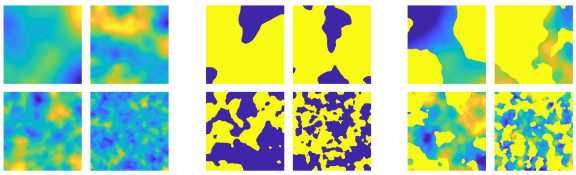

Figure 1: Example of independent samples from (left) plain Gaussian prior, (middle) plain level set prior and (right) mixed level set prior. The underlying Gaussian random fields have Matérn covariance with the regularity and amplitude parameters fixed, and the length-scale is decreased from left-to-right, top-to-bottom within each block.

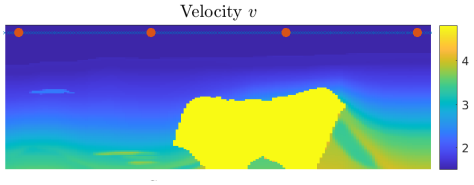

Figure 2: The true salt velocity field v we aim to infer, the location of the four sources $\{s_j\}$ and the set of receivers on which the solution is measured at each time

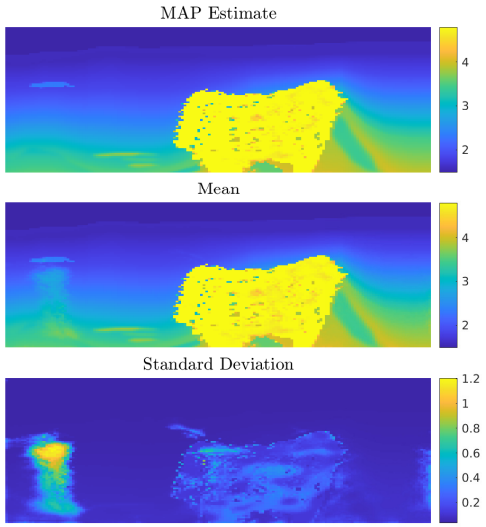

Figure 3: (Top) The MAP estimate, (middle) mean and (bottom) standard deviation.

or alternatively, from informally assuming the size of the noise is proportional to the size of the observed data. Note that the data and output of the forward map do not need to be probability densities here. However, for the stability of the resulting posterior, the forward map must be bounded away from zero; see (Dunlop, 2019) for a discussion. The condition can be ensured by using the same operator, P_σ. Additionally, one may prefer to use a model wherein $\eta \sim N(1, I)$ rather than its reciprocal; see Dunlop and Yang (2020) for more discussions.

Existence and well-posedness

We establish existence and well-posedness of the corresponding (Gibbs) posterior distributions (Dunlop and Yang, 2020):

Theorem 1 (Existence)**.** *Let π_0 be a Borel probability measure on X. Then for any choice $\Phi \in \{\Phi_{L^2}, \Phi_{H^{-1}}, \Phi_{W_2}, \Phi_M\}$,*

$$Z_\Phi(y) = \int_X \exp(-\Phi(u;y))\, \pi_0(\mathrm{d}u)$$

is strictly positive and finite, and

$$\pi_\Phi^y(\mathrm{d}u) := \frac{1}{Z_\Phi(y)} \exp\left(-\Phi(u;y)\right)\, \pi_0(\mathrm{d}u)$$

defines a Radon probability measure on X.

Theorem 2 (Well-posedness)**.** *Let π_0 be a Borel probability measure on X. Choose $\Phi \in \{\Phi_{L^2}, \Phi_{H^{-1}}, \Phi_{W_2}, \Phi_M\}$. Under mild assumptions, there exists $C_\Phi(r) > 0$ such that for all $y, y' \in Y$ with $\|y\|_{L^\infty(D;L^\infty(T))}, \|y'\|_{L^\infty(D;L^\infty(T))} < r$,*

$$d_{\mathrm{H}}(\pi_\Phi^y, \pi_\Phi^{y'}) \le C_\Phi(r)\|y - y'\|_Y.$$

Remark 1. The norm Y in Theorem 2 is H^{-1} for $\Phi_{H^{-1}}$ and Φ_{W_2}, but L^2 for Φ_{L^2}, Φ_M. The stability result implies that W_2- and H^{-1}-based likelihood functions are more robust with respect to high-frequency data noise, which was also analyzed by (Engquist et al., 2018) under the deterministic setting.

LEVEL-SET PRIOR FOR SALT INVERSION

Standard parameters of interest in FWI are technically unknown functionals defined on the spatial domain D. For this infinite-dimensional problem, Bayesian inversion aims to recover a field, in which the prior distribution is often chosen to impose properties such as regularity and length-scale on samples. Gaussian priors are often used to impose continuity and certain smoothness. However, the assumption is not always physical, especially for seismic problems where reflections dominate the real data.

Under such circumstances, non-Gaussian priors are more physical to be used. In particular, we introduce a new type of level-set prior for inference of salt model inversion, which we express as nonlinear transformations of Gaussian fields. Given a Gaussian measure $\nu_0 = N(m, C)$ and scalar values $u_+, u_- \in \mathbb{R}$, one could define a prior measure by the pushforward

$$\pi_0 = F^\sharp \nu_0, \quad F(v)(x) = u_+ \mathbb{1}_{v(x)>0} + u_- \mathbb{1}_{v(x)\le 0}.$$

That is, π_0 is the law of the thresholded Gaussian field $F(v), v \sim \nu_0$: samples from π_0 take the values u_+, u_- almost everywhere, with interface between the values given by the level set $\{v(x) = 0\}$. Such priors have been studied previously from a nonparametric Bayesian perspective (Iglesias et al., 2016; Dunlop et al., 2016)

Alternatively, one may desire some combination of the above

10.1190/segam2020-3428223.1

and plain Gaussian priors. For example, given a product Gaussian measure $\nu_0 = N(m_1, C_1) \times N(m_2, C_2)$, one could define a mixed level set prior:

$$\pi_0 = F^{\#}\nu_0, \quad F(v,w)(x) = u_+ \mathbb{1}_{v(x)>0} + w(x)\mathbb{1}_{v(x)\leq 0}.$$

Examples of samples are shown in Figure 1, which can be generalized to multiple interfaces by the vector level set method (Bertozzi et al., 2018).

Level set methods have been considered previously for deterministic FWI (Kadu et al., 2016), where the objective functions are similar to what we use here when seeking the modes (MAP estimates) of the posterior. However, one key difference is the form of regularization. In essence, there the solution is assumed to have a radial basis expansion. So regularization is performed by restricting to a low dimensional space spanned by the basis. In contrast, here, the regularization will be via the addition of an explicit Sobolev penalty to the loss arising from the underlying Gaussian. Note also that the Bayesian method provides much more than just MAP estimates, allowing, for example, computation of uncertainties in the solution, and propagation of these uncertainties to error bounds on quantities of interest.

NUMERICAL EXAMPLE

We consider the recovery of a salt model as illustrated in Figure 2 via the use of the mixed level set prior and the Wasserstein loss Φ_{W_2}. The continuous background field is assumed known, and we infer the level set function as well as the velocity in the salt region. We use a small number (four) of sources in order to illustrate better the uncertainty arising in the solution. Rather than the velocity c itself, the inversion is performed on the parameter $m = 1/c^2$, and it is this parameter upon which we place the mixed level set prior. The Laplace approximation to the posterior is used as a computationally tractable alternative to sampling, which approximates the posterior with a Gaussian distribution centered around the MAP point: $\pi_\Phi^y \approx N(u_\Phi, C_\Phi)$, where

$$u_\Phi = \arg\min_{u\in X} \Phi(u;y) + \frac{1}{2}\langle u, C_0^{-1}u\rangle, \quad C_\Phi^{-1} = \nabla_u^2\Phi(u_\Phi;y) + C_0^{-1}.$$

Computationally, a low-rank approximation to the Hessian $\nabla_u^2\Phi$ is made via randomized SVD. The diagonal of the regularized inverse C_Φ is approximated with a stochastic method. Additionally, the indicator functions in the definition of the mixed level set method are replaced with mollified versions in order to preserve the differentiability of the forward map.

Figure 3 shows the MAP estimate $1/\sqrt{F(u_\Phi)}$ for the velocity, as well as the pointwise mean and standard deviation. Note that since the inversion is performed on the parameter m, the nonlinearity of the map $1/\sqrt{F(\cdot)}$ means that the distribution on the velocity will not be Gaussian, and hence the MAP estimate and mean do not coincide. We see that both the MAP estimate and the mean recover the shape of the salt region and its velocity value. There is a moderate level of uncertainty in the salt region, and a high level of uncertainty in the region below the small inclusion on the left of the domain; the latter is to be expected, as it has been observed that recovery is typically more difficult below reflectors (Yang and Engquist, 2018b).

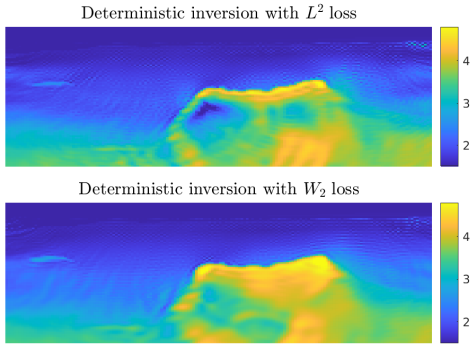

Figure 4: Reconstructions arising from deterministic inversion using (top) L^2 loss and (bottom) W_2 loss.

For comparison with existing deterministic methods, Figure 4 shows the reconstructions obtained by directly optimizing the loss functions Φ_{L^2} and Φ_{W_2}. Though the W_2 loss outperforms the L^2 loss, both fail to recover the entire salt body and the velocity value within that region. Hence, even before taking uncertainty into account, the MAP estimate above appears to be preferable and may be found for approximately the same computational cost.

CONCLUSION

We considered four potentials originating from the objective functions for deterministic full-waveform inversion and showed how they could be viewed as arising, or approximately arising, from specific noise models for the data. We then formulated corresponding Bayesian inverse problems after quantifying our prior beliefs with a probability measure, leading to the construction of the Gibbs posterior distributions. The stability of these posteriors regarding perturbations in the data is shown here. In particular, the posterior arising from the Wasserstein loss is more stable under the corruption of the data by high-frequency noise. In the numerical example, we considered the use of a level set prior, which leads to superior reconstructions compared to standard deterministic approaches. Moreover, the Bayesian approach provides an interpretable form of uncertainty in the solution.

ACKNOWLEDGEMENT

This work is partially supported by National Science Foundation grant DMS-1913129, and the U.S. Department of Energy Office of Science, Advanced Scientific Computing Research (ASCR), Scientific Discovery through Advanced Computing (SciDAC) program. We thank Prof. Georg Stadler and Prof. Andrew Stuart for constructive discussions.

REFERENCES

Bertozzi, A. L., X. Luo, A. M. Stuart, and K. C. Zygalakis, 2018, Uncertainty quantification in graph-based classification of high dimensional data: SIAM/ASA Journal on Uncertainty Quantification, **6**, 568–595.

Dashti, M., and A. M. Stuart, 2013, The Bayesian approach to inverse problems: arXiv preprint arXiv:1302.6989.

Dunlop, M. M., 2019, Multiplicative noise in Bayesian inverse problems: Well-posedness and consistency of MAP estimators: arXiv preprint arXiv:1910.14632.

Dunlop, M. M., M. A. Iglesias, and A. M. Stuart, 2017, Hierarchical Bayesian level set inversion: Statistics and Computing, **27**, 1555–1584.

Dunlop, M. M., and Y. Yang, 2020, Stability of Gibbs posteriors from the Wasserstein loss for Bayesian full waveform inversion: arXiv preprint arXiv:2004.03730.

El Moselhy, T. A., and Y. M. Marzouk, 2012, Bayesian inference with optimal maps: Journal of Computational Physics, **231**, 7815–7850.

Engquist, B., and B. D. Froese, 2013, Application of the Wasserstein metric to seismic signals: arXiv preprint arXiv:1311.4581.

Ren, K., and Y. Yang, 2019, The quadratic Wasserstein metric for inverse data matching: arXiv preprint arXiv:1911.06911.

Engquist, B., and Y. Yang, 2018, Seismic inversion and the data normalization for optimal transport: arXiv preprint arXiv:1810.08686.

Fichtner, A., and J. Trampert, 2011, Resolution analysis in full waveform inversion: Geophysical Journal International, **187**, 1604–1624.

Gouveia, W. P., and J. A. Scales, 1998, Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis: Journal of Geophysical Research: Solid Earth, **103**, 2759–2779.

Huang, G., H. Wang, and H. Ren, 2014, Two new gradient precondition schemes for full waveform inversion: arXiv preprint arXiv:1406.1864.

Iglesias, M., M. Park, and M. V. Tretyakov, 2018, Bayesian inversion in resin transfer molding: Inverse Problems, **34**, 105002.

Iglesias, M. A., Y. Lu and A. M. Stuart, 2015, A Bayesian level set method for geometric inverse problems: arXiv preprint arXiv:1504.00313.

Isaac, T., N. Petra, G. Stadler, and O. Ghattas, 2015, Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet: Journal of Computational Physics, **296**, 348–368.

Izzatullah, M., T. van Leeuwen, and D. Peter, 2019, Bayesian uncertainty estimation for full waveform inversion: A numerical study: SEG Technical Program, Expanded Abstracts, 1685–1689.

Kadu, A., T. van Leeuwen, and W. A. Mulder, 2016, Salt reconstruction in full-waveform inversion with a parametric level-set method: IEEE Transactions on Computational Imaging, **3**, 305–315.

Liu, J., H. Chauris, and H. Calandra, 2011, The normalized integration method-an alternative to full waveform inversion?: Near Surface 2011-17th EAGE European Meeting of Environmental and Engineering Geophysics, cp-253.

Motamed, M., and D. Appelo, 2019, Wasserstein metric-driven Bayesian inversion with applications to signal processing: International Journal for Uncertainty Quantification, **9**, 395-414.

Syring, N., and R. Martin, 2019, Calibrating general posterior credible regions: Biometrika, **106**, 479–486.

Tarantola, A., 1987, Method for data fitting and model parameter estimation, Inversion problem theory: Elsevier Science.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM, Vol. **89**.

Villani, C., 2003, Topics in optimal transportation (No. 58): American Mathematical Soc.

Yang, Y., and B. Engquist, 2018, Analysis of optimal transport and related misfit functions in full-waveform inversion: Geophysics, **83**, no. 1, A7–A12.

Yang, Y., and B. Engquist, 2018, Model recovery below reflectors by optimal-transport FWI: SEG Technical Program, Expanded Abstracts, 1178–1182.

Zhu, H., S. Li, S. Fomel, G. Stadler, and O. Ghattas, 2016, A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration: Geophysics, **81**, no. 5, R307–R323.