Sampling-Free Bayesian Inference for Local Refinement in Linear Inversion Problems With a Latent Target Property

Charlotte Semin-Sanchis and Odd Kolbjørnsen

***Abstract*— We present a sampling-free probabilistic inversion of latent target property based on the principles of expectation propagation where we estimate the joint distribution of the target variable in a local region. The prior model matches the prior distribution in the local-focused region but integrates our model parameters outside the focus region using approximate distributions. The approximate distribution includes large spatial structure information while maintaining the dimension of the inversion small. In addition, we map and solve the inversion into a new feature space where we can exclude components where the data have little influence, thereby decreasing the dimensionality of the inversion, and therefore, the inversion runtime. We test the method on seismic amplitude-versus-offset (AVO) inversion examples for the prediction of facies classes, as well as on the estimation of vuggy porosity in computed tomography (CT) images of core from a carbonate reservoir. We demonstrate that our method achieves good-quality predictions while significantly reducing the computational demand, making it particularly interesting to run large-scale inversion studies.**

***Index Terms*— Bayesian inference, expectation propagation, Gaussian process, linear inversion problems, rocks, seismic inversion, variational inference.**

I. Introduction

MANY ill-posed inverse problems that arise through remote sensing and indirect measurements, such as seismic and medical imaging, conform to a structure where the parameters sought have a two-level structure. The acquired data (**d**) respond to a particular property of the investigated media, the intermediate property (**m**), which itself responds to another property of the media, the target property (**r**). In X-ray tomography, the observations (**d**) are line integrals of the absorption (**m**), whereas the target property (**r**) is tissue type [1] or density. In a seismic inversion, the objective is the estimation of rock properties, such as rock type, porosity, and fluid filling (**r**), but the observations (**d**), the seismic data in this case, are the response to intermediate properties, such as sound velocity and density (**m**) [2].

Manuscript received 23 December 2022; revised 26 June 2023; accepted 14 July 2023. Date of publication 3 August 2023; date of current version 29 August 2023. This work was supported in part by the Geophysical Inversion to Geology (GIG) Consortium, in part by Aker BP, in part by ConocoPhillips, in part by Equinor, in part by Lundin, in part by TotalEnergies, in part by Vår Energi, in part by Wintershall Dea, and in part by Sharp Reflections. *(Corresponding author: Charlotte Semin-Sanchis.)*

Charlotte Semin-Sanchis is with the Norwegian Computing Center, 0373 Oslo, Norway (e-mail: cha@nr.no).

Odd Kolbjørnsen is with the Group for Statistics and Data Science, Department of Mathematics, University of Oslo, 0851 Oslo, Norway (e-mail: oddkol@math.uio.no).

Digital Object Identifier 10.1109/TGRS.2023.3301717

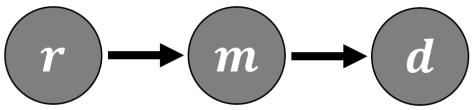

Fig. 1. Relation between target property (**r**), intermediate property (**m**), and observations (**d**).

The objective of inversion problems is to make inferences about **r** based on the data **d**. Thus, the problem formulation is split into two parts: the first one involves a local relationship that connects the target property to the intermediate local property, while the second one is a global relationship that connects the intermediate property to the observations. The relations are illustrated in Fig. 1. The first relation that connects the target and intermediate properties is local, meaning that there is a point-to-point relationship between them. This implies that a target property in one location only affects the intermediate property in that specific location. This relation might be theoretical or empirical and need not be unique, i.e., different formulations of this relation may exist. The parameter may be discrete or continuous. In the case of the second relation, there exists a physical relationship connecting the intermediate property to the observations. This is considered as global as the intermediate property in one location impacts the observed data within a potentially large neighborhood around that specific location.

A common way to handle this problem is to use a two-step approach of inversion. First, an inverse problem is solved to estimate the intermediate property, next this property is inverted into the target property. The benefit of this approach is that it separates the two problems. One that only considers the physics of the indirect measurements and another that considers the relationship between properties of the media. The two fields of knowledge are often just partly overlapping, and as a result, the separation is dictated by the boundaries of the respective disciplines which are convenient. The challenge of the two-step approach is to transfer the knowledge and uncertainty related to the measurements all the way to the target property. In this process, the spatial constraints and allowances are the hardest. The intermediate property has a point-to-point relation to the target property, but this is in general not true for the estimate of the intermediate property obtained by inverting the physics. This estimate lacks the resolution a point-to-point relation requires, and also the

uncertainty in the inversion is not well represented using a pointwise description.

An alternative is to employ a one-step approach, which estimates the target property $\mathbf{r}$ directly from the data $\mathbf{d}$. In such cases, the Bayesian framework [3] is particularly well suited to propagate knowledge and uncertainty. Let us use the notation p to denote a generic probability distribution. Then, the initial knowledge of the target properties is summarized in the prior distribution $p(\mathbf{r})$, the link between the intermediate property and the target property is given by a likelihood $p(\mathbf{m} \mid \mathbf{r})$ and the link to the physical observation is given by a second likelihood $p(\mathbf{d} \mid \mathbf{m})$. The posterior information about the target variable $p(\mathbf{r} \mid \mathbf{d})$ is given by the marginal posterior, through the relation

$$p(\mathbf{r} \mid \mathbf{d}) \propto \int p(\mathbf{r})\, p(\mathbf{m} \mid \mathbf{r})\, p(\mathbf{d} \mid \mathbf{m}) d\mathbf{m}. \tag{1}$$

However, for large-scale problems with spatial interactions, the models are often complex, and it is a challenge to quantify the information contained by the posterior distribution. Brute force methods, such as Markov chain Monte Carlo (McMC), are time consuming and memory inefficient. The way around this consists of finding an adequate approximation, which is computationally feasible.

The general framework we present in this article is built on the one-step approach, wherein we directly invert the data into target properties. For seismic inversion, the framework developed for the direct inversion of seismic amplitudes into rock properties. Quantitative prediction from seismic is historically defined through the two-step approach [4], [5]. Grana and Della Rossa [2] and Buland et al. [6] show how the uncertainty from the seismic inversion can be transferred into uncertainty about the rock properties for the continuous and discrete parameters, respectively. Although these methods transfer the uncertainty correctly in a point-to-point setting, they do not utilize the spatial structure of the problem. Reference [7] proposes a model based on state-space models for unobserved categorical variables [8]. Other authors improve the approximation of the likelihood model to be used in this setting [9], [10], and the speed of these methods relies on efficient computations of model and data utilizing 1-D dependencies. Models that include 2-D or 3-D dependencies in the problem either settle for a solution without any assessment of the uncertainty [11] or utilize a computationally intensive McMC procedure as an approximation method [12], [13], [14]. An alternative to McMC is the sampling-free Bayesian inference approach using machine learning, described in [15]. This involves using a trained neural network to estimate the statistical properties of the posterior distribution directly from the data. This method assesses one large-scale property, while our method focuses on local target properties that can be aggregated to estimate larger scale properties. The machine learning approach has fast execution, but the construction of a training dataset may be computationally intensive. Finally, another possibility for Bayesian inference is variational inference, which tends to be faster and easier to scale to large data to McMC, as discussed in [16] and [17].

In geosciences, [18] uses variational inference to assess the posterior distribution. A general weakness of variational inference is that the uncertainty in the estimates tends to be underestimated [19]. Another sampling-free approach is expectation propagation presented in [20]. For the problem of seismic inversion, this philosophy is followed in [21], [22], and [23]. The aim of these papers is to compute the marginal posterior distribution of a local property, rather than solving the full joint posterior distribution, which uses approximate distributions to perform integrals. In the local approach, a key factor is that detailed analysis is performed using a neighborhood of the location considered. However, the neighborhood needs to be large enough to capture all the information provided by the local property in the observations. This has a computational cost since such algorithm complexity grows exponentially with the neighborhood size [24].

The method proposed in this article extends the well-established Bayesian inversion method [24], a one-step approach using variational inference for direct inversion of the data into target properties. Rimstad and Omre [24] have already demonstrated their capability to estimate high-quality posterior predictions in seismic inversion, including a comparison with examples using McMC methods considered as the best possible approach. Our contribution focuses on the computational effort. Therefore, we develop [24] further with the objective of enhancing computational efficiency while preserving the same level of quality. In our approach, we introduce three zones of localization—the target region, which we want to investigate, the neighboring region, where the distribution of the parameters is influenced by the parameters of the target region and the rest. We denote the approach "focused inversion" to highlight the increase of details in the approximation when zooming in on the target area. In the case where the prior is stationary, we show that the problem has the form of convolutional computations, i.e., the same set of computations is utilized in every spatial location. The formulation enables us to include large spatial structure information into the inversion process while maintaining the size of the target region small. Moreover, by combining this with the dimension reduction of noninformative features, we effectively reduce the computational runtime.

In Section II, we provide the theoretical framework for focused inversion. First, we present the likelihood integral obtained by integrating all intermediate properties in the neighboring region. Then, we detail how we approximate the likelihood from intermediate properties to preserve information while allowing for dimension reduction. We first use the expectation propagation principle to capture the spatial structure information from a wide prior model and then use the likelihood principle to validate the mapping of the linear model to a new and more compact representation, facilitating dimension reduction.

In Section III, we show two synthetic examples: the seismic inversion for lithology-fluid predictions and porosity predictions in rock computed tomography (CT) images. The first example uses a relatively uncomplicated model and a small dataset, such that it becomes possible to invert the data with benchmark methods within a reasonable time. Thus, we can compare the performance of the focused inversion against state-of-the-art methods. The size and complexity of

the second example do not allow for exact computations. In this example, we test the sensitivity of focused inversion parametrizations and evaluate both the runtime speedup and the quality of the predictions against the approximated ground truth for a 2-D inversion problem.

II. Model Specification and Approximated Likelihood

A. Likelihood Model

The focused inversion method extends the existing, well-established Bayesian amplitude-versus-offset (AVO) inversion methods [22], [24], [25] based on multivariate Gaussian distributions. Hence, it is appropriate to make similar assumptions and employ the multivariate Gaussian distributions. While it is feasible to relax the Gaussian model assumptions, as shown in [26], doing so would necessitate using sampling procedures to estimate the posterior distribution, which can significantly increase the computational demand in high dimensions settings. This approach contradicts the purpose of the method presented in this article.

Thus, we use the following notation: $N(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the probability density function for a multidimensional Gaussian random variable $\mathbf{x}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, while the transpose of a vector $\mathbf{x}$ will be denoted $\mathbf{x}^{\mathbf{T}}$. We consider a linear data acquisition model where the link between the observed data $\mathbf{d} = (d_1, d_2, \ldots, d_{Ld})^{\mathbf{T}}$ and the intermediate property $\mathbf{m} = (m_1, m_2, \ldots, m_{Lm})^{\mathbf{T}}$ is given by the modeling matrix $\mathbf{G} \in \mathbb{R}^{L_d \times L_m}$, potentially obscured with an $R^{L_m \times 1}$ error term $\mathbf{e} \sim N(\mathbf{e}; \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{e}})$ such that

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{e}. \tag{2}$$

From this relationship, it follows that the likelihood function of the data given the intermediate variable is

$$p(\mathbf{d} \,|\, \mathbf{m}) = N(\mathbf{d}; \mathbf{G}\mathbf{m}, \boldsymbol{\Sigma}_{\mathbf{e}}). \tag{3}$$

On the contrary, the relationship between the target property $\mathbf{r}$ and the intermediate property $\mathbf{m}$ is not necessarily linear and can either be theoretical or empirical. However, this relationship is local, meaning that target properties in a specific location i influence only the intermediate property at that location. Note that both $\mathbf{r}$ and $\mathbf{m}$ can contain multiple properties in one location, as discussed in Section I. Thus, if we denote by Γ the mapping function from target to intermediate properties, we have the local relation

$$\mathbf{m}_i = \boldsymbol{\Gamma}(\mathbf{r}_i, \mathbf{e}_i) \tag{4}$$

where $\mathbf{e}_i$ is additional local parameters considered nuisance parameters as they will influence the value of the intermediate variable but are not of interest by themself. Randomizing over these nuisance parameters, we approximate the likelihood of the intermediate variable given the target parameter by the Gaussian distribution

$$p(\mathbf{m} \,|\, \mathbf{r}) = N\left(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m}|\mathbf{r}}, \boldsymbol{\Sigma}_{\mathbf{m}|\mathbf{r}}\right) \tag{5}$$

where the conditional mean and covariance matrix are dependent on the target variable in a nontrivial way. We discuss how these can be assessed in the following.

To calculate the posterior distribution of target properties in (1), we need to integrate both likelihoods $p(\mathbf{d} \,|\, \mathbf{m})$ and $p(\mathbf{m} \,|\, \mathbf{r})$. In general, we cannot assume that the relation in (2) is local and that the mapping in (4) is linear. As a result, the likelihood in (3) cannot be considered local and the likelihood in (5) cannot be considered linear either. Consequently, the joint problem of integrating both likelihoods can be considered as both nonlocal due to (2) and nonlinear due to (5). This implies that the likelihood cannot be deduced analytically, and sampling approaches are prohibited due to the nonlocal nature of the problem. If this is the case, a simpler approach may be preferable. Thus, focusing the efforts on solving a subset of the target variables makes it possible to quantify the solution. We denote the focused subset of the target variables by

$$\mathbf{r}_w = (r_1, r_2, \ldots, r_w)^{\mathbf{T}} \tag{6}$$

where r_k are the unobservable target properties in the selected set of spatial locations. The subset is often referred to as the configuration window. The relevant likelihood is then $p(\mathbf{d} \,|\, \mathbf{r}_w)$. This likelihood is obtained by integrating all intermediate parameters

$$p(\mathbf{d} \,|\, \mathbf{r}_w) \propto \int p(\mathbf{d} \,|\, \mathbf{m}) p(\mathbf{m} \,|\, \mathbf{r}_w) d\mathbf{m}. \tag{7}$$

Here, we introduce a reasonable approximation wherein the target variable $\mathbf{r}_w$ impacts the intermediate variable $\mathbf{m}$ in a limited region R larger than w. Thus, if we select R such that this assumption holds, i.e., $p(\mathbf{m} \,|\, \mathbf{m}_R, \mathbf{r}_w) \sim p(\mathbf{m} \,|\, \mathbf{m}_R)$, then we use the factorization

$$p(\mathbf{m} \,|\, \mathbf{r}_w) = p(\mathbf{m} \,|\, \mathbf{m}_R) p(\mathbf{m}_R \,|\, \mathbf{r}_w) \tag{8}$$

to limit the likelihood integral in (7) such that

$$p(\mathbf{d} \,|\, \mathbf{r}_w) \propto \int p(\mathbf{d} \,|\, \mathbf{m}_R) p(\mathbf{m}_R \,|\, \mathbf{r}_w) d\mathbf{m}_R \tag{9}$$

where $\mathbf{m}_R$ is the local representation of the intermediate property in region R. To summarize, the three levels in our approximation are the window w being the target region, the region R in which the intermediate variable $\mathbf{m}$ is influenced by the target variable $\mathbf{r}_w$, and the rest. Thus, to solve the integral in (7), we first approximate $p(\mathbf{d} \,|\, \mathbf{m}_R)$ and then the integral (9). This will provide the approximation of the posterior distribution through the Bayes formula

$$p(\mathbf{r}_w|\mathbf{d}) \propto p(\mathbf{d} \,|\, \mathbf{r}_w) p(\mathbf{r}_w). \tag{10}$$

The current approximation and methodology are common to other established methods, such as [24]. However, this article's contribution lies in the inclusion of a greater level of detail in the approximation within these regions while reducing the computational demand. To achieve this goal, we will discuss how we approximate $p(\mathbf{d} \,|\, \mathbf{m}_R)$ and the integral (9) in Sections II-B and II-C. More specifically, we will provide an approach, which gives insight into which information is carried through the different levels.

B. Integration of the Outer Region

In this section, we detail how we approximate $p(\mathbf{d} \,|\, \mathbf{m}_R)$ for any region R. To preserve information, we integrate the intermediate variables of the outer region, that is set of locations outside the region R. We use the following procedure: we first consider the inversion of the intermediate variable. If we

assume $\mathbf{m}$ to follow a multivariate Gaussian distribution such that $\mathbf{m} \sim N(\mathbf{m}; \boldsymbol{\mu}_\mathbf{m}, \boldsymbol{\Sigma}_\mathbf{m})$, then the posterior distribution of the intermediate variable, $\mathbf{m}|\mathbf{d} \sim N(\mathbf{m}; \boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}}, \boldsymbol{\Sigma}_{\mathbf{m}|\mathbf{d}})$, under the linear model of expression (2) is Gaussian with the parameters

$$\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}} = \boldsymbol{\mu}_\mathbf{m} + \boldsymbol{\Sigma}_\mathbf{m}\mathbf{G}^\mathbf{T}\left(\mathbf{G}\boldsymbol{\Sigma}_\mathbf{m}\mathbf{G}^\mathbf{T} + \boldsymbol{\Sigma}_\mathbf{e}\right)^{-1}(\mathbf{d} - \mathbf{G}\boldsymbol{\mu}_\mathbf{m}) \quad (11)$$

$$\boldsymbol{\Sigma}_{\mathbf{m}|\mathbf{d}} = \boldsymbol{\Sigma}_\mathbf{m} - \boldsymbol{\Sigma}_\mathbf{m}\mathbf{G}^\mathbf{T}\left(\mathbf{G}\boldsymbol{\Sigma}_\mathbf{m}\mathbf{G}^\mathbf{T} + \boldsymbol{\Sigma}_\mathbf{e}\right)^{-1}\mathbf{G}\boldsymbol{\Sigma}_\mathbf{m}. \quad (12)$$

With explicit expressions now available for both the prior and the posterior distributions, we can use them to approximate the likelihood in region R. We first need to set the prior and posterior Gaussian parameters for the region R. To achieve this, we use the principle of expectation propagation and let the prior and posterior means, $\boldsymbol{\mu}_{\mathbf{m}_R}$ and $\boldsymbol{\mu}_{\mathbf{m}_R|\mathbf{d}}$, and covariances, $\boldsymbol{\Sigma}_{\mathbf{m}_R}$ and $\boldsymbol{\Sigma}_{\mathbf{m}_R|\mathbf{d}}$, match the means and covariances of the intermediate variable limited to the region R. Finally, we use a common technique to estimate the likelihood $p(\mathbf{d}\,|\,\mathbf{m}_R)$ that consists of approximating it by the ratio of the posterior to the prior like in [7], which gives

$$p(\mathbf{d}\,|\,\mathbf{m}_R) \propto \frac{N\left(\mathbf{m}_R; \boldsymbol{\mu}_{\mathbf{m}_R|\mathbf{d}}, \boldsymbol{\Sigma}_{\mathbf{m}_R|\mathbf{d}}\right)}{N\left(\mathbf{m}_R; \boldsymbol{\mu}_{\mathbf{m}_R}, \boldsymbol{\Sigma}_{\mathbf{m}_R}\right)}. \quad (13)$$

This is the approximation we use to integrate the outer layer. This approach also allows us to focus the inversion on a region R that is narrower than the one affected by the target variable $\mathbf{r}_w$.

Although (13) is well defined, it is convenient to have the likelihood in terms of a set of independent observations in later integrations. Therefore, we propose a mapping to a new feature space.

C. Mapping to a New Feature Space

In this section, we look for a transformation $\tilde{\mathbf{G}}$ that honors the likelihood estimated in (13). Since both the prior and the posterior in (13) are Gaussian, then the likelihood principle [27] allows us to compute a set of linear-Gaussian observations $\tilde{\mathbf{d}}$ and the associated error $\tilde{\mathbf{e}}$ with the same information content as (13). Thus, the transformed observations $\tilde{\mathbf{d}}$ also denoted as observed features are related to the intermediate property in the region R by the relation

$$\tilde{\mathbf{d}} = \tilde{\mathbf{G}}\mathbf{m}_R + \tilde{\mathbf{e}}. \quad (14)$$

The local representation of the intermediate property is the vector $\mathbf{m}_R \in \mathbf{R}^{L_R}$, $\tilde{\mathbf{G}}$ is a $L_{\tilde{\mathbf{d}}} \times L_R$ feature matrix while the error $\tilde{\mathbf{e}}$, and observed features $\tilde{\mathbf{d}}$ are $\mathbf{R}^{L_{\tilde{\mathbf{d}}}}$-dimensional vectors. The notation $\tilde{\cdot}$ is used for any variable expressed in the feature space. The error term $\tilde{\mathbf{e}}$ follows a Gaussian distribution such that $\tilde{\mathbf{e}} \sim N(\tilde{\mathbf{e}}; \mathbf{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}})$ with $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ being a diagonal $L_{\tilde{\mathbf{d}}} \times L_{\tilde{\mathbf{d}}}$ matrix. Note that the error relates to the observed features, and therefore, the covariance structure differs from the one in the data domain in (2). Furthermore, expressions for $\tilde{\mathbf{d}}$, $\tilde{\mathbf{G}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$, are derived in the Appendix.

Features in the new representation are ordered according to a decreasing signal-to-noise ratio. We can choose to leave out the features which bring the least information and thereby obtain a more compact representation. If no dimension reduction is applied, the dimension of the observed features matches the number of elements in the region R, i.e., $L_{\tilde{\mathbf{d}}} = L_R$. Otherwise, the observation representation is more compact, and $1 \leq L_{\tilde{\mathbf{d}}} < L_R$, which is typical for inverse problems since there are often unresolved features.

To summarize, we have derived a set of features $\tilde{\mathbf{d}}$ whose dimension can be chosen smaller than the dimension of the observations in the data domain, i.e., $L_{\tilde{\mathbf{d}}} < L_R$, and whose likelihood $p(\tilde{\mathbf{d}}\,|\,\mathbf{m}_R) = p(\mathbf{d}\,|\,\mathbf{m}_R)$. Thus, we have presented an approximation of the likelihood $p(\mathbf{d}\,|\,\mathbf{m}_R)$ where the focused inversion approach allows for a reduced size of the region R, as well as for dimension reduction through the mapping to the feature space. These two parameters control the dimensionality of the focused inversion problem and therefore, they can both contribute to the reduction of the inversion runtime.

D. Local Inversion and Approximated Posterior Model

Now, let us consider the approximation of the integral (9). Having captured the global properties in the feature space, we are now free to select an improved approximation for $p(\mathbf{m}_R|\mathbf{r}_w)$ that we will denote by $p^*(\mathbf{m}_R|\mathbf{r}_w)$. By construction, we know that the distribution of $\mathbf{m}_R$ is influenced by $\mathbf{r}_w$. For this reason, we again approximate $p(\mathbf{m}_R|\mathbf{r}_w)$ by a Gaussian distribution

$$p^*(\mathbf{m}_R\,|\,\mathbf{r}_w) = N(\mathbf{m}_R; \boldsymbol{\mu}_R(\mathbf{r}_w), \boldsymbol{\Sigma}_R(\mathbf{r}_w)). \quad (15)$$

The functions for $\boldsymbol{\mu}_R(\mathbf{r}_w)$ and $\boldsymbol{\Sigma}_R(\mathbf{r}_w)$ are case specific. If $\mathbf{r}_w$ only contains discrete components, the resulting distribution of $\mathbf{m}_R$ is a mixture of Gaussian distributions. Reference [28] discusses how to merge the modes of the mixture distribution to provide the optimal distribution for expectation propagation. A general case is considered in [22]. Current progress in deep learning would suggest normalizing flows [29] as one option for approximating $\boldsymbol{\mu}_R(\mathbf{r}_w)$ and $\boldsymbol{\Sigma}_R(\mathbf{r}_w)$.

Using Gaussian linear computations, we can now derive the likelihood of observed features given the target properties $p(\tilde{\mathbf{d}}|\mathbf{r}_w)$, which is the specificity of the focused inversion approach

$$p\left(\tilde{\mathbf{d}}|\mathbf{r}_w\right) = N\left(\tilde{\mathbf{d}}; \tilde{\mu}_{\tilde{\mathbf{d}}|\mathbf{f}_w}, \tilde{\boldsymbol{\Sigma}}_{\tilde{\mathbf{d}}|\mathbf{f}_w}\right) \quad (16)$$

where

$$\tilde{\boldsymbol{\mu}}_{\tilde{\mathbf{d}}|\mathbf{f}_w} = \tilde{\mathbf{G}}\boldsymbol{\mu}_R(\mathbf{r}_w) \quad (17)$$

$$\tilde{\boldsymbol{\Sigma}}_{\tilde{\mathbf{d}}|\mathbf{f}_w} = \tilde{\mathbf{G}}\boldsymbol{\Sigma}_R(\mathbf{r}_w)\tilde{\mathbf{G}}^\mathbf{T} + \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}. \quad (18)$$

Expressions for the observed features $\tilde{\mathbf{d}}$ and the features modeling and error covariance matrices $\tilde{\mathbf{G}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ are derived in Appendix A. It appears clearly that the lower dimensional elements $\tilde{\mathbf{d}}$, $\tilde{\mathbf{G}}$, and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ yield lower dimensional model parameters in (17) and (18), and therefore, faster computations of the likelihood in (16).

E. Approximate Posterior Distribution

Having derived the approximate likelihood, we use relation (10) to derive the posterior distribution. In the case where the target variable is discrete, there is an explicit expression for the prior and the likelihood can be multiplied to provide the posterior. When the target variable is continuous there are multiple options since the dimension of the problem is small.

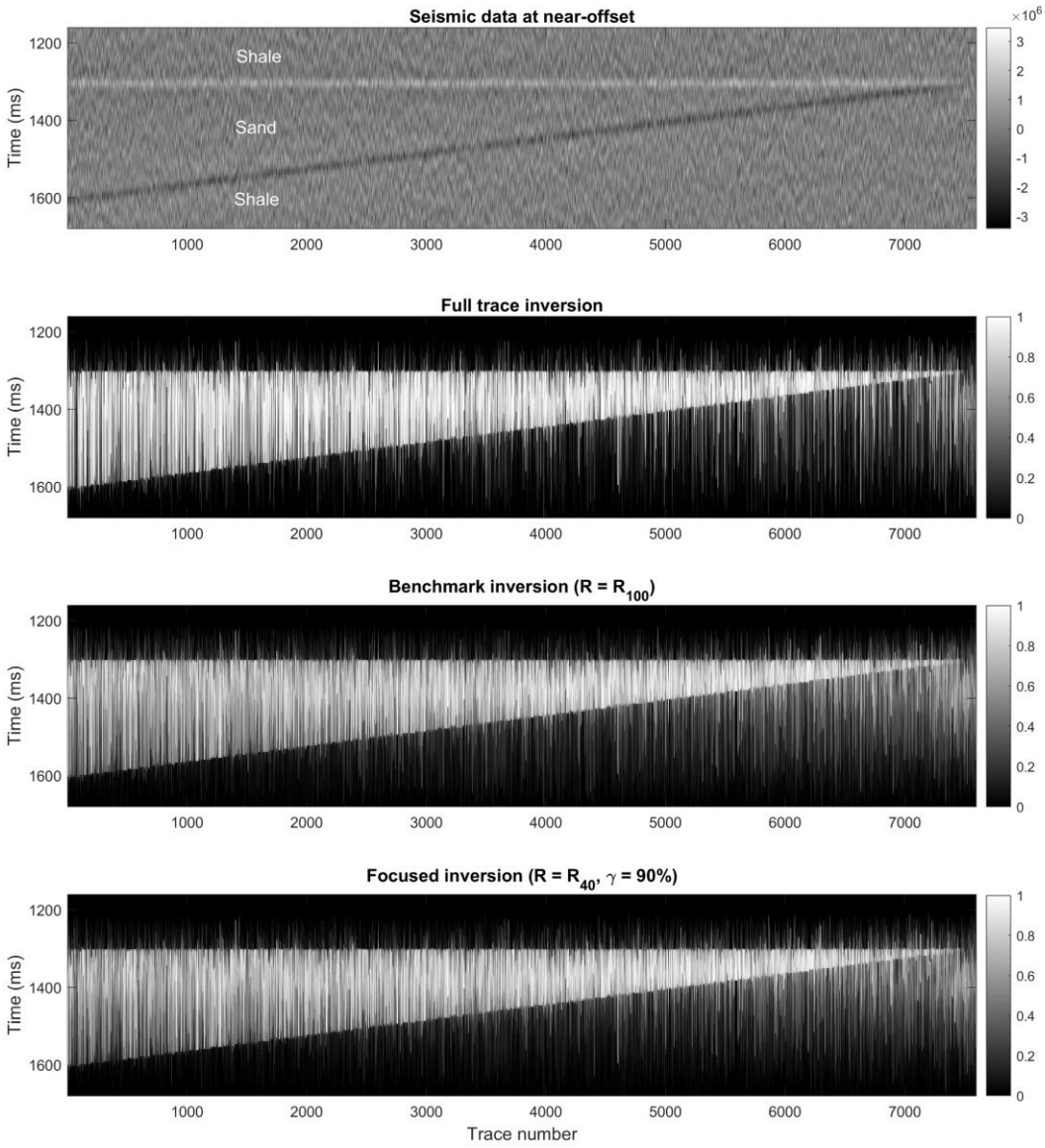

Fig. 2. Top: seismic wedge data at near-offset (12° acquisition angle) consisting of a layer of sand trapped between two layers of shale. Bottom: posterior probability for sand estimated with three different methods—full-trace inversion (middle top), unfocused reference inversion with R_{100} (middle bottom), and focused inversion with R_{40} (40% of nominal R_{100}) and $\gamma = 90\%$ (bottom).

One option is to sample the prior distribution and use the approximate likelihood as an important weight.

The methodology detailed in this article focuses so far on the computation of the properties for one specific window. A natural generalization is to apply the methodology to a sliding window that will loop through an entire dataset, thereby repeating the computations for all windows. Focusing on the integration of the outer region, we see that the same global linear inversion can be reused for all windows w and corresponding regions R.

F. Stationary Case

Many simplifications occur if the problem has stationary properties. Consider now the case where the linear map $\mathbf{G}$ is a convolution and the random fluctuations around the prior mean and the error distribution are stationary. Then the prior and posterior used for three outer regions can be computed using Fourier analysis, see [30] for details about seismic inversion.

In the stationary case, the feature vectors in (29) and the error matrix in (36) will not change when the window location moves, and thus, these need only be computed once. The input to (33) can be derived for all translations of the window using convolutions. The transformed data are then derived using the computation in (33) per location and per feature. Thus, the full set of operations scales with the number of features times the number of grid elements.

In the stationary case, the likelihood in (15) is stationary as well and we can reuse the full likelihood in (16). An interesting feature of the approach is that even if the likelihood is stationary, the prior of $\mathbf{r}_w$ need not be stationary. So, we can use stationary computations also to derive results that are nonstationary. There are, however, some limitations to this because if there are strong nonstationarities in the distribution of $\mathbf{r}_w$ as the window moves, the optimal distribution for expectation propagation will not be stationary.

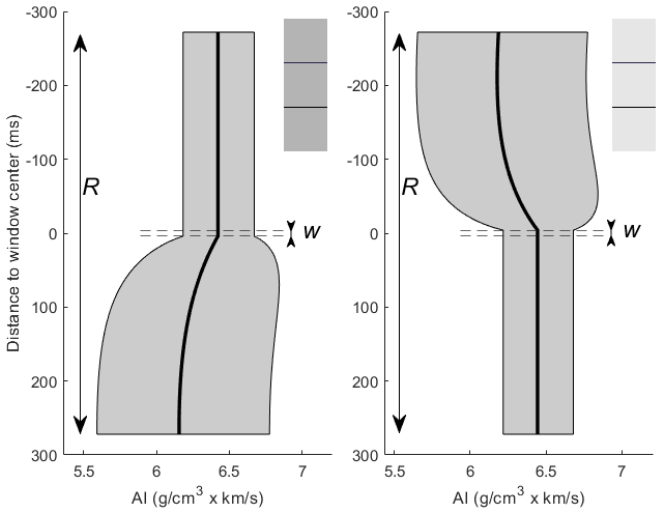

Fig. 3. Two example distributions of acoustic impedance (AI). The three samples' lithology configuration $\mathbf{r}_w$ represented by the boxes in the top right corner, consisting of shale 1 only (left-hand side) and shale 2 only (right-hand side). The distance to the window center is given along the vertical axis. The thick black line is the distribution mean while the uncertainty between 2.5% and 97.5% is represented by the area in gray.

G. Discussion

Even though the distribution of the nuisance parameters in expression (4) is unknown, it is possible to estimate the likelihood $p(\mathbf{m} \mid \mathbf{r})$ using collocated samples of the target and the intermediate variables. In our work, we estimate the mean and covariance parameters in expression (5). Note that even though the relation in expression (4) is local, the covariance in expression (5) will not be diagonal since the nuisance parameters most often are spatially correlated.

The question about how to choose the region R given the window w is essential in the approach. An implicit assumption, which is frequently made in local computations is $R = w$, see [10]. This simplification is not required in a local framework and is in general not true (an example is given in the following). The important impact of introducing the region R is that we consider the intermediate property to be influenced by the target property in a larger region than the window w. To assess the size of R, it is natural to compare the mean and standard deviation of the two distributions, $p(\mathbf{m}|\mathbf{m}_R, \mathbf{r}_w)$ and $p(\mathbf{m}|\mathbf{m}_R)$. The region R should be then selected such that there is no applicable difference between the two models outside R.

To highlight the importance of having $R \neq w$, we use an example from seismic inversion. In this example, we consider a geological model where a layer of sand simulating the reservoir is trapped between two layers of shale, as illustrated in Fig. 2 (top figure). We denote by shale 1 the shale layer above the sand and by shale 2 the shale layer below the sand. In our notation, $\mathbf{r}_w$ is the lithology configuration of a three samples window, where the potential values $\mathbf{r}_k$ are the categorical variables {sand, shale 1, and shale 2}. The parameter $\mathbf{m}_w$ contains the rock physics properties of each lithology $\mathbf{r}_k$ within the window, located at the center of $\mathbf{m}$ components in Fig. 3. For the sake of argument, we assume that the distribution of rock physics properties for shale 1 and shale 2 is identical but different from the sand one, as illustrated in Fig. 4. The distribution in question is $p(\mathbf{m}|\mathbf{m}_w, \mathbf{r}_w)$. Consider the components of $\mathbf{m}$ which are below the window, i.e., from 4 to 300 ms in Fig. 3. If $\mathbf{r}_w$ contains shale 1 (left-hand side of Fig. 3), then sand can still be expected below the window. On the contrary, if $\mathbf{r}_w$ contains shale 2 (right-hand side of Fig. 3), we have already gone through the transition to shale at the bottom of the sand layer, and therefore, we can no longer expect sand below the window. Thus, there is a difference in the distributions of $\mathbf{m}$, which cannot be explained by the values of rock physics properties in the window $\mathbf{m}_w$.

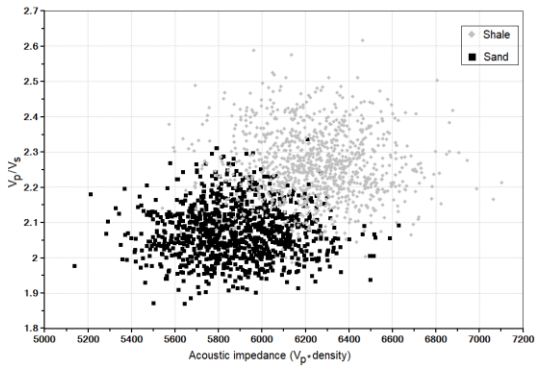

Fig. 4. Rock physics parameters scatter plot for shale in black and sand in gray. Shale and sand rock physics parameters have equal variance, but different means.

Even though the likelihood approximation is local, the results can be integrated into a spatial setting. One such case is provided in [24]. Another way to integrate the results in a global analysis is to use the approximation in the framework for "projection approximation of the joint likelihood" developed in [10]. The argument of integral approximation through expectation propagation provides an alternative to the blunt truncation used in the original formulation.

III. Synthetic Examples

In this section, we first consider two seismic inversion examples. By opting for relatively uncomplicated models and small datasets, it becomes possible to compute within a reasonable time an ideal inversion based on the exact likelihood. Thus, we will be able to demonstrate the benefits of the focused inversion in terms of posterior distribution quality and reduced computational effort. Then, we consider a second example where we predict porosity in rock CT images. An ideal inversion based on the exact likelihood would be too time consuming, and to the best of our knowledge, there are no well-established methods for the estimation of maximum a posteriori probability. For this reason, we do not compare it with other inversion methods. Instead, we want to test the sensitivity of focused inversion to different parametrizations and evaluate the quality of the predictions against the approximated ground truth for a 2-D inversion problem.

A. Assessment of the Posterior Model

As mentioned earlier, the focused inversion method is controlled by two parameters: the size of the region R and the amount of dimensionality reduction. In the synthetic examples,

we test different combinations of parametrization, where the following holds.

1) R_p is smaller or equal to the nominal region R, where the subscript p indicates the retained percentage. p is set to increase from 10% to 100% by a step of 10% such that R_{40} corresponds to 40% of the nominal region R and $R_{100} = R$. The choice of the nominal region R depends on the application.
2) γ, as defined in (38) in the Appendix, is the fraction of explained variance included in the features. γ is set to increase from 40% to 100% by a step of 10%. Note that a full rank ($\gamma = 100\%$) discards eigenvalues of 1 such that the dimensionality of the inversion is reduced, which leads to $L_{\tilde{\mathbf{d}}} < L_R$.

The posterior distribution obtained for each (R_p, γ) combination will be one of the test distributions p^*. Furthermore, the quality of these distributions will be assessed by the Kullback–Leibler divergence score and the average prediction power defined in (39) and (40) in the Appendix.

B. Seismic AVO Inversion

We apply our methodology to seismic AVO inversion that consists of lithology predictions from seismic data. In seismic surveys, a seismic source generates sound waves that travel to the subsurface and bounce back to the surface after reflecting off rock layers, where they are measured by receivers. The acquisition angle which is determined by the distance between the source and the receiver, and the depth of a target, corresponds to the angle of incidence of the reflected wave. The amplitudes of the seismic data, which capture the subsurface response, vary with the acquisition angle and are crucial for AVO analysis.

For this example, we compare focused inversion against two other inversion methods. The first one is the full-trace inversion. To predict lithologies, we set up all the possible lithologies trace configurations, calculate the likelihood of a seismic trace given each possible trace configuration and eventually calculate the marginal posterior probability of each categorical data of K at any depth for this trace. The full-trace inversion is the most exact inversion method in the sense that it does not make any approximations, and therefore, computes the exact likelihood and posterior probabilities, given the available data. It requires, however, to loop through and invert the covariance matrix for a very large number of trace configurations, which gives a very long processing runtime. Therefore, this method is an unrealistic solution at an industrial scale and alternatives like focused inversion need to be used. However, because our test example is a simple test case limited to a small seismic dataset, it is possible to run the full-trace inversion. We will use it as our benchmark method and compare the performance of focused and unfocused inversions against it. The full-trace inversion will be the reference distribution p used to evaluate $p^*(x)$ in the Kullback–Leibler divergence score, $D(p(x)||p^*(x))$.

The second inversion method is the one presented in [24] that we will refer to as the "unfocused" method in this article. This method is already in use in the industry, and therefore, is a realistic alternative to compare against for industrial-scale projects. This method also estimates the target properties contained in a window w and defines the distribution of $\mathbf{m}$ for a region R. For comparison purposes, we choose the same length for w and test the same intermediate regions R_p as in focused inversion. Note, however, that there is no possibility for dimension reduction here. This will be the other test distribution p^*.

To test the inversion methods, we generate a synthetic wedge dataset that consists of a sand layer simulating a reservoir trapped between two layers of shale for three different marine acquisition angles 12°, 20°, and 28°. The top horizon corresponds to the transition from shale to sand at 1300 ms. The second horizon corresponds to a transition from sand to shale that gradually moves from 1600 to 1300 ms by a time increment of 4 ms such that the sand zone thickness gradually decreases from 300 to 0 ms (i.e., no zone). In this case, we have two lithologies as target properties, i.e., $\mathbf{r}_k \in K = \{\text{sand, shale}\}$, each characterized by rock physics properties $\mathbf{m}_k$ like the elastic P- and S-waves velocities V_p and V_s, density parameter ρ. To generate the dataset, we sample the rock physics parameters $\mathbf{m}_i$ and colored Gaussian noise $\boldsymbol{e}_i$ at each time location $\boldsymbol{i}$ such that the signal-to-noise ratio is an average of 6 at near offset (12° angle). Rock physics properties for shale and sand are chosen to have equal variance but different means, as shown in Fig. 4. The noise model is the sum of a white component and a correlated component function of the seismic wavelet. Finally, we use the forward model in (2) to generate the synthetic data wedge for acquisition angles 12°, 20°, and 28° representing the near, mid, and far offset datasets, respectively. To mimic the earth absorption, we downscale the wavelet amplitude at a far offset such that the signal-to-noise ratio is smaller. For each thickness of the sand zone, 100 traces are independently sampled such that a total of 7600 realizations are used as input for inversion. The 2-D flatten | representation of the nearest offset (12°) is shown in Fig. 2 (top figure). The seismic reflections from one layer to another remain, however, relatively small, and the dataset relatively noisy.

As a prior model for the inversion, we set up a stratigraphic model that assumes the top reservoir reflector to be between 1200 and 1400 ms, which corresponds to +/−100 ms around the true location. The bottom reservoir reflector is assumed to be between 1200 and 1700 ms which also corresponds to 100 ms above the shallowest and below the deepest true location. We choose to use a target property window with five samples, $\mathbf{r}_w = (r_{-2}, r_{-1}, r_0, r_1, r_2)^T$ with $r_k \in K$, which makes a total of $2^5 = 32$ possible configurations. This number is reduced to 21 possible configurations after enforcing a strict ordering of the stratigraphic layers, e.g., no possibility to go back to sand below a transition from sand to shale. Finally, we choose R to be equal to the window w length plus half wavelet length above and below the window.

We assess the quality of focused and unfocused inversion methods for their respective parametrization with the Kullback–Leibler divergence score summarized in Table I and the average prediction power in Table II. In Tables I and II, the columns show the scores obtained as a function of region R_p. The top seven rows show the scores for focused inversion with rank γ ranging from 40% to 100%, while the bottom row shows the scores for unfocused inversion. Note that the

TABLE I
KULLBACK–LEIBLER DIVERGENCE ($\times 10^{-3}$)

		Region R_p									
		R_{10}	R_{20}	R_{30}	R_{40}	R_{50}	R_{60}	R_{70}	R_{80}	R_{90}	R_{100}
Focused inversion, Relative rank γ (%)	40	7.60	6.06	10.3	9.06	9.59	8.49	8.95	9.55	10.3	10.0
	50	5.52	5.98	5.00	5.60	7.88	7.93	7.44	6.95	7.23	6.70
	60	5.08	5.53	4.83	4.42	4.16	4.22	4.12	4.05	6.22	5.90
	70	4.97	5.22	4.77	4.33	4.10	4.07	4.00	3.97	4.24	5.88
	80	4.54	4.27	4.53	4.20	4.05	4.00	3.88	3.88	3.84	3.76
	90	4.41	4.02	3.79	3.67	3.88	3.72	3.63	3.59	3.60	3.61
	100	4.34	3.92	3.61	3.47	3.36	3.31	3.29	3.29	3.28	3.25
Unfocused inversion		5.81	5.39	4.93	4.64	4.31	4.10	3.92	3.68	3.52	3.37

TABLE II
AVERAGE PREDICTION POWER ($\times 10^{-1}$)

		Region R_p									
		R_{10}	R_{20}	R_{30}	R_{40}	R_{50}	R_{60}	R_{70}	R_{80}	R_{90}	R_{100}
Focused inversion, Relative rank γ (%)	40	8.60	8.59	8.62	8.65	8.67	8.62	8.64	8.59	8.79	8.71
	50	8.77	8.74	8.76	8.73	8.71	8.70	8.81	8.84	8.79	8.89
	60	8.81	8.80	8.92	8.91	8.90	8.90	8.88	8.85	8.81	8.90
	70	8.82	8.90	8.92	8.91	8.91	8.90	8.88	8.85	8.83	8.91
	80	8.94	8.93	8.93	8.92	8.91	8.90	8.89	8.86	8.94	8.92
	90	8.94	8.94	8.94	8.93	8.92	8.91	8.97	8.96	8.95	8.93
	100	8.99	8.99	8.99	8.99	8.99	8.99	8.98	8.97	8.95	8.93
Unfocused inversion		8.79	8.82	8.85	8.86	8.89	8.90	8.91	8.92	8.93	8.94

reference distribution in the Kullback–Leibler divergence is the full-trace inversion, which is the method that provides the most accurate probabilities based on the available data. Therefore, the objective for focused inversion in terms of prediction quality is to keep the Kullback–Leibler divergence as small as possible and ensure at least the same level as that of unfocused inversion.

For unfocused inversion, when the region R_p is reduced from 100% to 10%, the Kullback–Leibler divergence increases from 3.37×10^{-4} to 5.81×10^{-4} which corresponds to a 72% increase and the average prediction power decreases from 0.894 to 0.879. This is what we expect when the amount of data to compare with gets smaller. For focused inversion, we generally observe the same trend but to a smaller extent. Reducing the rank, γ from 100% to 80% does not degrade significantly the Kullback–Leibler divergence and hardly changes the average prediction power. Using smaller regions R_p correspond to a 20%–34% increase in the Kullback–Leibler divergence for ranks 100% to 80%. This is exactly what we aim at when focusing on the most informative features only.

These numbers indicate that we can use focused inversion with a smaller region and a reduced rank γ down to 80%, and still achieve the same level of prediction accuracy as with the unfocused inversion, with a benefit of a processing speed up. Note that this is also valid for traces with very thin sand layers.

The posterior probabilities for the sand layer are shown in Fig. 2 for the full-trace inversion (middle top), for unfocused inversion with nominal region R, i.e., R_{100} (middle bottom) and for focused inversion with a focused region R_{40} and rank γ of 90% (bottom). For the latest method, we choose this parametrization because the Kullback–Leibler divergence is equal to the one obtained with unfocused and region R_{100}. Full-trace inversion, which produces predictions that most accurately match the data, clearly provides the best sand predictions with generally well-predicted top and bottom reservoir reflectors and more certain sand predictions in the reservoir. We notice, however, a few occurrences, for instance between realizations 5500 and 6500, and around 7000, where sand is overpredicted and the full-trace inversion is quite certain about this. On the contrary, there are other occurrences, around traces

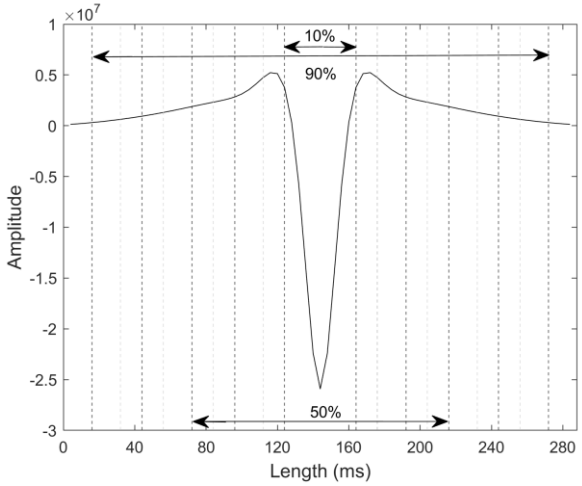

Fig. 5. Near-offset wavelet used for seismic data sampling and inversion. The vertical dot lines show which part of the wavelet is used when the region R is reduced from 100 % (full wavelet) down to 10% by a step of 10%.

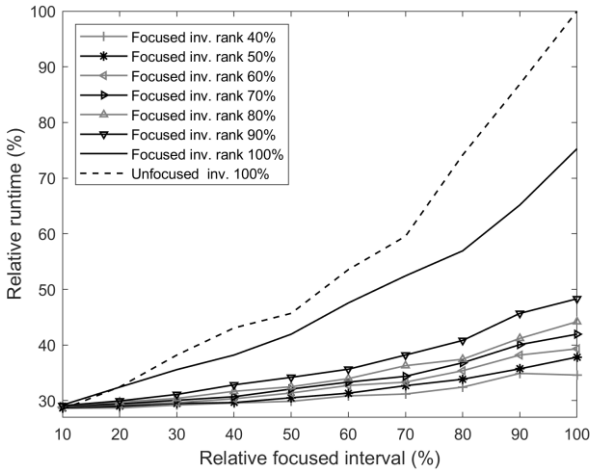

Fig. 6. Relative runtimes as a function of the relative region R_p. Unfocused inversion is the dashed curve, while the other curves are focused inversion for γ between 40% and 100%.

79, 100, or 400 for instance, where the probability for sand is close to 0 in the sand layer. This typically occurs when the seismic data are heavily corrupted by noise to the point where reflections at layers interface are no longer discernible and can be interpreted as noise, even when using a synthetic data example. For unfocused and focused inversions, sand predictions are comparable. In general, there have less certain probabilities than for full-trace inversion, especially for the first 2000 traces. It also appears that sand is overpredicted when the sand layer is very thin or nonexistent. Their average prediction powers are equal to 0.89 which is smaller than that of full-trace inversion at 0.92, as expected.

Comparison of Kullback–Leibler and prediction power scores with unfocused inversion at the nominal region R_{100}, shows that we can obtain the same prediction quality by choosing a more restrictive parametrization. Thus, in this example, we can confidently use focused inversion with a focused region down to 40% and a rank down to 90% as an alternative to unfocused inversion. In terms of wavelet, this is equivalent to cutting it down to the limit of the sidelobe, as shown in Fig. 5.

Fig. 6 shows the inversion runtimes relative to unfocused inversion runtime at the nominal region R_{100}. In this example, going from an unfocused inversion with the nominal region R_{100} to focused inversion with a focused region set to 40% and relative rank to 90% speeds up the inversion by a factor of 3 without deteriorating the quality of the posterior probabilities. They both obtain similar Kullback–Leibler divergence and average prediction powers numbers, therefore making the choice of focused inversion particularly interesting. In addition, we note the strong impact of rank reduction on inversion runtime in this example to keep the most informative features only. When γ incrementally increases from 40% to 100%, the number of retained eigenvalues in the feature space corresponds to 11%, 13%, 16%, 19%, 23%, 28%, and 89%, respectively. More specifically, it illustrates the case, where $L_{\tilde{\mathbf{d}}} < L_R$ even though $\gamma = 100\%$.

We illustrate our methodology with another example, which is the one used in [24]. In this case, full-trace inversion becomes too computationally intensive and instead, we use McMC with a limit of 50 000 iterations as a state-of-the-art benchmark method. The synthetic dataset is a seismic cube of 100×100 traces and consists of three layers. The intermediate layer is the reservoir containing both brine–sand and gas–sand, trapped between the overburden (top layer) and basement (bottom layer). Input seismic data are provided for near, mid, and far acquisition angles. For the inversion, we choose a prior stratigraphic model and rock physics prior properties as in [24]. The stratigraphic model allows for these three layers including uncertainty on the location of the horizons, thus containing a total of four facies classes. These are the target properties to be estimated. The rock physics properties for brine–sand and gas–sand are rather close and may be challenging to separate from each other.

In addition to McMC that we use as a state-of-the-art benchmark method, we invert this dataset with focused inversion for a focused region set to 20% and the rank set to 100%, and with the standard unfocused inversion used in [24] for a window length of five samples. We expect McMC predictions to be more accurate than the other two but at the cost of an extreme computational effort which makes it an unrealistic solution for large-scale projects. Thus, the objective here is to compare the posterior predictions of the two other methods against each other. Table III summarizes the Kullback–Leibler divergence when using McMC as a reference method and the average prediction power, both calculated for brine–sand and gas–sand only. Additionally, Table III includes the inversion runtime. The reason why we only include the reservoir facies classes to evaluate the inversions is that the brine–sand and gas–sand classes represent the target properties of primary interest. Thus, we ensure that the assessment results are not influenced by the large amount of overburden and basement, which are comparatively easier to predict.

We first invert a cross section of 100 traces. Focused and unfocused inversions perform equally well in terms of posterior probabilities and significantly better than the prior distribution. Notably, because of the small size of the cross section dataset, unfocused inversion is as fast as focused inversion.

TABLE III
LARGE-SCALE SEISMIC INVERSION EXAMPLE: SAND PREDICTIONS

	p^*	$D(McMC\|\|p^*)$	$APP(p^*)$	Runtime
Cross-section	McMC	0.000	0.92	7.5 h
	Unfocused	0.12	0.76	0.6 s
	Focused ($R_{20}, \gamma = 100\%$)	0.13	0.73	0.6 s
	Prior	0.45	0.46	Not relevant
Full cube	Unfocused	Not available	0.77	13.0 s
	Focused ($R_{20}, \gamma = 100\%$)	Not available	0.80	8.02 s
	Prior	Not available	0.59	Not relevant

The gain in inversion time is compensated by the increased setup time in this case. McMC, however, requires more than 7 h whereas the other two use less than 1 s. For this reason, we are only able to invert the full seismic cube of 100×100 traces for focused and unfocused inversions, and the Kullback–Leibler divergence is no longer used to assess the quality of the posterior distribution. The average prediction power, however, demonstrates a clear uplift in prediction quality when comparing prior probabilities to both focused and unfocused posterior probabilities. Contrary to the cross section, the posterior probabilities for focused inversion are slightly better than for unfocused inversion for the full cube. This suggests that both methods generally perform equally well, with minor variations depending on the inverted dataset. However, the main difference lies in the runtime, as focused inversion significantly reduces the runtime by 38%.

In these seismic data inversion examples, we have compared unfocused and focused inversions against either full-trace inversion or McMC inversion, which can be considered state-of-the-art benchmark methods, as these approaches provide the best possible achievable posterior probabilities. The huge computational demand makes it, however, unrealistic for large-scale datasets, as demonstrated by the later example. This is why the well-established unfocused inversion is used. We have developed this method further and proposed focused inversion as a computationally more efficient alternative. They both produce posteriors close to full-trace or McMC inversions, but we showed that focused inversion can significantly reduce the inversion runtime in this example.

C. Inversion of Carbonate Core Samples CT Images

We now consider a 2-D example, the estimation of vuggy porosity in CT images of a core from a carbonate reservoir. For some parts of the rock, ultrahigh-resolution CT images are available where one pixel corresponds to 7 μm. However, for most of the rocks, we have a resolution corresponding to medical CT which is 0.5 mm per pixel. Here, we will test different parametrizations of focused inversion and compare the quality of the predictions against the high-resolution images. Thus, we want to demonstrate that using focused inversion to predict porosity from standard-resolution images is relevant.

In prospect evaluation, the porosity of the subsurface rock is a key factor in determining the economics. In most subsurface rocks, the pore space is not visible to the human eye. However, in carbonate rocks, parts of the porosity are contained in pores, which are large enough to be seen by the eye, denoted vuggy porosity. Standard porosity tests often avoid regions with large pores, because this can potentially create weaknesses in the samples so that they may break when undergoing the measurement procedure. Thus, the vuggy porosity must be accounted for by other means. The vuggy porosity also needs to be quantified prior to slabbing the rock due to the brittleness of the rock which may create apparent pore space. This is why the use of CT of carbonate cores is an important source of information.

In this example, we use a set of 109 high-resolution 320×234 images to generate a semisynthetic dataset of 109 blurred low-resolution images that will be used as input data for the focused inversion. This gives a dataset, where we have a controlled experiment but have a realistic resolution. Blurring is achieved first by the convolution of high-resolution images with a point spread function and then by adding Gaussian random and correlated noise. The objective here is to predict the porosity from the blurred images $\mathbf{d}$. The target property configuration $\mathbf{r}_w$ is a 3×3 window

$$\mathbf{r}_w = \begin{pmatrix} r_{-1,-1} & r_{-1,0} & r_{-1,1} \\ r_{0,-1} & r_{0,0} & r_{0,1} \\ r_{1,-1} & r_{1,0} & r_{1,1} \end{pmatrix} \tag{19}$$

where $r_{k,k} \in K$ indicates the presence of porosity, i.e., K = {"void of rock," "rock"}. This represents a total of $2^9 = 512$ configurations. The marginal probability for no rock indicative

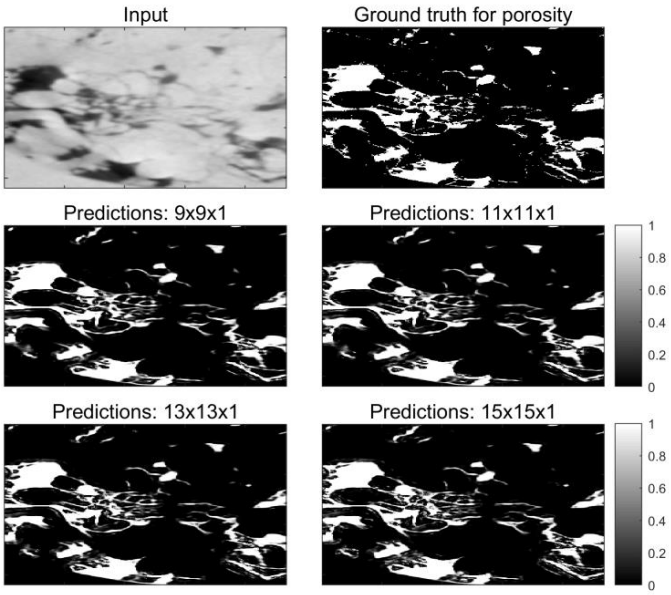

Fig. 7. Example of porosity areas prediction for different sizes of R. Input CT image (top left), ground truth (top right), and posterior probabilities for regions R set to 9×9 (middle left), 11×11 (middle right), 13×13 (bottom left), and 15×15 (bottom right).

of porosity at a given 2-D index is then obtained by integrating the posterior probabilities of all window configurations $\mathbf{r}_w$, where the central element $r_{0,0}$ is the "void of rock" state. In this example, we estimate a stationary prior model of size R for each window from high-resolution images. The larger the region R is the noisier, as it is likely that the area includes rocks even when centered around a window indicative of no rock. Thus, setting up a good prior model is challenging and for this reason, we could reasonably expect better predictions for smaller regions.

We invert the set of low-resolution images to predict the absence of rock indicative of porosity with the focused inversion method and for different (R, γ) parametrizations. We test the region size for R to be 9×9, 11×11, 13×13, and 15×15, and the relative rank γ to increase from 30% to 100%. Again, $\gamma = 100\%$ means that only noninformative features are ignored which can still reduce the dimensionality of the inversion, and therefore, the runtime. A rank γ of 30% means harsher dimension reduction where the most important information is kept but high frequencies or details are ignored. Fig. 7 shows an example of porosity prediction for a CT image (shown in the top left), where spaces void of rocks indicative of porosity correspond to the dark areas. Filtering the high-resolution version of this image provides us with a ground truth estimate of porosity corresponding to the white parts, i.e., with probability 1 (top right). Other subfigures show the marginal posterior probability for porosity for different regions R and for $\gamma = 100\%$, i.e., without loss of any information. We observe a rather good estimation of porosity in all cases. Smaller regions better capture isolated porosity spots like in the top right-hand side corner of the images, while larger regions tend to be more uncertain in thin no porosity areas that are squeezed between porosity areas, like in the bottom left-hand side part of Fig. 7.

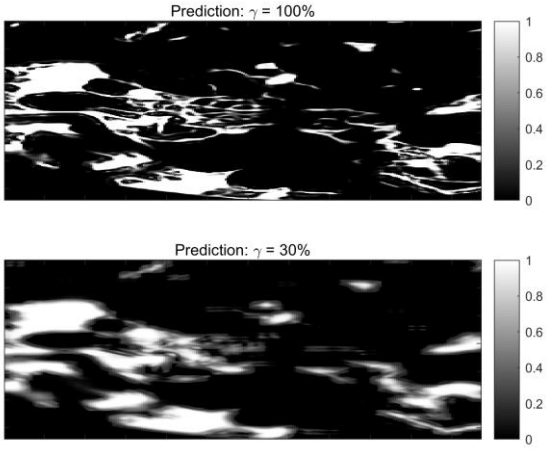

Fig. 8. Illustration of rank reduction on posterior probabilities for porosity. In this example, R is set to 9×9 and the relative rank γ is 100% (top) and 30% (bottom).

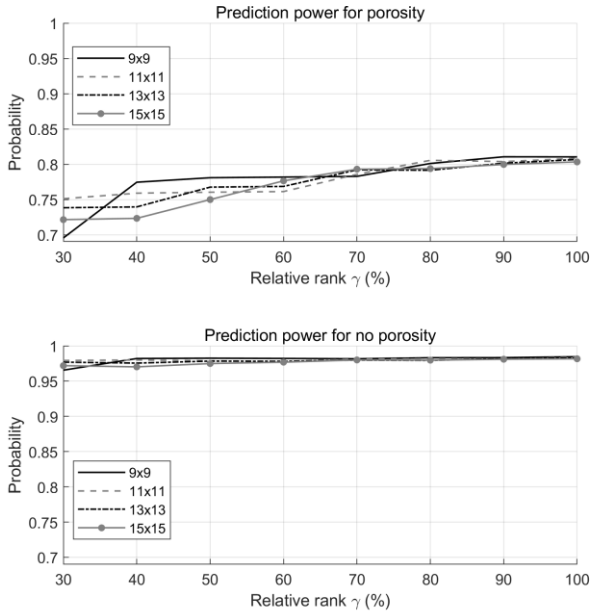

Fig. 9. Prediction power for porosity (top) and no porosity (bottom) as a function of relative rank γ, ranging from 30% to 100%.

The effect of dimension reduction can be observed in Fig. 8 which shows the porosity posterior probabilities for the CT image shown in Fig. 7 and region R set to 9×9. The relative rank γ is set to 100% at the top and 30% at the bottom of Fig. 8, respectively. In this case, the number of retained eigenvalues decreases from 64 when $\gamma = 100\%$ to 4 when $\gamma = 30\%$ out of 81. Keeping the largest eigenvalues ensure the predictions of large porosity areas, but by ignoring smaller ones, we lose finer information necessary to accurately predict transitions between porosity and no porosity, as well as the detection of isolated porosity spots.

We generalize the inversion results of all images and quantify the quality of the predictions with prediction power in Fig. 9 for no rock indicative of porosity $PP_{\text{norock}}(p^*(x))$ (top) and for rock $PP_{\text{rock}}(p^*(x))$ (bottom) as a function of rank γ, ranging from 30% to 100%. As most images mainly consist of rock cells, it is more relevant to use categorical

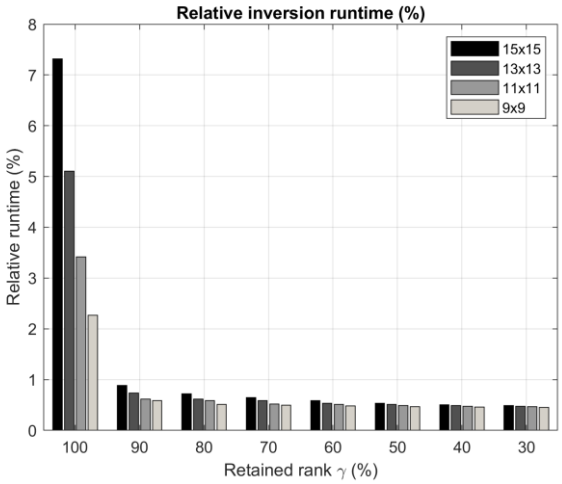

Fig. 10. Relative inversion runtimes for different regions as a function of relative rank γ in (%). The reference runtime, used to calculate relative runtimes, is the one for region 15×15 without any dimension reduction.

prediction power, rather than the average prediction power in this case. The prediction of rock is unproblematic with a prediction power always larger than 0.96. In addition, it appears to be a little sensitive to dimension reduction and regions. The prediction of porosity is relatively good, around 0.8, for all tested regions if the rank reduction is limited such that γ is at least 80%. Below this value, the prediction power progressively declines as the rank reduction gets harsher. The impact of region reduction can really be observed from 50% and below.

Fig. 10 shows the relative runtimes for the different regions R calculated as the inversion runtime normalized by a reference runtime. The reference runtime is chosen to be the runtime for the largest comparison area, 15×15, and without any dimension reduction. As mentioned earlier, this is different from full rank, $\gamma = 100\%$, which ignores noninformative features and thereby reduces the dimensionality of the inversion. Fig. 10 shows that we obtain the most significative runtime speedup when going from no dimension reduction to a relative rank $\gamma = 90\%$. Using small regions R yields to shorter runtimes but we also observe that the speedup tends to equalize as the rank reduction gets harsher. For this example, prediction powers showed that we could use $\gamma = 80\%$ without impacting the prediction's quality, which corresponds to a significative speed up of 99% when compared with our reference runtime. Going even further down in rank reduction would start to impact the quality of the porosity predictions without major runtime speedups.

All in all, it appears that the best predictions quality—runtime speedup tradeoff is obtained for the smallest region R, 9×9, and by reducing the rank such that $\gamma = 80\%$. More generally, this reasonable tradeoff shows the benefits of focused inversion and makes the choice of focused inversion to be a serious alternative for 2-D applications.

In this CT image example, we have assessed the sensitivity of focused inversion to different parameterizations. We have evaluated the impact on porosity predictions both qualitatively, using the prediction power metric, and computationally, by considering the runtime. Applying the methodology to images also allows for visual comparisons. Thus, a substantial rank reduction would affect the level of detail in predictions, while an excessively large region could pose difficulties in estimating a good prior model. Good quality predictions can be achieved with rather small regions and moderate rank reduction. In addition, another benefit of focused inversion is demonstrated through a significant speedup in computational runtime, as observed in this example.

IV. Conclusion

We have presented a sampling-free probabilistic inversion based on the principles of expectation propagation, where we estimate the joint distribution of the target variable in a local outer region. The prior model matches the prior distribution in the local-focused region but integrates model parameters outside the focus region using approximate distributions. The approximate distribution includes large spatial structure information while maintaining the dimension of the inversion small. In addition, we map and solve the inversion into a new feature space where we can exclude components where the data have little influence, thereby decreasing the dimensionality of the inversion, and therefore, the runtime. This method has been tested on two examples. First, the seismic inversion problem of lithologies-fluid prediction using seismic data and geologic information, where predictions have been compared with full-trace exact likelihood inversion and McMC inversion. Second, the estimation of vuggy porosity in CT images of the core from the carbonate reservoir, with a comparison to high-resolution images. We demonstrated through the Kullback–Leibler divergence and prediction power scores that focused inversion can perform equally well and reach the same prediction quality as reference methods while reducing the processing runtime. This makes it to be a realistic candidate method to run large-scale inversion projects.

Appendix

A. Mapping to Features Space

Mapping from inversion results to observations is a key element of the method to capture spatial structure into a compact representation of the data. We compute $\tilde{\mathbf{d}}$, $\tilde{\mathbf{G}}$, and $\tilde{\mathbf{\Sigma}}_e$ from the prior and posterior means $\boldsymbol{\mu}_{\mathbf{m}_R}$ and $\boldsymbol{\mu}_{\mathbf{m}_R|\mathbf{d}}$ and from the prior and posterior covariance matrices $\mathbf{\Sigma}_{\mathbf{m}_R}$ and $\mathbf{\Sigma}_{\mathbf{m}_R|\mathbf{d}}$.

The starting point is the generalized eigenvalue problem that finds a generalized eigenvalue λ and a generalized eigenvector $\boldsymbol{c}$ of $\mathbf{\Sigma}_{\mathbf{m}_R}$ and $\mathbf{\Sigma}_{\mathbf{m}_R|\mathbf{d}}$ that obeys

$$\mathbf{\Sigma}_{\mathbf{m}_R|\mathbf{d}}\mathbf{c} = \lambda \mathbf{\Sigma}_{\mathbf{m}_R}\boldsymbol{c}. \tag{20}$$

The solution to this problem is a set of L_R linearly independent eigenvectors $\boldsymbol{c}_i$ and eigenvalues λ_i such that

$$\mathbf{\Sigma}_{\mathbf{m}_R|\mathbf{d}}\boldsymbol{c}_i = \lambda_i \mathbf{\Sigma}_{\mathbf{m}_R}\boldsymbol{c}_i \quad \text{for} \quad i = 1, \ldots, L_R \tag{21}$$

where the eigenvectors $\boldsymbol{c}_i$ provide a basis which adheres to the following properties:

$$\boldsymbol{c}_i^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{m}_R|\mathbf{d}}\boldsymbol{c}_j = 0, \quad i \neq j \tag{22}$$

$$\boldsymbol{c}_i^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{m}_R}\boldsymbol{c}_j = 0, \quad i \neq j \tag{23}$$

$$\boldsymbol{c}_i^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{m}_R|\mathbf{d}}\boldsymbol{c}_i = \lambda_i \boldsymbol{c}_i^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{m}_R}\boldsymbol{c}_i. \tag{24}$$

Thus, (22) and (23) show that the features made by the generalized eigenvectors are independent in both the prior and the posterior distributions. Furthermore, in (24) we see that the generalized eigenvalue indicates how large a fraction of the variance in the prior which remains in the posterior. We can order the generalized eigenvalues such that

$$0 \leq \lambda_1 \leq \cdots \leq \lambda_{L_R} \leq 1. \tag{25}$$

The latter inequality follows from that the prior dominates the posterior variance. Equation (21) follows the relation:

$$\left(\boldsymbol{\Sigma}_{\mathbf{m}_R} - \boldsymbol{\Sigma}_{\mathbf{m}_R|\mathbf{d}}\right)\boldsymbol{c}_\mathbf{i} = (1-\lambda_i)\boldsymbol{\Sigma}_{\mathbf{m}_R}\boldsymbol{c}_\mathbf{i} \tag{26}$$

where the factor $(1-\lambda_i)$ is interpreted as the reduction in variance when going from the prior to the posterior. Since $\boldsymbol{\Sigma}_{\mathbf{m}_R}$ is invertible, the generalized eigenvalues and eigenvectors can be computed as a standard eigenvalue problem

$$\boldsymbol{\Sigma}_{\mathbf{m}_R}^{-1}\boldsymbol{\Sigma}_{\mathbf{m}_R|\mathbf{d}}\mathbf{c} = \lambda \boldsymbol{c}. \tag{27}$$

Thus, the feature matrix $\tilde{\mathbf{G}}$ we wish to build consists of the first $L_{\tilde{\mathbf{d}}}$ generalized eigenvectors of the equation above where $1 \leq L_{\tilde{\mathbf{d}}} \leq L_R$. Starting with the smallest eigenvector we build the matrix

$$\tilde{\mathbf{G}} = \left[\boldsymbol{c}_\mathbf{1}, \boldsymbol{c}_\mathbf{2}, \ldots, \boldsymbol{c}_{\mathbf{L}_{\tilde{\mathbf{d}}}}\right]^\mathbf{T}. \tag{28}$$

Obviously, the eigenvectors corresponding to $\lambda_i = 1$ should be excluded since this indicates that there is no difference in the uncertainty between the prior and the posterior. Furthermore, excluding eigenvectors corresponding to eigenvalues slightly less than one means excluding components where the data have little influence. Let us now consider the feature

$$f_i = \boldsymbol{c}_i^T \mathbf{m}_{\boldsymbol{R}}. \tag{29}$$

Since both the prior and the posterior matrices are diagonalized by the feature matrix, we see that the transformed problem is diagonal, and therefore, each feature can be treated independently. The equivalent observation error can be computed for each feature independently using the relation

$$\tilde{\mathbf{d}}_i = f_i + \tilde{e}_i \tag{30}$$

where $\tilde{e}_i \sim N(0, \sigma_i^2)$ is the noise component. From (29), we have $f_i \sim N(f_i; \mu_i, \tau_i^2)$, where $\mu_i = \boldsymbol{c}_i^T \boldsymbol{\mu}_R$ and $\tau_i^2 = \boldsymbol{c}_i^\mathbf{T}\boldsymbol{\Sigma}_{\mathbf{m}_R}\boldsymbol{c}_i$.

The posterior expression for this feature is $f_i|\tilde{d}_i \sim N(f_i|\tilde{\mathbf{d}}_i; \mu_{i|d}, \tau_{i|d}^2)$, where the posterior mean and variance are known and equal to $\mu_{i|d} = \boldsymbol{c}_i^T \boldsymbol{\mu}_{R|d}$ and $\tau_{i|d}^2 = \lambda_i \tau_i^2$, respectively.

However, standard posterior computations for Gaussian variables give

$$f_i|\tilde{d}_i \sim N\left(f_i; \mu_i + \frac{\tau_i^2}{\tau_i^2+\sigma_i^2}\left(\tilde{d}_i - \mu_i\right); \tau_i^2\left(1 - \frac{\tau_i^2}{\tau_i^2+\sigma_i^2}\right)\right). \tag{31}$$

Thus, combining the posterior mean and variance with expressions in (31), we can back-compute the observation $\tilde{d}_i$ and the observation error σ_i^2 such that

$$\tilde{\mathbf{d}}_i = \mu_i + \frac{\tau_i^2+\sigma_i^2}{\tau_i^2}\left(\mu_{i|d} - \mu_i\right) \tag{32}$$

$$\sigma_i^2 = \tau_i^2 \cdot \frac{\lambda_i}{1-\lambda_i}. \tag{33}$$

To summarize, the linearized model for observed features is given by

$$\tilde{\mathbf{d}} = \tilde{\mathbf{G}}\mathbf{m}_R + \tilde{e} \tag{34}$$

where $\tilde{\mathbf{G}}$ consists of the first $L_{\tilde{\mathbf{d}}}$ generalized eigenvectors in (28), $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ is the diagonal $L_{\tilde{\mathbf{d}}} \times L_{\tilde{\mathbf{d}}}$ error covariance matrix with variance defined in (33) such that

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_{L_{\tilde{\mathbf{d}}}}^2\right) \tag{35}$$

and $\tilde{\mathbf{d}}$ the observed features defined in (32) and given by

$$\tilde{\mathbf{d}} = \left(\tilde{\mathbf{G}}\boldsymbol{\Sigma}_{\mathbf{m}_R}\tilde{\mathbf{G}}^\mathbf{T} + \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}\right)\left(\tilde{\mathbf{G}}\boldsymbol{\Sigma}_{\mathbf{m}_R}\tilde{\mathbf{G}}^\mathbf{T}\right)^{-1}\tilde{\mathbf{G}}\left(\boldsymbol{\mu}_{R|d} - \boldsymbol{\mu}_R\right) + \tilde{\mathbf{G}}\boldsymbol{\mu}_{\mathbf{m}_R}. \tag{36}$$

Furthermore, we introduce the parameter

$$\gamma = \frac{\sum_{i=1}^{L_{\tilde{\mathbf{d}}}}(1-\lambda_i)}{\sum_{i=1}^{L_R}(1-\lambda_i)} \tag{37}$$

which should be interpreted as the fraction of explained variance included in the model and used in this article as an inversion parameter.

B. Measures for Assessment of the Posterior Model

To quantify the impact of the focused inversion model approximation on the inversion results, we use two statistical measures: the Kullback–Leibler divergence score and average the prediction power.

Let K denote the set of L_K categorical target variables such that $\mathrm{K} = \{r_k\}_{k=1,\ldots,L_K}$. The Kullback–Leibler divergence measures the information lost when the posterior probability p is approximated by the posterior probability p^*, where p is nonzero. Then, the Kullback–Leibler divergence score is calculated by

$$D\left(p(x)||p^*(x)\right) = \sum_{x\in\mathrm{K}} p(x)\ln\frac{p(x)}{p^*(x)}. \tag{38}$$

The second statistical measure is the prediction power of a target variable r_k. Since we use synthetic data, we have access to the true configuration of categorical variables from which the synthetic data have been generated. Thus, if we denote by $\mathcal{L}_k$ a discretized top-down grid along a vertical profile, where r_k exists and by $L_{\mathcal{L}_k}$ the number of grid elements in $\mathcal{L}_k$, then the prediction power of a target variable r_k estimated by posterior probability p^*, $PP_{r_k}(p^*)$, is calculated by summing the posterior probabilities for r_k for grid cells where r_k exists, such as

$$PP_{r_k}\left(p^*\right) = \frac{1}{L_{\mathcal{L}_k}}\sum_{n\in\mathcal{L}_k} p^*(x(n) = r_k). \tag{39}$$

Similarly, we define the average prediction power of the posterior probability p^*, $APP(p^*)$, as the sum of posterior probabilities for categorical variables where these variables exist

$$APP\left(p^*\right) = \frac{1}{\sum_{k=1}^{L_K} L_{\mathcal{L}_k}}\sum_{k=1}^{L_K}\sum_{n\in\mathcal{L}_k} p^*(x(n) = r_k). \tag{40}$$

ACKNOWLEDGMENT

The authors would like to thank Ragnar Hauge and Per Røe at the Norwegian Computing Center, Oslo, Norway, for fruitful feedback on the method development and testing.

REFERENCES

[1] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Phys. Med. Biol.*, vol. 21, no. 5, pp. 733–744, Sep. 1976.

[2] D. Grana and E. D. Rossa, "Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion," *Geophysics*, vol. 75, no. 3, pp. O21–O37, 2010.

[3] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2005.

[4] B. Goodway, T. Chen, and J. Downton, "Improved AVO fluid detection and lithology discrimination using Lamé petrophysical parameters; '$\lambda\rho$', '$\mu\rho$', & '$\lambda\mu$' fluid stack, from P and S inversions," in *Proc. 67th Annu. Int. Meeting, SEG, Expanded Abstr.*, 1997, pp. 183–186.

[5] P. Avseth, T. Mukerji, A. Jørstad, G. Mavko, and T. Veggeland, "Seismic reservoir mapping from 3-D AVO in a North Sea turbidite system," *Geophysics*, vol. 66, no. 4, pp. 1157–1176, 2001.

[6] A. Buland, O. Kolbjørnsen, R. Hauge, Ø. Skjæveland, and K. Duffaut, "Bayesian lithology and fluid prediction from seismic prestack data," *Geophysics*, vol. 73, no. 3, pp. C13–C21, May 2008.

[7] A. L. Larsen, M. Ulvmoen, H. Omre, and A. Buland, "Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model," *Geophysics*, vol. 71, no. 5, pp. R69–R78, Sep. 2006.

[8] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*. New York, NY, USA: Springer, 2006.

[9] D. V. Lindberg and H. Omre, "Blind categorical deconvolution in two-level hidden Markov models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7435–7447, Nov. 2014.

[10] T. Fjeldstad and H. Omre, "Bayesian inversion of convolved hidden Markov models with applications in reservoir prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1957–1968, Mar. 2020.

[11] M. Kemper and J. Gunning, "Joint impedance and facies inversion—Seismic inversion redefined," *First Break*, vol. 32, no. 9, pp. 89–95, Sep. 2014.

[12] K. Rimstad and H. Omre, "Approximate posterior distributions for convolutional two-level hidden Markov models," *Comput. Statist. Data Anal.*, vol. 58, pp. 187–200, Feb. 2013.

[13] M. Ulvmoen and H. Omre, "Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations: Part 1—Methodology," *Geophysics*, vol. 75, no. 2, pp. R21–R35, Mar. 2010.

[14] K. Waters and M. Kemper, "Multi-scenario multi-realisation seismic inversion for probabilistic seismic reservoir characterization," in *Proc. 80th EAGE Conf. Exhib.*, Copenhagen, Denmark, 2018, pp. 1–5.

[15] T. M. Hansen and C. C. Finlay, "Use of machine learning to estimate statistics of the posterior distribution in probabilistic inverse problems—An application to airborne EM data," *J. Geophys. Res., Solid Earth*, vol. 127, no. 11, Nov. 2022, Art. no. e2022JB024703, doi: 10.1029/2022JB024703.

[16] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, no. 2, pp. 183–233, Nov. 1999.

[17] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, nos. 1–2, pp. 1–305, 2007.

[18] M. A. Nawaz, A. Curtis, M. S. Shahraeeni, and C. Gerea, "Variational Bayesian inversion of seismic attributes jointly for geologic facies and petrophysical rock properties," *Geophysics*, vol. 85, no. 4, pp. MR213–MR233, Jul. 2020.

[19] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, Apr. 2017.

[20] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell.*, 2013, pp. 362–369.

[21] O. Kolbjørnsen, A. Buland, and R. Hauge, "Method of modelling a subterranean region of the earth by performing a Bayesian inversion," U.S. Patent 2011 0 172 977 A1, Sep. 3, 2009.

[22] M. Jullum and O. Kolbjørnsen, "A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties," *Geophysics*, vol. 81, no. 3, pp. 75–87, 2016.

[23] O. Kolbjørnsen et al., "Bayesian AVO inversion to rock properties using a local neighborhood in a spatial prior model," *Lead. Edge*, vol. 35, no. 5, pp. 431–436, May 2016.

[24] O. Kolbjørnsen, A. Buland, R. Hauge, P. Røe, A. O. Ndingwan, and E. Aker, "Bayesian seismic inversion for stratigraphic horizon, lithology, and fluid prediction," *Geophysics*, vol. 85, no. 3, pp. 207–221, 2020.

[25] A. Buland and H. Omre, "Bayesian linearized AVO inversion," *Geophysics*, vol. 68, no. 1, pp. 185–198, Jan. 2003.

[26] K. Rimstad and H. Omre, "Bayesian generalized Gaussian inversion of seismic data," Presented at the Petroleum Geostatistics Conf., Sep. 2015, doi: 10.3997/2214-4609.201413591.

[27] A. W. F. Edwards, *Likelihood: An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference*. Cambridge, U.K.: Cambridge Univ. Press, 1972.

[28] A. R. Runnalls, "Kullback–Leibler approach to Gaussian mixture reduction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 989–999, Jul. 2007, doi: 10.1109/TAES.2007.4383588.

[29] I. Kobyzev, S. J. D. Prince, and M. A. Brubaker, "Normalizing flows: An introduction and review of current methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 3964–3979, Nov. 2021.

[30] A. Buland, O. Kolbjørnsen, and H. Omre, "Rapid spatially coupled AVO inversion," in *Proc. SEG Tech. Program Expanded Abstr.*, Jan. 2003, pp. 193–199.

Charlotte Semin-Sanchis received the M.S. degree in electrical engineering from the Grenoble Institute of Technology, Grenoble, France, in 2004, and the Ph.D. degree in electrical engineering from the University of Tromsø, Tromsø, Norway, in 2010. Her Ph.D. thesis was on signal processing techniques for the enhancement of marine seismic data.

From 2010 to 2016, she was a Senior Research and Development Engineer with geoscience companies, including Fugro Norway AS, Oslo, Norway, and CGG, Oslo. She was involved in the development and implementation of new innovative signal processing methods for the analysis, improvement, and reconstruction of seismic data and the development of new seismic data acquisition designs. Since 2016, she has been a Senior Research Scientist with the Norwegian Computer Center, Oslo, where she is involved mostly in Bayesian inversion of seismic amplitude-versus-offset data. She has authored several articles and holds patent applications and granted patents.

Odd Kolbjørnsen received the M.S. degree in industrial mathematics from the Norwegian Institute of Technology, Trondheim, Norway, in 1996, and the Ph.D. degree in mathematical statistics from the Norwegian University of Technology and Science, Trondheim, in 2002. His Ph.D. thesis was on inverse problems formulated within the Bayesian framework.

From 2002 to 2014, he was with the Norwegian Computer Center, Oslo, Norway, where he was involved in geostatistical methods and seismic amplitude-versus-offset inversions. He is currently an Advisor of data science with the Aker BP, Lysaker. He also holds a 20% position as an Associate Professor with the Group for Statistics and Data Science, Department of Mathematics, University of Oslo, Oslo. His research interest includes the applications of advanced analytics in the field of geoscience.

Dr. Kolbjørnsen served as an Associate Editor for *Mathematical Geosciences* from 2010 to 2015.