Edinburgh Research Explorer

Variational Bayesian Inversion of Seismic Attributes Jointly for Geological Facies and Petrophysical Rock Properties

VARIATIONAL BAYESIAN INVERSION OF SEISMIC ATTRIBUTES JOINTLY FOR GEOLOGICAL FACIES AND PETROPHYSICAL ROCK PROPERTIES

Muhammad Atif Nawaz, School of Geosciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom, Muhammad.Atifnawaz@ed.ac.uk

Andrew Curtis, School of Geosciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom, and Exploration and Environmental Geophysics Group, Eth Zurich, Switzerland, Andrew.Curtis@ed.ac.uk .

Mohammad Sadegh Shahraeeni, Total, Pau, France, Mohammad.Shahraeeni@total.com

Constantin Gerea, Geoscience Research Center, Total, Aberdeen, Scotland, United Kingdom, Constantin.Gerea@total.com

ABSTRACT

Seismic attributes (derived quantities) such as P-wave and S-wave impedances and P-wave to S-wave velocity ratios may be used to classify subsurface volume of rock into geological facies (distinct lithology-fluid classes) using pattern recognition methods. Seismic attributes may also be used to estimate subsurface petrophysical rock properties such as porosity, mineral composition and pore-fluid saturations. Both of these estimation processes are conventionally carried out independent of each other and involve significant uncertainties, which may be reduced significantly by a joint estimation process. We present an efficient probabilistic inversion method for joint estimation of geological facies and petrophysical rock properties. Seismic attributes and petrophysical properties are jointly modeled using a *Gaussian mixture* (GM) distribution whose parameters are initialized by unsupervised learning using well-log data. Rock physical models may be used in our method to augment the training data if the existing well data are limited, however this is not required if sufficient well data are available. The inverse problem is solved using the Bayesian paradigm that models uncertainties in the form of probability distributions. Probabilistic inference is performed using *variational optimization* which is a computationally efficient deterministic alternative to the commonly used sampling based stochastic inference methods. With the help of a real data application from the North Sea we show that our method is computationally efficient, honors expected spatial correlations of geological facies, allows reliable detection of convergence, and provides full probabilistic results without stochastic sampling of the posterior distribution.

INTRODUCTION

3D seismic data offers an extensive coverage of the subsurface and provides essential information required to build models of subsurface fluid reservoirs. Such models are used for reserves estimation and for making decisions regarding development of subsurface resources. At the very least, the structural architecture of a reservoir may be defined based on geological interpretation of 3D seismic data. Additional information in the form of spatial distribution of geological facies (discrete lithology-fluid classes) and petrophysical rock properties (continuous physical properties of rocks such as porosity and permeability) is also required for quantitative reservoir characterization. However, such information cannot be inferred from seismic data directly, and must be obtained from other sources of information such as well data. Since well data are usually limited and sparse, we need to map these properties over the entire reservoir. Such a mapping is usually performed by inverting seismic data to ensure that the mapped properties are consistent with the seismic data. For a given geological facies, petrophysical rock properties are often well correlated with seismic attributes; the latter refers to the quantities of interest or elastic rock properties that can be derived from seismic waveform data, such as P-wave and S-wave impedances. Therefore, seismic waveform data and their attributes provide useful constraints on the spatial distribution of both geological facies and petrophysical rock properties.

Examples of seismic attributes are P-wave and S-wave velocities (V_p and V_s) and impedances (I_p and I_s), the ratio of P-wave to S-wave velocity ($\gamma \equiv V_p/V_s$), Poisson's ratio (σ), density (ρ), Lame's coefficients (λ and μ), and amplitude variation with offset (AVO) attributes such as intercept (A), gradient (B) and their product ($A * B$). Examples of petrophysical properties are porosity (φ), volume of clay (V_{cl}) in siliciclastic reservoirs, and pore space water saturations (S_w). Although seismic attributes are generally estimated from the observed seismic waveform data, we refer to them as the *observed data* since these are considered as fixed inputs to our method. The elastic rock properties (or seismic attributes) and the petrophysical rock properties are together referred to as rock properties. Petrophysical rock properties and geological facies are henceforth together referred to as *reservoir properties* or *model parameters* of interest.

Estimating petrophysical rock properties from seismic attributes is a non-unique inverse problem, but it can be regularized in a meaningful way if the solution can be constrained by the distribution of geological facies. Further, discrimination of geological facies from the seismic attributes may be improved if petrophysical rock properties are estimated and as such can be regarded as (uncertain) data along with the seismic attributes. Thus knowledge of either facies or petrophysical properties helps in the discrimination or estimation of the other. Since both of these are unknown, their inference from seismic attributes is a joint, usually nonlinear problem. In this paper, we solve this nonlinear problem in an iterative fashion, by alternately estimating one of these unknowns from the current estimate of the other in each iteration, with the objective of improving the overall joint model.

For the sake of simplicity, Bayesian inversion often assumes that seismic attributes observed or measured at a location depend on the reservoir properties at that location only – the so-called *localized likelihoods* assumption. Such an assumption is commonly used in previous research (e.g., Larsen et al., 2006; Ulvmoen and Omre, 2010; Ulvmoen et al., 2010; Walker and Curtis, 2014; Nawaz and Curtis, 2017). Unfortunately, band-limited nature of seismic data contravenes this assumption and induces strong spatial correlations in seismic images of the subsurface. Another common assumption for the sake of computational efficiency and analytical convenience is that geological facies are spatially independent which ignores spatial correlations in geology. Such an approach has also been implicitly or explicitly used in the literature (e.g., Shahraeeni and Curtis, 2011; Shahraeeni et al., 2012; Grana, 2018) with the hope that the spatial continuity of facies may be recovered from the spatial continuity of seismic data. These assumptions make the inverted reservoir properties vulnerable to noise in the input seismic attributes. Spatial coupling (probabilistic dependence between neighboring locations) based on prior information may be introduced in the reservoir properties (model parameters of interest) to reconstruct desired spatial correlations in their posterior distributions. The difficulty with this approach is that exact Bayesian inference is intractable in real-scale models with spatial coupling between the model parameters. Thus, approximate inference becomes inevitable in this case. Stochastic sampling using *Markov-chain Monte Carlo* (McMC) is widely employed as an approximate inference method for solving spatial inverse problems. However, since McMC is a suite of general methods, it is computationally expensive and

requires significantly large number of samples to converge to the unknown true distribution in high dimensional problems. Approximate inference may be performed more efficiently for many problems of practical interest under a suitable set of assumptions using exact sampling (e.g., Walker and Curtis, 2014), or completely avoiding posterior sampling by using variational principles (e.g., Nawaz and Curtis, 2017, 2018, 2019) or by using stochastic sampling within a variational framework (e.g., Zhang and Curtis, 2020a, 2020b).

Nawaz and Curtis (2017) used spatial inference on a 2D Hidden Markov Model (HMM) in order to recover the marginal posterior distributions of facies from noisy data, including spatial correlations from prior information on expected continuity of facies. Their method uses the localized likelihoods assumption whereby knowing the geology at a location makes the data observed at that location independent of geology elsewhere in the model. Examples of previous research in which the localized likelihoods assumption has been relaxed in 1D Bayesian inversion methods are Lindberg and Omre (2014, 2015), and Grana et al. (2017). Nawaz and Curtis (2018) introduced the concept of *quasi-localized likelihoods* (QLL) – a less stringent assumption than localized likelihoods. In that method, multi-dimensional probabilistic dependence is allowed between seismic attributes at a location and the facies in some arbitrary but pre-specified neighborhood of that location. Nawaz and Curtis (2019) completely removed the assumption of localized likelihoods such that model parameters may be conditioned on any data irrespective of their observation location. They used a discriminative approach that models the posterior distribution directly using supervised learning. This is in contrast to the generative approach that we use in the current paper where we model the posterior distribution through the joint distribution of the model parameters and the observed data. The discriminative approach provides additional sophistication – for example Nawaz and Curtis (2019) used it to discriminate between noise and signal in the data within the inversion process. However, that method requires generation of and learning from training examples which may be regarded as an unnecessary additional step when such sophistication is not required.

This paper extends the method of Nawaz and Curtis (2018) by inverting seismic attributes (elastic rock properties) jointly for petrophysical rock properties and geological facies while honoring prior information on their spatial correlations. We achieve this using *variational Bayesian inversion* (VBI) – an efficient probabilistic inference method based on numerical optimization that allows reliable detection of its convergence. This avoids extensive sampling during inference, yet provides fully probabilistic Bayesian results.

Below we introduce the Bayesian framework for probabilistic inversion that combines the prior information and the data likelihood (the information content in the data regarding unknown model parameters). We then discuss the prior probability model for spatially coupled facies. Then, we describe the practical limitations of exact Bayesian computation in realistic-scale models, and discuss variational Bayesian (VB) inference as an approximate inference method. In this section we describe how Bayesian inference can be performed within an optimization framework without requiring stochastic samples. Then we provide a real data example from the North Sea, where we first show the

inversion results for a gas reservoir on well-log data and then across 2D seismic attributes section. Finally, we provide a discussion on the method, and finally the conclusions.

Before proceeding, we define the notation used in this paper. We use a linear index denoted by lower case letters such as i and j to define the locations (or cells) in our model, or equivalently vertices in the underlying graph. Sets are represented with italic, regular (non-boldface) capital (English or Greek) letters, e.g., $\mathcal{V}$ and $\mathcal{G}$. We use the term vector for a one-dimensional row or column matrix. We use boldface font with lower case (English or Greek) letters for vectors, e.g., $\boldsymbol{r}$ or $\boldsymbol{\beta}$, and upper case English or Greek letters for matrices, e.g., $\boldsymbol{R}$. A subscript used with such letters connotes meanings indicated in the text. The identity matrix is represented as $\boldsymbol{I}$. A superscript T stands for transpose of a vector or matrix. Bracketed superscripts indicate an estimate of a quantity at the iteration number specified in brackets during the course of an iterative update, e.g., $\theta^{(t)}$ represents an estimate of some quantity θ after t iterations of an iterative algorithm. A hat, or caret, over a parameter (or random variable) denotes its estimator, e.g., $\hat{\theta}$ represents an estimator of θ. Other commonly used statistical and set theoretic notations include: '$\sim$' for a random variable which reads "*is distributed as*", '\' for set difference, '$\cup$' for the union of two sets, '$\cap$' for the intersection of two sets, and '$|\cdot|$' for cardinality (or number of elements) of a set.

BAYESIAN INVERSION

We want to infer petrophysical rock properties $\boldsymbol{r}$ and facies $\boldsymbol{\kappa}$ jointly from the seismic attributes $\boldsymbol{d}$ along with their associated uncertainty of prediction. In terms of probability theory, we seek the so called *posterior distribution* $\mathcal{P}(\boldsymbol{r}, \boldsymbol{\kappa} | \boldsymbol{d})$ of unknown model parameters $\boldsymbol{r}$ and $\boldsymbol{\kappa}$ conditioned on the realized data $\boldsymbol{d}$. For this purpose, we use the *generative modeling* approach that formulates an observational model (also called a *forward model*) to describe the relationship between data $\boldsymbol{d}$ and the model parameters $\boldsymbol{r}$ and $\boldsymbol{\kappa}$. The forward model is usually a deterministic or stochastic relationship that can be used to express the likelihood $\mathcal{P}(\boldsymbol{d} | \boldsymbol{r}, \boldsymbol{\kappa})$ of data given the unknown model parameters. For the observed data, this conditional distribution is called the *data likelihood*. The posterior distribution $\mathcal{P}(\boldsymbol{r}, \boldsymbol{\kappa} | \boldsymbol{d})$ and the data likelihood $\mathcal{P}(\boldsymbol{d} | \boldsymbol{r}, \boldsymbol{\kappa})$ are related according to Bayes' theorem as

$$\mathcal{P}(\boldsymbol{r}, \boldsymbol{\kappa} | \boldsymbol{d}) = \frac{\mathcal{P}(\boldsymbol{d} | \boldsymbol{r}, \boldsymbol{\kappa})\mathcal{P}(\boldsymbol{r} | \boldsymbol{\kappa})\mathcal{P}(\boldsymbol{\kappa})}{\mathcal{P}(\boldsymbol{d})} = \frac{\mathcal{P}(\boldsymbol{d}, \boldsymbol{r} | \boldsymbol{\kappa})\mathcal{P}(\boldsymbol{\kappa})}{\mathcal{P}(\boldsymbol{d})}, \tag{1}$$

where $\mathcal{P}(\boldsymbol{\kappa})$ represents the prior distribution of facies, $\mathcal{P}(\boldsymbol{r} | \boldsymbol{\kappa})$ represents the conditional prior distribution of the petrophysical properties $\boldsymbol{r}$ given a particular facies model $\boldsymbol{\kappa}$, and $\mathcal{P}(\boldsymbol{d})$ represents the marginal probability of data $\boldsymbol{d}$. Since the data $\boldsymbol{d}$ are observed as a single realization of noisy underlying random data variables, the denominator $\mathcal{P}(\boldsymbol{d})$ in equation 1 is an unknown constant that ensures normalization of the posterior distribution $\mathcal{P}(\boldsymbol{r}, \boldsymbol{\kappa} | \boldsymbol{d})$ to be a valid probability distribution. It is commonly referred to as the *evidence*, and is given by

$$\mathcal{P}(\boldsymbol{d}) = \sum_{\boldsymbol{\kappa}} \int \mathcal{P}(\boldsymbol{d}|\boldsymbol{r},\boldsymbol{\kappa})\mathcal{P}(\boldsymbol{r}|\boldsymbol{\kappa})\mathcal{P}(\boldsymbol{\kappa})d\boldsymbol{r}\,. \tag{2}$$

Evaluation of the evidence $\mathcal{P}(\boldsymbol{d})$ is intractable for realistic scale models, and must be estimated using approximate probabilistic inference. We postpone further discussion about estimation of $\mathcal{P}(\boldsymbol{d})$ until next section. We first describe a model for the prior distribution $\mathcal{P}(\boldsymbol{\kappa})$ of facies, and then we combine the conditional prior distribution $\mathcal{P}(\boldsymbol{r}|\boldsymbol{\kappa})$ and the likelihood $\mathcal{P}(\boldsymbol{d}|\boldsymbol{r},\boldsymbol{\kappa})$ to form the joint distribution $\mathcal{P}(\boldsymbol{d},\boldsymbol{r}|\boldsymbol{\kappa})$ of rock properties $\boldsymbol{d}$ and $\boldsymbol{r}$ given facies $\boldsymbol{\kappa}$.

Facies prior model

Prior information on the expected patterns of geological facies in space is often available from any previously available data or from modeling of geological processes that might have produced the geological structures in a given depositional environment, and can be expressed effectively using a training image (TI) ([Mariethoz and Caers, 2014](#)). A TI is a pictorial manifestation of spatial continuity of subsurface geological features. Such information may be injected into the Bayesian inversion as prior information parameterized as a *Markov random field* (MRF), a graphical representation of probabilistic dependence among various facies in space.

A MRF is a graphical model $\mathbb{G}(\mathcal{V},\mathcal{E})$, or simply $\mathbb{G}$, containing a set of *vertices* $\mathcal{V}$ (or *nodes*) that represent variables, and a set of edges $\mathcal{E}$ between these vertices that represent probabilistic dependence between the connected vertices (Figure 1). The vertices corresponding to the data (i.e., seismic attributes) are called *observed vertices* and those corresponding to the unknown parameters (reservoir properties) in each model cell are represented as *unobserved vertices*. A MRF that contains hidden vertices is known as a *hidden Markov random field* (HMRF). The set of vertices connected to a given vertex i is called the *neighborhood* of i, denoted by $\mathcal{N}_{\backslash i}$. The subscript $\backslash i$ indicates that $\mathcal{N}_{\backslash i}$ is exclusive of the vertex i itself. The neighborhood is expressed as $\mathcal{N}_i$ when it includes the vertex i.

A MRF assumes that given the model parameters in the neighborhood of a vertex (or model cell), the model parameters at that vertex are conditionally independent of those in the rest of the model – the so-called *(1st-order) Markov assumption*. Accordingly, the geological heterogeneity may be regarded as globally random, while the geology at neighboring locations is more likely to be similar than those at distant locations.
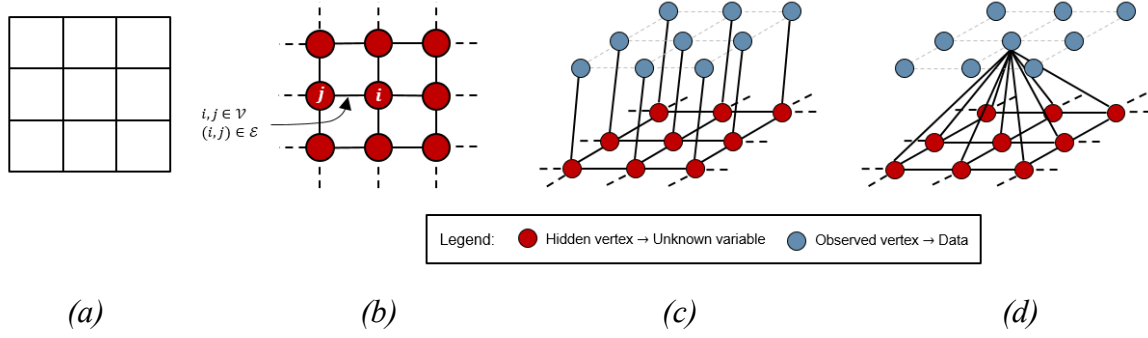
Figure 1: (a) Representation of a standard gridded (cellular) model, and (b, c and d) a probabilistic graphical model (PGM), where vertices in the latter (shown as circles) represent random variables and the edges (links between vertices) indicate probabilistic dependence between the connected vertices (or the associated random variables). Red circles represent hidden vertices or unobserved variables (model parameters) and the blue circles represent observed vertices (data). (c) A typical HMRF with localized likelihoods (LL) where each unobserved variable is conditioned on the observed variable at the same location only. (d) A HMRF with the quasi-localized likelihoods (QLL) assumption of [Nawaz and Curtis (2018)](#), where the hidden variable at each location is conditioned on the observed variables within a pre-specified neighborhood around that location. In this paper we use the QLL assumption which is a relaxation of the LL assumption. The neighborhood of any hidden vertex (red circle) in (b)-(d) consists of the four hidden vertices that share an edge with that vertex.

According to the *Hammerseley-Clifford* theorem (proved by [Besag, 1974](#)), the joint distribution $\mathcal{P}(\boldsymbol{\kappa})$ of facies over a MRF decomposes into potential functions, $\psi_{ij}(\kappa_i, \kappa_j)$ called *edge potentials*, and may be expressed in the form of a *Gibbs distribution* given by

$$\mathcal{P}(\boldsymbol{\kappa}) = \frac{1}{\mathcal{Z}} \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(\kappa_i, \kappa_j), \tag{3}$$

where $\mathcal{Z}$ is a constant that ensures normalization of the joint distribution to be a valid probability distribution and is given by the sum of the numerator over all possible configurations of $\boldsymbol{\kappa}$, i.e.

$$\mathcal{Z} = \sum_{\boldsymbol{\kappa}} \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(\kappa_i, \kappa_j). \tag{4}$$

The potential functions $\psi_{ij}(\kappa_i, \kappa_j)$ may be estimated by scanning the training image and building histograms for various combinations of facies $\boldsymbol{\kappa}$ over pixels with offset distance and direction depending on the graph structure. The prior conditional probability of occurrence of facies κ_i at a location i in the model given the facies $\boldsymbol{\kappa}_{\mathcal{N}_{\backslash i}}$ in its neighborhood $\mathcal{N}_{\backslash i}$ is therefore given by

$$\mathcal{P}\left(\kappa_i|\boldsymbol{\kappa}_{\mathcal{N}_{\setminus i}}\right) \propto \prod_{j \in \mathcal{N}_{\setminus i}} \psi_{ij}\left(\kappa_i, \kappa_j\right), \tag{5}$$

which defines the spatial coupling of facies in terms of the edge potentials $\psi_{ij}\left(\kappa_i, \kappa_j\right)$.

Likelihood model

Two main approaches are used for modeling the relationship between data and model parameters: physics-based modeling and the data-driven modeling. Physics-based models define a mapping from the model parameters to the observed data based on the physics of the problem. Semi-empirical approaches are often used to fit such models as they usually contain free parameters that are tuned such that the derived model matches observed examples of model parameter values and corresponding data. Examples are the parameterized empirical Gardner relationship between density and seismic velocity (Gardner et al., 1974), and the soft-sand and stiff-sand rock physics models (Dvorkin and Nur, 1996) with Gaussian distributed noise. Such models typically require a small number (often three or four) of parameters to be calibrated to fit petro-elastic data (e.g., V_p and φ) from siliciclastic rocks. On the other hand, the data driven approach defines and fits a non-parametric model to the observed samples – a model which cannot be defined in terms of a finite number of parameters. An example of a data-driven model is *non-parametric kernel mixture density* (Grana, 2018) that fits a pre-specified base function (the kernel function) at each data point to approximate any complex probability distribution.

The physics based approach may allow intuitive interpretation of the observed data, for example, fitting the soft-sand and stiff-sand models to petro-elastic data (e.g., V_p and φ) may help determine the compactness of the rocks under investigation. However, for this to be possible the models need to be simple, and consequently they may not capture salient features of any particular dataset. This may lead to inaccurate estimation of posterior (post-inference) uncertainties of the model parameters conditioned to the observed data. The data-driven models incorporate little or no physical intuition about the relationship between model parameters and observed data, however they are flexible in the level of detail that they can capture. Also, in contrast to physics-based models which are often valid only for a particular type of geology, data-driven models may be applied to any geology. However, data-driven models may easily over-fit the data and consequently result in biased posterior estimates of the model parameters.

We use a middle ground: a *Gaussian mixture model* (GMM), which is a semi-parametric way of representing an arbitrarily complex and possibly multimodal distribution. A GMM defines a *Gaussian mixture* (GM) distribution as a linear combination (weighted sum) of Gaussian probability density functions (PDF). It is similar to the Kernel mixture density with Gaussian kernels, but it typically requires a much smaller number of kernels than the number of data points to be fit. For a random variable $\boldsymbol{x}$, a GM distribution with T components may be expressed by the following PDF:

$$\mathcal{P}(\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t \, g_t(\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t \, N(\boldsymbol{x} \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \qquad (6)$$

where $g_t(\boldsymbol{x}) = N(\boldsymbol{x} \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ represents a Gaussian PDF with mean $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$, and α_t is the weight of the tth component of the mixture. A GM distribution is a universal approximator of PDFs: given a sufficient number of Gaussian kernels with appropriate parameters, it can approximate any complex PDF to any desired non-zero accuracy (McLachlan and Peel, 2000).

GM distributions have been widely used to model the distribution of rock properties in geophysical literature (e.g., Meier et al., 2007a, 2007b, 2009; Grana and Della Rossa, 2010; Shahraeeni and Curtis, 2011; Grana et al., 2017; Nawaz and Curtis, 2017, 2018). Shahraeeni and Curtis (2011) used a *Mixture Density Network* (MDN) (Bishop, 1995), which is a type of neural network that can be trained to emulate a desired conditional distribution with a GM distribution. They used it to compute cell-wise posterior distributions of petrophysical rock properties given the seismic attributes in each model cell after the network is trained on well data. In the current work, we use a variant of the *expectation maximization* (EM) algorithm (Dempster et al., 1977; Nawaz and Curtis, 2018) to model the joint distribution of all rock properties (elastic and petrophysical) as a GM distribution. The posterior distribution of petrophysical rock properties given the seismic attributes may then be obtained analytically; by marginalizing or by conditioning on the joint distribution depending on whether the data (seismic attributes or elastic rock properties) uncertainties are included in the model or not, respectively. As opposed to the MDN approach that uses supervised learning from training examples, the presented method is based on unsupervised learning and is computationally more efficient as it avoids the computational cost of generating and learning from training examples.

A rock physics model is usually used to relate elastic properties and corresponding petrophysical properties. However, if sufficient well coverage is available the joint distribution of rock properties may be estimated directly from the well data, i.e. without requiring a rock physics model. This allows estimation of the correlation between any combination of rock properties, and their variances. The conditional prior distribution $\mathcal{P}(\boldsymbol{r}|\boldsymbol{\kappa})$ of petrophysical rock properties $\boldsymbol{r}$ given geological facies $\boldsymbol{\kappa}$ is usually modeled using well logs that have been up-scaled at the dominant seismic wavelength relative to seismic attributes $\boldsymbol{d}$ (Grana and Della Rossa, 2010), and the likelihood $\mathcal{P}(\boldsymbol{d}|\boldsymbol{r}, \boldsymbol{\kappa})$ is usually modeled using rock physics models (Bosch et al., 2010; Grana and Della Rossa, 2010; Lang and Grana, 2018; Grana, 2018) calibrated with the well data and local geological information. We adopt a different approach: we model both of the conditional prior $\mathcal{P}(\boldsymbol{r}|\boldsymbol{\kappa})$ and the likelihood $\mathcal{P}(\boldsymbol{d}|\boldsymbol{r}, \boldsymbol{\kappa})$ jointly using up-scaled well-logs in the form of a joint distribution $\mathcal{P}(\boldsymbol{d}, \boldsymbol{r}|\boldsymbol{\kappa}, \boldsymbol{\Theta})$ of elastic attributes $\boldsymbol{d}$ and petrophysical properties $\boldsymbol{r}$ given the facies $\boldsymbol{\kappa}$, defined in terms of a set of parameters $\boldsymbol{\Theta}$ which we will define and estimate below. Therefore, the current method does not require a rock physics model to be used. However, if well coverage is limited, available well data may be augmented by using an appropriate rock physics model prior to the estimation of the joint PDF of rock properties.

We adopt the quasi-localized likelihoods model of <u>Nawaz and Curtis (2018)</u> where rock properties $\boldsymbol{d}_i$ and $\boldsymbol{r}_i$ in each cell i are conditioned on the facies $\boldsymbol{\kappa}_{\mathcal{N}_i}$ in some pre-specified neighborhood $\mathcal{N}_i$ of i. The quasi localized likelihoods defined in this manner, $\mathcal{P}(\boldsymbol{d}_i, \boldsymbol{r}_i | \boldsymbol{\kappa}_{\mathcal{N}_i}, \Theta)$, may be very high dimensional depending on the size of the neighborhood structure $\mathcal{N}_i$. This may increase the computational cost of the method significantly. However, since facies in the neighboring locations tend to be similar in a MRF model, there is a high probability that any one facies dominates other facies within any neighborhood. This suggests that we can reduce the dimensionality of quasi-localized likelihoods by defining the most probable facies $\hat{\kappa}_i$ in cell i as the one that maximizes the sum of some estimate of marginal probabilities $\hat{\mathcal{P}}(\kappa_j)$ of facies κ_j at locations $j \in \mathcal{N}_i$, i.e.

$$\hat{\kappa}_i = \underset{\kappa}{\operatorname{argmax}} \sum_{j \in \mathcal{N}_i} \hat{\mathcal{P}}(\kappa_j) = \underset{\kappa}{\operatorname{argmax}} \sum_{j \in \mathcal{N}_i} \sum_{\hat{\boldsymbol{\kappa}}_{\mathcal{N}_{\backslash j}}} \mathcal{P}\left(\kappa_j | \hat{\boldsymbol{\kappa}}_{\mathcal{N}_{\backslash j}}\right), \tag{7}$$

where $\mathcal{P}\left(\kappa_j | \hat{\boldsymbol{\kappa}}_{\mathcal{N}_{\backslash j}}\right)$ is the prior probability of facies κ_j at a location j given some estimate $\hat{\boldsymbol{\kappa}}_{\mathcal{N}_{\backslash j}}$ of the facies $\boldsymbol{\kappa}_{\mathcal{N}_{\backslash j}}$ in the neighborhood $\mathcal{N}_{\backslash j}$ of j given by equation 5.

Since the prior distribution $\mathcal{P}(\boldsymbol{\kappa})$ of facies is expressed as a Gibbs distribution, it factorizes over edges in the model according to equation 3. A similar factorization of $\mathcal{P}(\boldsymbol{d}, \boldsymbol{r} | \boldsymbol{\kappa}, \Theta)$ can be achieved by assuming conditional independence of rock properties ($\boldsymbol{d}$ and $\boldsymbol{r}$) given the facies $\boldsymbol{\kappa}$ such that

$$\mathcal{P}(\boldsymbol{d}, \boldsymbol{r} | \boldsymbol{\kappa}, \Theta) = \prod_{i \in \mathcal{V}} \mathcal{P}(\boldsymbol{d}_i, \boldsymbol{r}_i | \boldsymbol{\kappa}_{\mathcal{N}_i}, \Theta) \cong \prod_{i \in \mathcal{V}} \mathcal{P}(\boldsymbol{d}_i, \boldsymbol{r}_i | \hat{\kappa}_i, \Theta). \tag{8}$$

The probability of $\boldsymbol{d}$ given $\boldsymbol{\kappa}$ may then be expressed as

$$\mathcal{P}(\boldsymbol{d} | \boldsymbol{\kappa}, \Theta) = \prod_{i \in \mathcal{V}} \int \mathcal{P}(\boldsymbol{d}_i, \boldsymbol{r}_i | \boldsymbol{\kappa}_{\mathcal{N}_i}, \Theta) d\boldsymbol{r}_i \cong \prod_{i \in \mathcal{V}} \int \mathcal{P}(\boldsymbol{d}_i, \boldsymbol{r}_i | \hat{\kappa}_i, \Theta) \, d\boldsymbol{r}_i$$

$$\equiv \prod_{i \in \mathcal{V}} \varphi_i(\hat{\kappa}_i), \tag{9}$$

where $\varphi_i(\kappa_i) \equiv \int \mathcal{P}(\boldsymbol{d}_i, \boldsymbol{r}_i | \hat{\kappa}_i, \Theta) \, d\boldsymbol{r}_i$ is a potential function of κ_i referred to as the *vertex potential* in a MRF model. It models the likelihood of observing seismic attributes $\boldsymbol{d}_i$ and current estimate of petrophysical properties $\boldsymbol{r}_i$ at a location i which may be regarded as the up-scaled response of facies $\boldsymbol{\kappa}_{\mathcal{N}_i}$ within the neighborhood of i (<u>Nawaz and Curtis, 2018</u>). If the estimate of marginal probability $\hat{\mathcal{P}}(\kappa_j)$ in equation 7 is obtained from the current estimate of posterior marginal distribution of facies in cell i, the approximations 8 and 9 correspond to the notion of *empirical Bayes*.

Petrophysical rock properties are usually obtained from well log data, and are therefore much higher in resolution compared to the seismic attributes. To account for the difference in resolution, the rock properties $\boldsymbol{x}_i = [\boldsymbol{d}_i, \boldsymbol{r}_i]$ at a location i are assumed to be a

weighted linear combination of the corresponding high-resolution rock properties $\boldsymbol{h}_j$ at the neighboring locations $j \in \mathcal{N}_i$ such that

$$x_i = \sum_{j \in \mathcal{N}_i} \beta_j \boldsymbol{h}_j + \boldsymbol{\varepsilon}_i, \tag{10}$$

where $\boldsymbol{x}_i$ is a $p \times 1$ vector of p dimensional vector of rock properties (seismic attributes $\boldsymbol{d}_i$ and the petrophysical properties $\boldsymbol{r}_i$), β_j are the regression coefficients, and $\boldsymbol{\varepsilon}_i$ is a vector of errors which are assumed to be jointly distributed according to a Normal distribution $N(0, \boldsymbol{\Sigma}_\varepsilon)$. The regression coefficients β_j in this expression act as coefficients of a spatial averaging filter, and may be estimated within the inversion process (Nawaz and Curtis, 2018), or may be fixed a priori based on vertical averaging of well-logs at the seismic wavelengths.

We use a Gaussian mixture (GM) distribution to model $\mathcal{P}(\boldsymbol{d}_i, \boldsymbol{r}_i | \hat{\kappa}_i, \Theta)$ that is defined as a linear combination of a given number of Gaussian kernels, usually referred to as the components of the mixture distribution. Defining $\boldsymbol{x}_i \equiv [\boldsymbol{d}_i, \boldsymbol{r}_i]^T$, i.e. a vector of rock properties in cell i, the GM distribution is expressed as

$$\mathcal{P}(\boldsymbol{x}_i | \hat{\kappa}_i = k, \Theta) = \sum_{t=1}^{T_k} \alpha_{t,k} \, g_{t,k}(\boldsymbol{x}_i), \qquad \forall i \in \mathcal{V}, \tag{11}$$

where T_k is the number of mixture components (which may be different for each facies k), $\alpha_{t,k}$ is the component weight and is included in Θ, and $g_{t,k}(\boldsymbol{x}_i)$ is the Gaussian kernel for the tth component and facies $\hat{\kappa}_i = k$. The Gaussian kernels $g_{t,k}(\boldsymbol{x}_i)$ are given by

$$g_{t,k}(\boldsymbol{x}_i) = g_{t,k}\left(\begin{bmatrix} \boldsymbol{d}_i \\ \boldsymbol{r}_i \end{bmatrix}\right) = N\left(\begin{bmatrix} \boldsymbol{\mu}_d \\ \boldsymbol{\mu}_r \end{bmatrix}_{t,k}, \begin{bmatrix} \boldsymbol{\Sigma}_{d,d} & \boldsymbol{\Sigma}_{d,r} \\ \boldsymbol{\Sigma}_{r,d} & \boldsymbol{\Sigma}_{r,r} \end{bmatrix}_{t,k}\right), \qquad \forall i \in \mathcal{V}, \tag{12}$$

where N represents the probability density function (PDF) of the Normal distribution, $\boldsymbol{\mu}$'s and $\boldsymbol{\Sigma}$'s are means and block covariance matrices of the kernel (and are also included in Θ) with subscripts indicating the data $\boldsymbol{d}$ or the petrophysical properties $\boldsymbol{r}$ components of $\boldsymbol{x}_i$. The expression for a Gaussian kernel may also be expressed explicitly as

$$g_{t,k}(\boldsymbol{x}_i) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}_{t,k}|^{-1/2} \exp\left\{\frac{-1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_{t,k})^T \boldsymbol{\Sigma}_{t,k}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_{t,k})\right\}, \\ \forall i \in \mathcal{V} \tag{13}$$

where p is the dimensionality of $\boldsymbol{x}_i$, and $\boldsymbol{\mu}_{t,k}$ and $\boldsymbol{\Sigma}_{t,k}$ are mean and covariance matrix of the kernel $g_{t,k}(\boldsymbol{x}_i)$ given by

$$\boldsymbol{\mu}_{t,k} = \begin{bmatrix} \boldsymbol{\mu}_d \\ \boldsymbol{\mu}_r \end{bmatrix}_{t,k} \tag{14}$$

and

$$\boldsymbol{\Sigma}_{t,k} = \begin{bmatrix} \boldsymbol{\Sigma}_{d,d} & \boldsymbol{\Sigma}_{d,r} \\ \boldsymbol{\Sigma}_{r,d} & \boldsymbol{\Sigma}_{r,r} \end{bmatrix}_{t,k}. \tag{15}$$

Since the joint conditional distribution $\mathcal{P}(\boldsymbol{d}, \boldsymbol{r}|\boldsymbol{\kappa}, \Theta)$ of seismic attributes $\boldsymbol{d}$ and rock properties $\boldsymbol{r}$ given facies $\boldsymbol{\kappa}$ (and the distribution parameters Θ) is modeled as a GM distribution, and the prior distribution of facies $\mathcal{P}(\boldsymbol{\kappa})$ is modeled as a MRF, the overall model of joint distribution $\mathcal{P}(\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{\kappa}|\Theta)$ of the data $\boldsymbol{d}$ and model parameters $\boldsymbol{r}$ and $\boldsymbol{\kappa}$ represents a *Gaussian mixture - Markov random field* (GM-MRF). The parameters Θ may be defined as

$$\Theta \equiv \{\alpha_{t,k}, \boldsymbol{\mu}_{t,k}, \boldsymbol{\Sigma}_{t,k}\}, \forall t, k. \tag{16}$$

We may initialize Θ using some training data (e.g., up-scaled well logs) and, as we show later, Θ may be updated as a part of the inversion process.

Posterior model

The posterior distribution in equation 1 may be written as

$$\mathcal{P}(\boldsymbol{r}, \boldsymbol{\kappa}|\boldsymbol{d}, \Theta) = \frac{\mathcal{P}(\boldsymbol{d}, \boldsymbol{r}|\boldsymbol{\kappa}, \Theta)\mathcal{P}(\boldsymbol{\kappa})}{\mathcal{P}(\boldsymbol{d}|\Theta)}. \tag{17}$$

Substituting equations 3 and 8 into equation 17 we get

$$\mathcal{P}(\boldsymbol{r}, \boldsymbol{\kappa}|\boldsymbol{d}, \Theta) = \frac{\mathcal{P}(\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{\kappa}|\Theta)}{\mathcal{P}(\boldsymbol{d}|\Theta)} \cong \frac{1}{\mathcal{Z}'} \prod_{i \in \mathcal{V}} \mathcal{P}(\boldsymbol{d}_i, \boldsymbol{r}_i|\hat{\kappa}_i, \Theta) \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(\kappa_i, \kappa_j), \tag{18}$$

where $\mathcal{P}(\boldsymbol{d}|\Theta)$ has been absorbed in the normalization constant $\mathcal{Z}'$ on the right hand side. This demonstrates that although we only assumed that the prior distribution $\mathcal{P}(\boldsymbol{\kappa})$ on facies $\boldsymbol{\kappa}$ is a MRF, the posterior distribution $\mathcal{P}(\boldsymbol{r}, \boldsymbol{\kappa}|\boldsymbol{d}, \Theta)$ and the joint distribution $\mathcal{P}(\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{\kappa}|\Theta)$ then also turn out to be MRFs. This is a consequence of the spatial conditional independence assumption on rock properties $\boldsymbol{d}$ and $\boldsymbol{r}$, and we show in the next section that such a factorization of the posterior distribution is crucial for making inference tractable for real-scale models.

VARIATIONAL BAYESIAN (VB) INFERENCE

Evaluating the denominator $\mathcal{P}(\boldsymbol{d}|\Theta)$ in equation 18 requires summation and/or integration over a very high dimensional space for most real scale models. For this reason, approximate inference using stochastic sampling (e.g., by using McMC) is performed (e.g., Grana and Della Rossa, 2010; Rimstad and Omre, 2013; Lindberg and Omre, 2014, 2015) but, as discussed earlier, it is computationally expensive. We instead use the *variational Bayes* (VB) method which uses the 'calculus of variation' to obtain a tractable functional approximation $\mathcal{Q}(\boldsymbol{r}, \boldsymbol{\kappa}|\boldsymbol{d})$, or simply $\mathcal{Q}$, of the intractable true posterior distribution

$\mathcal{P}(r, \kappa | d, \Theta)$. The approximate posterior distribution Q is chosen to belong to a family $\mathbb{Q}$ of distributions that are more easily manipulated, and is commonly referred to as the *auxiliary* or *variational distribution*. The approximation is achieved by minimizing the *Kullback-Leibler (KL) divergence* $KL(Q||\mathcal{P}(r, \kappa | d, \Theta)) \geq 0$ (also called *relative-entropy*, Shannon, 1948) between Q and $\mathcal{P}(r, \kappa | d, \Theta)$, which quantifies how different are its two argument distributions, and is given by

$$KL(Q \mid\mid \mathcal{P}(r, \kappa | d, \Theta)) = \mathbb{E}_Q \left[\log \frac{Q(r, \kappa | d)}{\mathcal{P}(r, \kappa | d, \Theta)} \right]$$

$$= \sum_\kappa \int Q(r, \kappa | d) \log \frac{Q(r, \kappa | d)}{\mathcal{P}(r, \kappa | d, \Theta)} \, dr. \tag{19}$$

VBI transforms probabilistic inference into numerical optimization which can be performed efficiently without requiring stochastic sampling while still providing full probabilistic results. For implementation details, see Nawaz and Curtis (2018). They used the *Expectation Maximization* (EM) algorithm (Dempster et al., 1977) as the optimization framework to solve the Bayesian inverse problem.

Expectation Maximization (EM) algorithm

Expectation Maximization (EM) is an iterative algorithm where each iteration comprises of two steps: the so-called E-step and the M-step, which alternately minimize $KL(Q||\mathcal{P}(r, \kappa | d, \Theta))$ with respect to Q and Θ, respectively. Nawaz and Curtis (2018) showed that the E-step of the EM algorithm can be solved using a message passing algorithm, called *belief propagation* (BP) (Pearl, 1982, 1988), or its variant, the *loopy belief propagation* (LBP) (Murphy et al., 1999; Yedidia et al., 2001a, 2001b; Koller and Friedman, 2009). The LBP algorithm in the E-step of the EM algorithm performs spatial inference by minimizing $KL(Q||\mathcal{P}(r, \kappa | d, \Theta))$ with respect to Q. This provides an estimate of the posterior distribution $Q(\kappa | d)$ of facies κ given seismic attributes d. The marginal conditional distribution of r_i given d_i and $\hat{\kappa}_i$ at a location i may be obtained by conditioning on $\mathcal{P}(d_i, r_i | \hat{\kappa}_i = k, \Theta)$ (see equation 11) by setting the parameters Θ equal to their current estimate $\Theta^{(l)}$ at any iteration l, which may be represented by a GM distribution as

$$\mathcal{P}(r_i | d_i, \hat{\kappa}_i = k, \Theta^{(l)}) = \sum_{t=1}^{T_k} \alpha_{t,k}^{(l)} \, g_{t,k}(r_i | d_i), \qquad \forall i \in \mathcal{V}, \tag{20}$$

where the bracketed superscript refers to the iteration number, and the Gaussian kernel $g_{t,k}(r_i | d_i)$ for the tth mixture component and facies $\hat{\kappa}_i = k$ is given by

$$g_{t,k}(r_i | d_i) = N\left(\left[\mu_{r|d}^{(l)} \right]_{t,k}, \left[\Sigma_{r|d}^{(l)} \right]_{t,k} \right), \qquad \forall i \in \mathcal{V}, \tag{21}$$

with mean $\boldsymbol{\mu}_{r|d}$ and covariance matrix $\boldsymbol{\Sigma}_{r|d}$ estimated from the current estimate $\Theta^{(l)}$ of the parameters Θ of the joint distribution of $\boldsymbol{d}$ and $\boldsymbol{r}$ (equation 11) by

$$\boldsymbol{\mu}_{r|d}^{(l)} = \boldsymbol{\mu}_r^{(l)} + \boldsymbol{\Sigma}_{r,d}^{(l)} \, \boldsymbol{\Sigma}_{d,d}^{(l)\,-1} \left(\boldsymbol{d} - \boldsymbol{\mu}_d^{(l)} \right) \tag{22}$$

and

$$\boldsymbol{\Sigma}_{r|d}^{(l)} = \boldsymbol{\Sigma}_{r,r}^{(l)} - \boldsymbol{\Sigma}_{r,d}^{(l)} \, \boldsymbol{\Sigma}_{d,d}^{(l)\,-1} \, \boldsymbol{\Sigma}_{d,r}^{(l)}. \tag{23}$$

Since petrophysical properties $\boldsymbol{r}$ are assumed to be conditionally independent given facies $\boldsymbol{\kappa}$, their joint posterior distribution $\mathcal{P}\left(\boldsymbol{r}|\boldsymbol{\kappa},\boldsymbol{d},\Theta^{(l)}\right)$ given $\boldsymbol{d}$ and $\boldsymbol{\kappa}$ over the entire graphical model $\mathbb{G}$ at any iteration l may be expressed as

$$\mathcal{P}\left(\boldsymbol{r}|\boldsymbol{\kappa},\boldsymbol{d},\Theta^{(l)}\right) = \prod_{i \in \mathcal{V}} \mathcal{P}\left(\boldsymbol{r}_i|\boldsymbol{d}_i,\hat{\kappa}_i = k,\Theta^{(l)}\right)$$

$$= \prod_{i \in \mathcal{V}} \sum_{t=1}^{T_k} \alpha_{t,k}^{(l)} \, g_{t,k}(\boldsymbol{r}_i|\boldsymbol{d}_i), \forall i \in \mathcal{V}. \tag{24}$$

The M-step of the EM algorithm at any iteration l computes an updated set of parameters $\Theta^{(l+1)}$ by minimizing $KL\left(Q||\mathcal{P}(\boldsymbol{r},\boldsymbol{\kappa}|\boldsymbol{d},\Theta)\right)$ with respect to Θ while keeping the variational distribution Q fixed. This results in the parameters $\Theta_{t,k} \equiv \left\{ \alpha_{t,k}, \boldsymbol{\mu}_{t,k}, \boldsymbol{\Sigma}_{t,k} \right\}$ of the joint GM distribution of $\boldsymbol{x} \equiv [\boldsymbol{d},\boldsymbol{r}]$ to be updated for all of the facies ($k \in \mathcal{G}$) and mixture components ($t = 1, \ldots, T_k$) as follows:

$$\alpha_{t,k}^{(l+1)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathcal{P}}\left(\kappa_i = k|\boldsymbol{d},\Theta_{t,k}^{(l)}\right) \tag{25}$$

$$\boldsymbol{\mu}_{t,k}^{(l+1)} = \frac{\sum_{i=1}^{n} \hat{\mathcal{P}}\left(\kappa_i = k|\boldsymbol{d},\Theta_{t,k}^{(l)}\right)\boldsymbol{x}_i}{\sum_{i=1}^{n} \hat{\mathcal{P}}\left(\kappa_i = k|\boldsymbol{d},\Theta_{t,k}^{(l)}\right)} \tag{26}$$

$$\boldsymbol{\Sigma}_{t,k}^{(l+1)} = \frac{\sum_{i=1}^{n} \hat{\mathcal{P}}\left(\kappa_i = k|\boldsymbol{d},\Theta_{t,k}^{(l)}\right) \cdot \left(\boldsymbol{x}_i - \boldsymbol{\mu}_{t,k}^{(l+1)}\right)\left(\boldsymbol{x}_i - \boldsymbol{\mu}_{t,k}^{(l+1)}\right)^T}{\sum_{i=1}^{n} \hat{\mathcal{P}}\left(\kappa_i = k|\boldsymbol{d},\Theta_{t,k}^{(l)}\right)}, \tag{27}$$

where $\hat{\mathcal{P}}\left(\hat{\kappa}_i = k|\boldsymbol{d},\Theta_{t,k}^{(l)}\right)$ is the current estimate (at iteration l) of the marginal distribution of facies $\hat{\kappa}_i = k$ at location i estimated in the E-step, and acts as weight for averaging the rock properties $\boldsymbol{x}_i \equiv [\boldsymbol{d}_i,\boldsymbol{r}_i]$ at a location i in order to honor the spatial dependence among facies $\boldsymbol{\kappa}$.

Approximate posterior distribution

On convergence of the EM algorithm, the variational distribution $Q(\kappa|d)$ approximates the true posterior distribution $\mathcal{P}(\kappa|d)$ of facies κ given seismic attributes d, such that the desired joint posterior distribution $\mathcal{P}(r, \kappa|d)$ may be approximated as

$$\mathcal{P}(r, \kappa|d) = \mathcal{P}(r|\kappa, d)\mathcal{P}(\kappa|d) \cong Q(r, \kappa|d) = \hat{\mathcal{P}}(r|\kappa, d, \hat{\theta})Q(\kappa|d), \qquad (28)$$

where $\hat{\theta}$ is the final estimate of parameters θ. Note that in the above expression the variational approximation $\mathcal{P}(\kappa|d) \cong Q(\kappa|d)$ on the form of posterior distribution is used only for the posterior distribution of facies, and no approximation on the form of the posterior distribution $\mathcal{P}(r|\kappa, d)$ of petrophysical properties r is assumed; only the value $\hat{\mathcal{P}}(r|\kappa, d, \hat{\theta})$ of $\mathcal{P}(r|\kappa, d, \theta)$ is approximated by the use of estimated parameters $\hat{\theta}$.
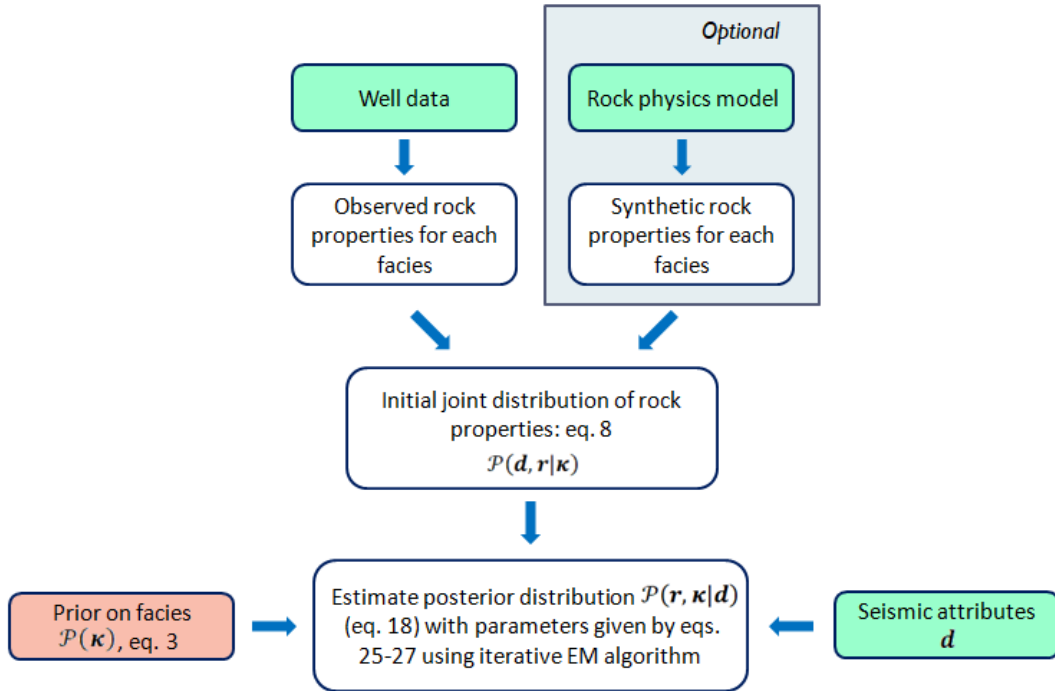

Figure 2: A flow-chart summary of the inversion method. Inputs are shown in green color: well data, rock physics model and seismic attributes. Prior information about facies is shown in red color, and estimated quantities and distributions are shown in white color. The arrows represent direction of flow of data in the workflow. The steps corresponding to rock physics modeling and the corresponding synthetic rock properties for each of the geological facies are enclosed in a blue colored box to emphasize that these are optional and may not be required if sufficient well data is available.

Figure 2 shows a flow chart summary of the overall method. For a discussion on computational efficiency of this variational method, we refer to Nawaz and Curtis (2018) since our current method is an extension of their VBI method to include continuous rock properties, and the computational efficiency of these methods is similar.

FIELD EXAMPLE: NORTH SEA

We apply the joint inversion method to estimate the spatial distribution of petrophysical rock properties and geological facies from well data and seismic attributes from the North Sea. The data available for testing our method includes vertical 2D sections of seismic attributes, P-wave impedance (I_p), S-wave impedance (I_s), and Vp/Vs ratios (V_p/V_s) (Figure 3) obtained from prestack seismic waveform inversion, and well logs from two wells, W1 and W2 (Figures 4 and 5), that are located on the available 2D seismic section. The seismic attributes were available from a previous inversion of seismic waveform data. We are interested in classifying the seismic attribute data into three geological facies: shale, brine-sand and gas-sand, which are jointly estimated together with petrophysical properties of interest: clay volume (V_{cl}), water saturation (S_w) and porosity (φ). The well log data were first analyzed and the three facies of interest (shale, brine-sand and gas-sand) were interpreted from the log data. Crossplots of pairs of elastic properties are shown in Figure 6 with the color scales set to (Figure 6a-6c) the volume of clay and (Figure 6d-6f) the facies interpreted from the well-log data. The gas-sand points are well separated while the brine-sand and shale points show a significant overlay.

The prior spatial distribution of facies was modeled as a MRF using a training image (TI) that represents a conceptual depiction of typical forms of expected geological structures and spatial distributions of facies in the subsurface (Figure 7). The TI encodes the spatial conditional distributions of facies graphically. The prior information was extracted from the training image in terms of prior probabilities $\mathcal{P}\left(\kappa_i | \boldsymbol{\kappa}_{\mathcal{N}\setminus i}\right)$ constructed from histograms of various facies configurations in the image using equation 5. The prior probabilities encapsulate the spatial conditional distributions of facies under the assumption that they are stationary over the entire model space. Since our input seismic attributes span a small 2D vertical section therefore stationarity is an acceptable assumption in this case. If, however, the aim is to invert a large region (or volume) of space or depth/time interval, the priors must be conditioned to the location using zonation or depth trends that capture the expected variability of facies patterns in space.
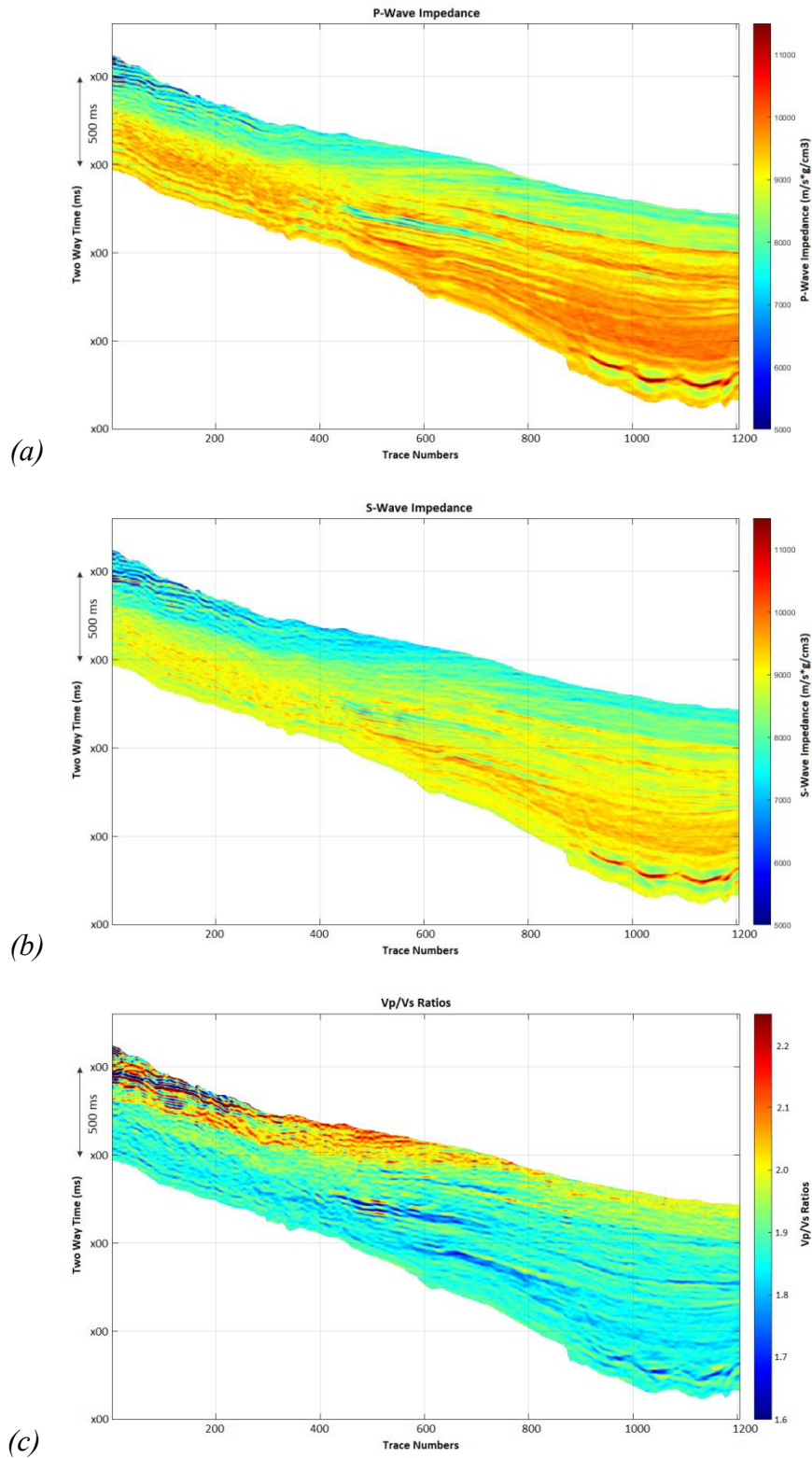
Figure 3: Seismic attributes (a) P-wave impedance, (b) S-wave impedance, and (c) Vp/Vs ratios, derived from a selected 2D section of waveform seismic AVO data using a deterministic inversion method. These attributes are used as inputs to our method for the joint inversion of geological facies and petrophysical rock properties.
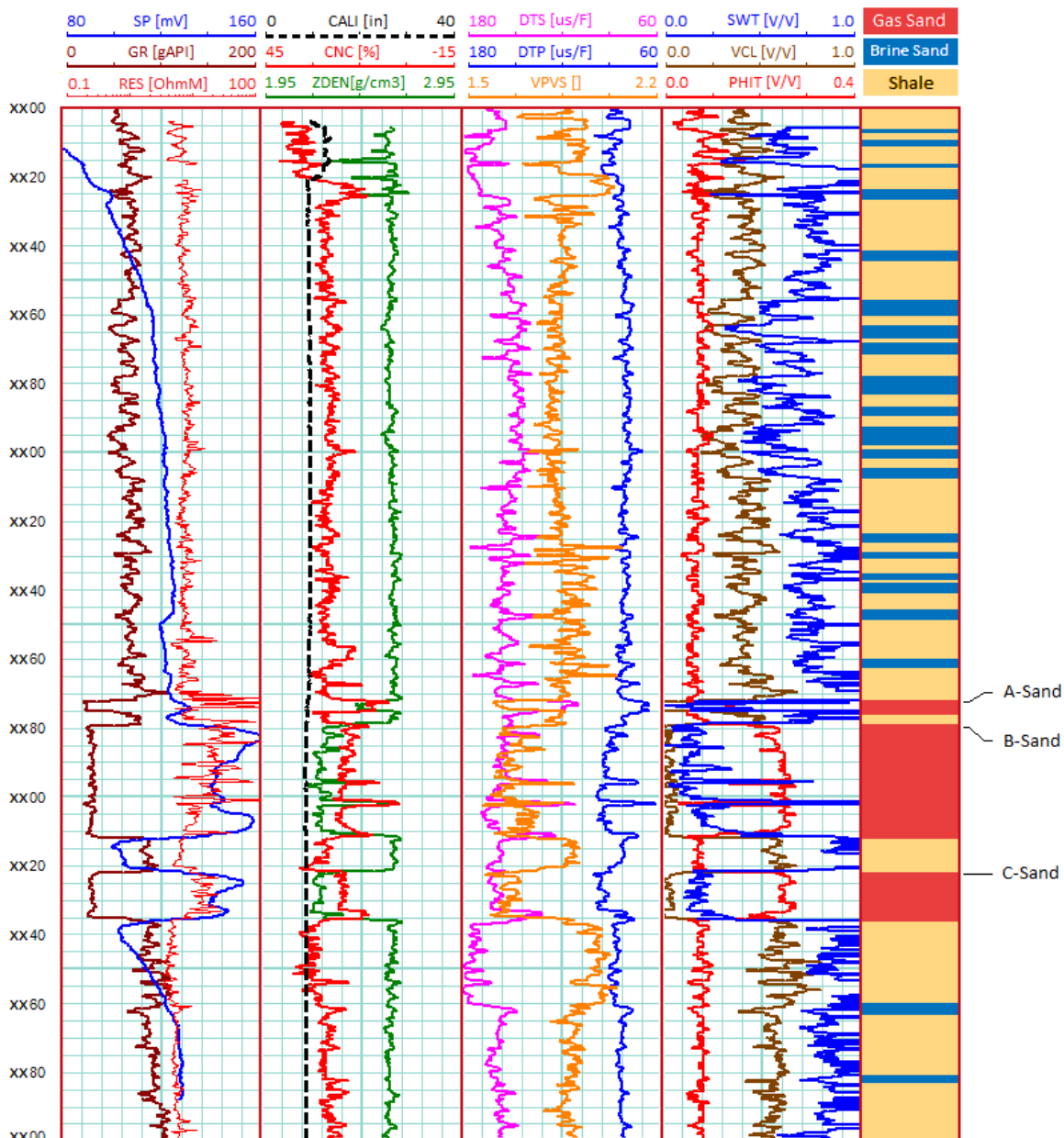
Figure 4: Well-log data and facies profiles in a well W1 in the study area. Standard well-log pneumonics are used for the well log curves as shown in the headers above the display columns. The color codes for three facies, i.e. yellow for shale, blue for brine-sand and red for gas-sand, are used as standard in all of the subsequent figures in this paper. The well log data from W1 are used as input for modeling the facies dependent prior joint distribution of elastic (seismic attributes) and petrophysical rock properties. Three reservoir layers encountered in W1 are marked with labels 'A', 'B' and 'C' for correlation.
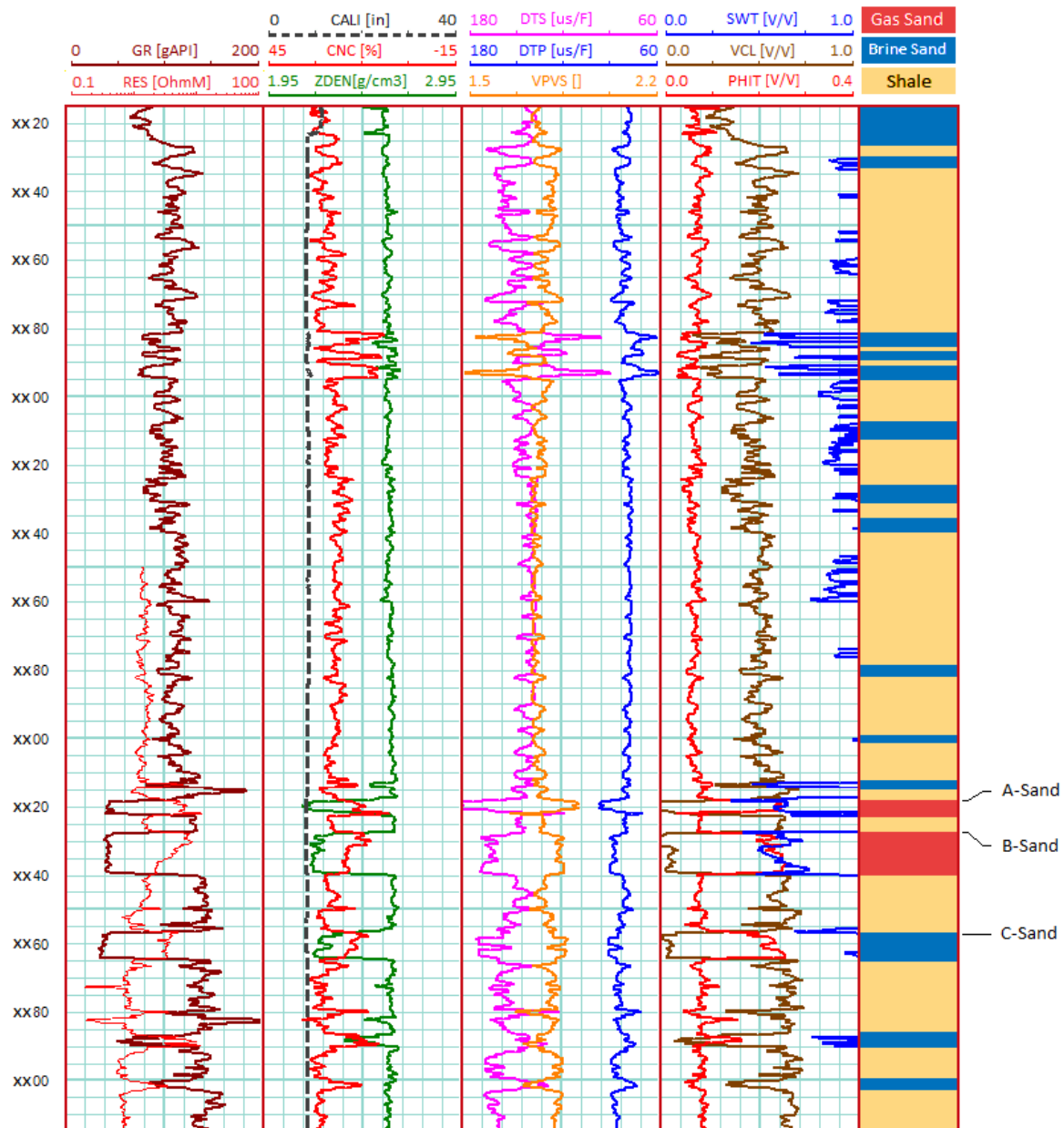
Figure 5: Well-log data and facies profiles in a well W2 in the study area. These data were not incorporated in the inversion process, and were used only for cross-validation (testing) of the results. Standard well-log pneumonics are used for the well log curves as shown in the headers above respective columns. Three reservoir layers encountered in W2 are marked with labels 'A', 'B' and 'C' for correlation.

The initial distribution of facies-dependent rock properties for seismic inversion was built from well log data. The well logs from W1 were used to model the prior distribution of rock properties. W1 encountered only dry gas in the reservoir layers (A, B and C), while W2 encountered brine in the reservoir layer 'C'. For this reason, log data from W2 within the 'C' interval were used for calibration of the prior distribution. Apart from the 'C' interval, W2 data were only used for validation (testing) of the inversion results. To reliably build the probability distribution of rock properties within a subsurface

section (or volume), a significant amount of well data are typically required. However, wells are often sparsely located and the well data are usually limited. In such a case, rock physics modeling and Monte Carlo (MC) simulation must be performed to augment the existing well data in order to build the prior distribution. If we construct a prior distribution using log data only from one well, it would not contain sufficient information to represent the entire model that is to be inverted. Thus, we first build a probabilistic rock physics model of the reservoir formations and then simulate rock properties from it to augment the existing well data.
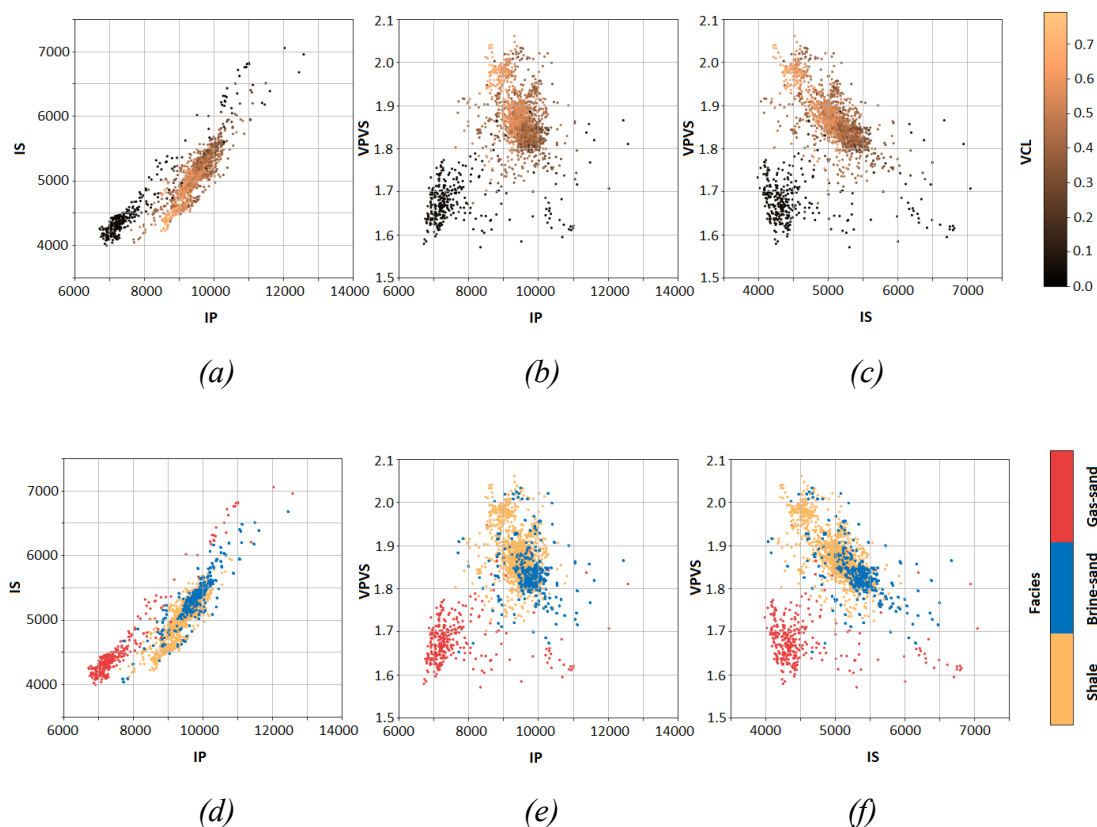

Figure 6: Crossplots between various combinations of P-wave impedance (I_p) and S-wave impedance (I_s) and the P-wave to S-wave velocity ratios (Vp/Vs) observed in the well log data: (a) I_p versus I_s, (b) I_p versus V_p/V_s, (c) I_s versus V_p/V_s, (d) I_p versus I_s, (e) I_p versus V_p/V_s, (f) I_s versus V_p/V_s. The crossplots (a)-(c) are color coded with respect to the volume of clay (V_{cl}) and (d)-(f) are color coded with respect to the interpreted facies. The gas-sand points are well separated from the other facies, while the brine-sand and shale points have a significant overlap.

We performed fluid substitution by synthetically replacing gas with brine in the reservoir sands to simulate the reservoir scenarios that are not actually encountered in W1. This requires a suitable rock physics model to be calibrated with the well data (Bosch et al., 2010). We investigated two related rock physics models: the *soft-sand* and *stiff-sand* models (Dvorkin and Nur, 1996). The soft-sand model assumes that the sand is unconsolidated and the cement is deposited away from the grain contacts, while the stiff-

sand model assumes that the sand is strongly consolidated due to the deposition of cement material at the grain contacts. The parameters of these models are the *coordination number* C_n, the *critical porosity* φ_c, and the *hydrostatic pressure P*. C_n refers to the average number of contacts that each grain has with its surrounding grains, and φ_c refers to the initial porosity at the time of deposition (before the implacement of cement). Figure 8 shows the $\varphi - V_p$ crossplot overlaid on the two models using different values for C_n and φ_c. Higher values of C_n and φ_c show a better fit of the well data with the soft-sand model than with the stiff-sand model. This suggests that the compaction of reservoir sands can be described by the *intermediate stiff-sand* model (Mavko et al., 2009).

The rock physics modeling involves a number of intermediate parameters, such as mineral and fluid properties, that introduce uncertainties in the desired elastic properties of brine-saturated rock. Such intermediate parameters are regarded as confounding variables and are assigned *Uniform* prior distributions listed in Table 1. MC simulation was then performed to sample these confounding variables, followed by upscaling of well logs and fluid substitution using Gassmann's equations (Berryman, 1999) to model brine and gas saturated rock with prior probabilities of brine-sand and gas sand taken from the training image. The simulated data were then combined with the existing well data to obtain augmented data that are expected a priori to represent the elastic properties of rocks in the entire model.
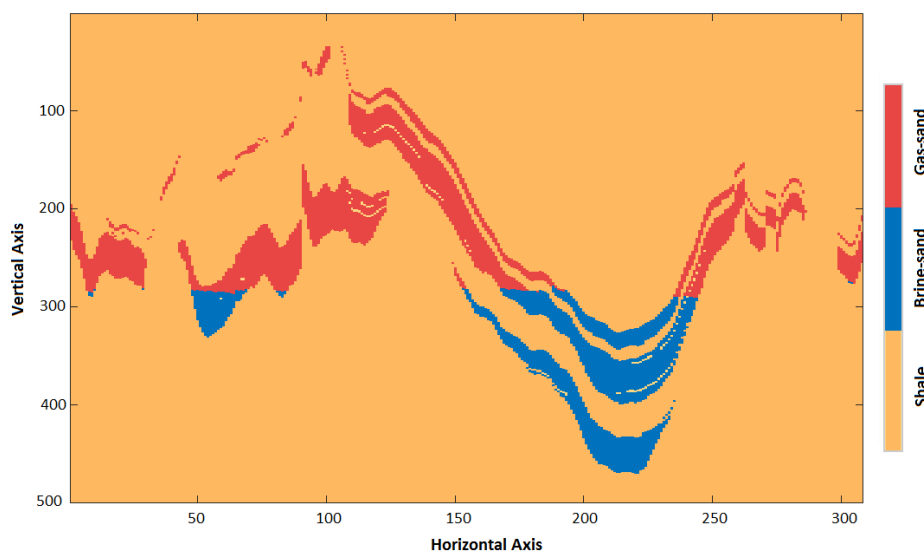

Figure 7: The training image used to model the spatial prior distribution of facies that is constructed from histograms of various facies configurations found in this image.
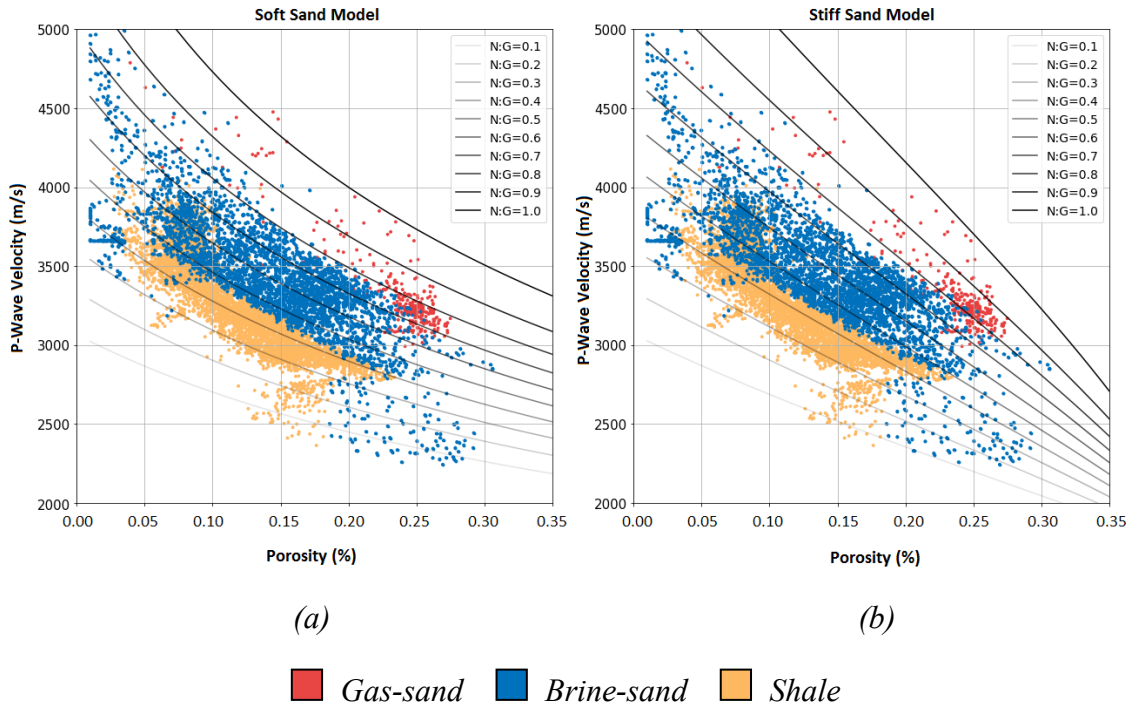
Figure 8: Porosity (φ) vs. P-wave velocity (V$_p$) crossplots with color codes based on the facies interpreted from the well data. The overlaid rock physics template (lines with different shades of grey) correspond to trends for different Net-to-Gross (N:G) ratios predicted using (a) the soft-sand and (b) the stiff-sand model. Each of the two rock physics models are calibrated using different set of parameters: the coordination number $C_n = 13$ *and the* critical porosity $\varphi_c = 0.5$ *for the soft-sand model, and* $C_n = 5$ *and the* critical porosity $\varphi_c = 0.4$ *for the stiff-sand model. This shows that the reservoir can be modeled using the Intermediate stiff-sand model ([Mavko et al., 2009](#)), i.e. either by a stiffer soft-sand model or a softer stiff-sand model.*

Figure 9 shows I_p versus V_p/V_s, and I_p versus I_s crossplots for a comparison between the original well data, the data after fluid substitution (brine replacing gas in the reservoir) using mean values of the confounding parameters, and the augmented data using MC simulations. The prior facies dependent joint distribution of the petrophysical and elastic rock properties (Figure 10) was modeled as a GM distribution using the augmented data. Each of these facies dependent GM distributions was modeled as a mixture of two Gaussian components in order to capture possible multimodal behavior of rock properties within each facies.

Before applying our method to invert elastic seismic attributes for petrophysical properties and facies, we first test the method by inverting the elastic logs from W2 for petrophysical properties and facies. This also validates the consistency of the prior distribution built using rock physics modeling against the log data from W2. Recall that the W2 data were not used in building the prior distribution. The joint inversion for petrophysical rock properties and facies was performed by updating the prior distribution of rock properties by conditioning on the seismic attributes (Figure 11) using the EM

algorithm as discussed earlier. The E-step of the EM algorithm approximates the posterior marginal distributions of facies using equations 20-24, while the M-step updates the parameters of the joint distribution of rock properties given facies estimated in the E-step using equations 25-27. The marginal conditional distribution of petrophysical properties given the observed elastic properties (elastic well logs in this case) may be computed for each facies at any iteration of the EM algorithm by conditioning on the joint distribution of rock properties given facies using equation 20. However, this is typically required only after convergence of the EM algorithm.

Table 1: Prior Uniform distribution ranges used for the intermediate rock physics parameters.

Rock Physics Parameter	Range
Coordination number, C_n	$5 - 13$
Critical porosity, φ_c	$0.4 - 0.5$
Hydrostatic pressure, P	$40 - 55$ MPa
Mineral density, ρ_m	$2.5 - 2.8$ g/cm^3
Mineral bulk modulus, K_m	$15 - 38$ GPa
Mineral shear modulus, μ_m	$5 - 44$ GPa
Brine density, ρ_b	$1.0 - 1.1$ g/cm^3
Brine bulk modulus, K_b	$2.2 - 2.8$ GPa
Gas density, ρ_g	$0.15 - 0.25$ g/cm^3
Gas bulk modulus, K_g	$0.04 - 0.06$ GPa
Error in volume of clay, ΔV_{cl}	$0.0 - 0.2$
Error in water saturation, ΔS_w	$0.0 - 0.1$
Error in porosity, $\Delta \varphi$	$0.0 - 0.1$

Figure 9: (a)-(c) I_p versus V_p/V_s crossplots in the first row, (d)-(f) I_s versus V_p/V_s crossplots in the second row, and (g)-(i) I_p versus I_s crossplots in the third row. The first column (a, d and g) displays the crossplots using log data from W1. The second column (b, e and h) displays the crossplots using the original well data together with the well data after replacing gas with brine in the sand layers using Gassmann fluid substitution modeling to show the effect of brine on the elastic properties of reservoir layers (A, B and C). The third column (c, f and i) displays the crossplots using Monte Carlo (MC) simulated data using the soft-sand model with intermediate rock physics parameters as shown in Table 1 to simulate a wide range of possible values that might not have been sampled in the well data.
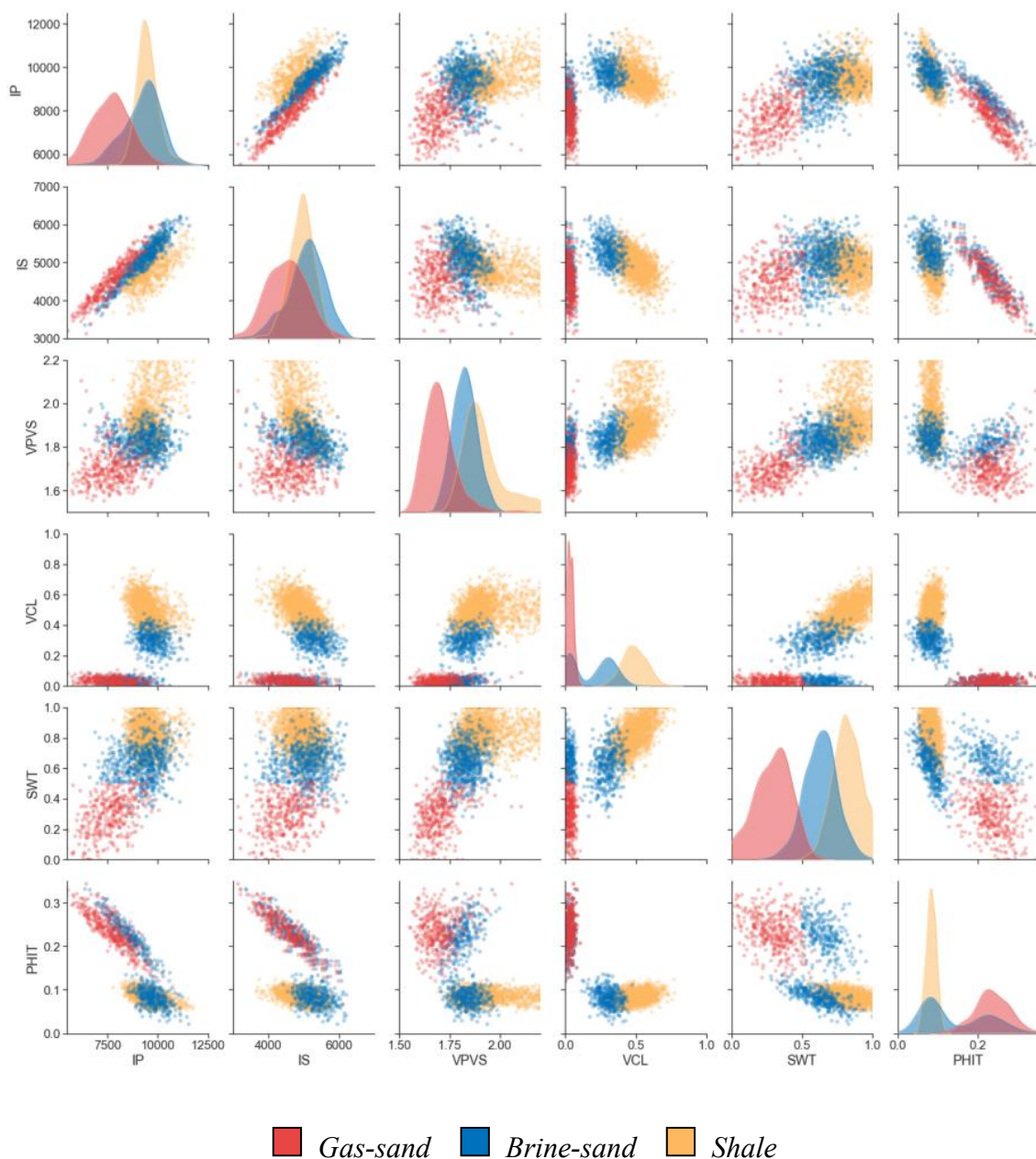
Figure 10: Matrix-plot of samples from components of the prior joint distribution of elastic and petrophysical rock properties. The first three components are the elastic properties: P-wave impedance I_p (IP log), S-wave impedance I_s (IS log) and the P-wave to S-wave velocity ratios V_p/V_s (VPVS log), and the last three components are the petrophysical properties: clay volume V_{cl} (VCL log), water saturation S_w (SWT log) and porosity φ (PHIT log). The diagonal plots represent smoothed histograms of each of the components, and the off-diagonal plots show facies dependent correlations between the respective components.
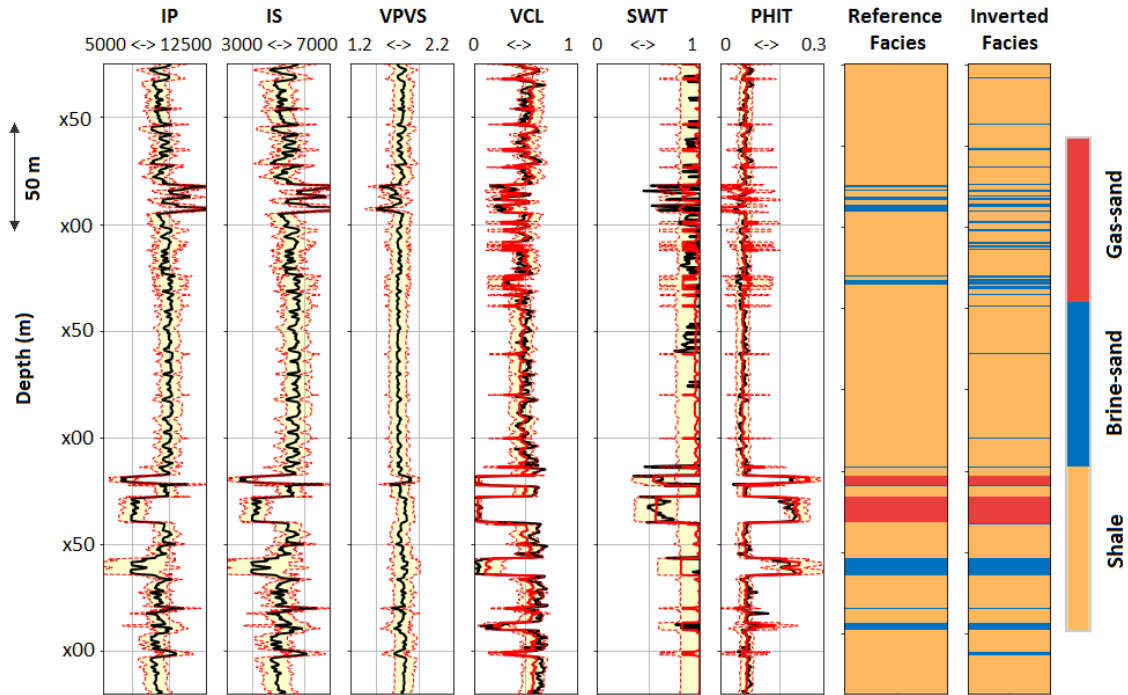
Figure 11: Well logs inversion results. The first three columns display the input elastic rock properties: P-wave impedance I_p (IP log), S-wave impedance I_s (IS log) and the P-wave to S-wave velocity ratios V_p/V_s (VPVS log), shown in the solid-black lines estimated from the sonic (DTP and DTS) and density (ZDEN) logs shown in Figures 4 and 5. The solid-black curves in columns 4-6 are the reference petrophysical well logs, and solid-red curves the mean inverted petrophysical properties: clay volume V_{cl} (VCL log), water saturation S_w (SWT log) and porosity φ (PHIT log). Column-7 displays the reference facies interpreted from the well data and column-8 shows the inverted facies. The yellow shaded regions bounded by the dashed-red curves represent the 2nd standard deviation of the posterior marginal distributions of the petrophysical rock properties in columns 4-6, and the 2nd standard deviation of the conditional marginals of the joint distribution of rock properties obtained by conditioning on the estimated posterior mean petrophysical properties and integrating out the elastic properties other than the one that is plotted in columns 1-3.

Testing the inversion method on the well log data provides a best case scenario for our method since the BP algorithm performs exact inference in the 1D case. Therefore, any inaccuracies in the inversion results in this case are not a result of any approximation used in probabilistic inference, but may be attributed to the approximations used in rock physics modeling. The inversion results are shown in Figure 11. The input to inversion are the measured elastic well logs (P-wave and S-wave impedances and VpVs ratios) that are shown as solid-black curves in the columns 1-3. The outputs are joint posterior distribution of the elastic and petrophysical rock properties and facies. The joint posterior GM distribution was conditioned on the observed elastic well logs using equations 20 and marginalized to obtain the posterior distribution of inverted petrophysical logs (VCL, SWT and PHIT). Each of the marginal posterior GM distributions of petrophysical properties

were approximated with univariate Gaussian distributions for display and interpretation purposes. The solid-red curves in columns 4-6 are means of posterior distribution of petrophysical properties. The yellow shaded regions bounded by the dashed-red curves in columns 1-6 are the 2nd standard deviation of the posterior distribution of corresponding rock properties. The actually observed petrophysical logs are shown as solid-black curves in columns 4-6 for comparison.

The standard deviation (Std.) of rock properties quantifies the natural variability of these properties, and also provides quantification of uncertainty of the predicted petrophysical properties. For precise inversion results, exactly 95.4% of the actual observed log samples should fall within the 2nd standard deviation of the posterior distribution. We define the percentage of actual petrophysical log samples contained within the 2nd standard deviation of the predicted distributions to the ideal value of 95.4% as the *confidence ratio* (CR). An ideal CR is therefore 1.0 which refers to perfect prediction of uncertainty for a Gaussian distribution. A CR value greater than 1.0 represents over-estimation of uncertainty, and vice versa. The CR for well data inversion of the petrophysical properties are shown in Table 2. The uncertainty is slightly under-estimated for the inverted petrophysical properties (with CR ranging between 0.93 and 0.98). It is interesting to note that since our method estimates the posterior conditional distributions of petrophysical properties from the joint distribution of elastic and petrophysical rock properties, it yields uncertainty in the input elastic properties under the joint distribution as well (as shown by the yellow shaded regions in columns 1-3).

Table 2: Accuracy measures for the petrophysical properties and facies inverted at well locations computed with respect to the actually measured (reference) log-curves and facies interpreted from well data. Confidence ratio and success rate are defined in the text.

| Property and Accuracy measure	Well-log inversion (W2)	Seismic inversion (W1)	Seismic inversion (W2)
Volume of clay, V_{cl}: Confidence	0.93	0.82	0.73
Volume of clay, V_{cl}: Correlation	0.91	0.59	0.72
Water saturation, S_w: Confidence	0.96	0.82	0.91
Water saturation, S_w: Correlation	0.81	0.68	0.61
Porosity, φ: Confidence ratio	0.98	0.77	0.89
Porosity, φ: Correlation	0.93	0.60	0.81
Shale prediction: Success rate	0.94	0.83	0.82
Brine-sand prediction: Success rate	0.76	0.60	0.66
Gas-sand prediction: Success rate	0.98	0.80	0.96
Overall facies prediction: Success	0.90	0.74	0.81
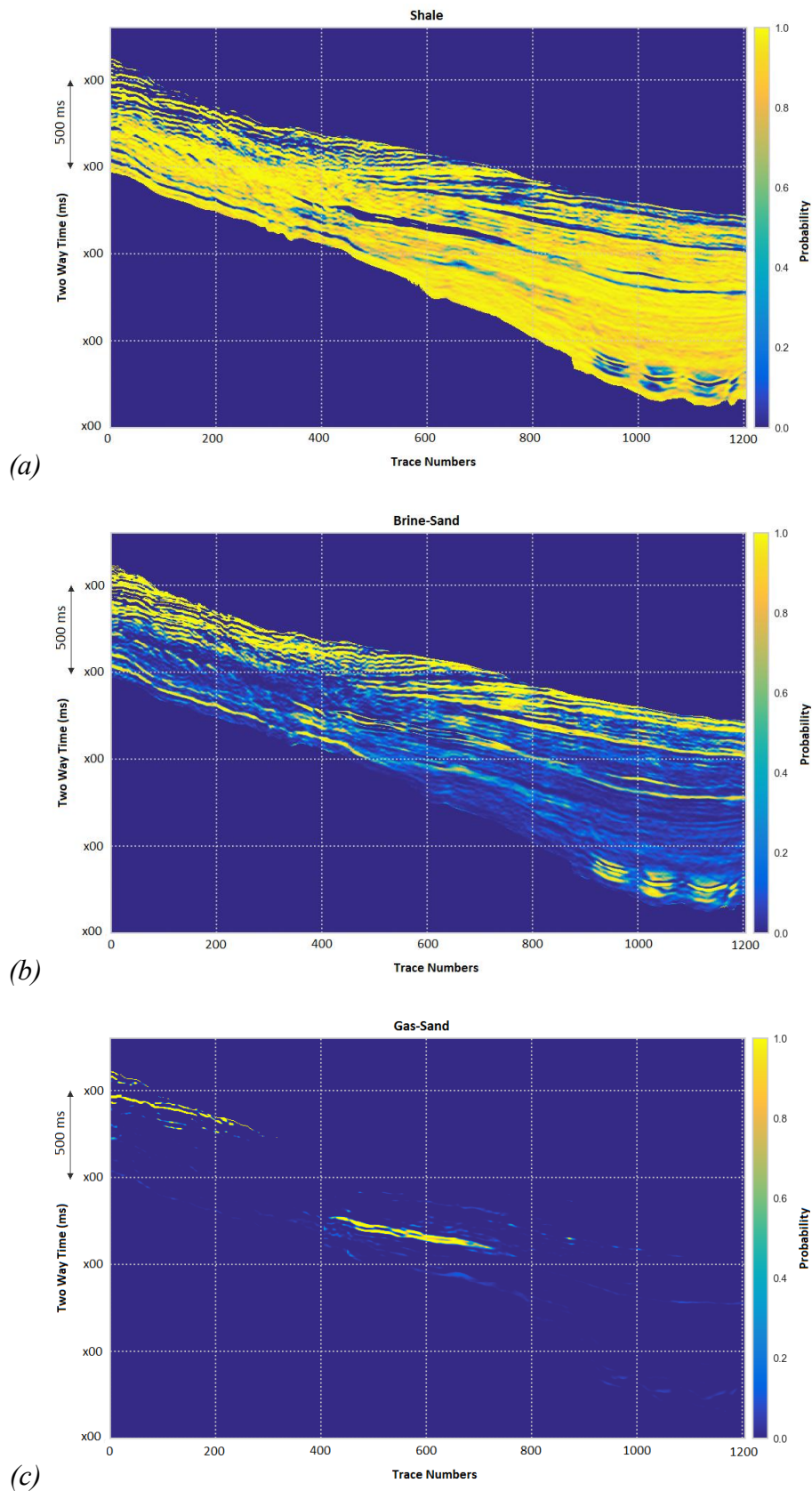
Figure 12: Cell-wise posterior marginal distributions of (a) shale, (b) brine-sand, and (c) gas-sand. Yellow color represents high probability (value=1.0) and dark blue color represents low probability (value=0.0).
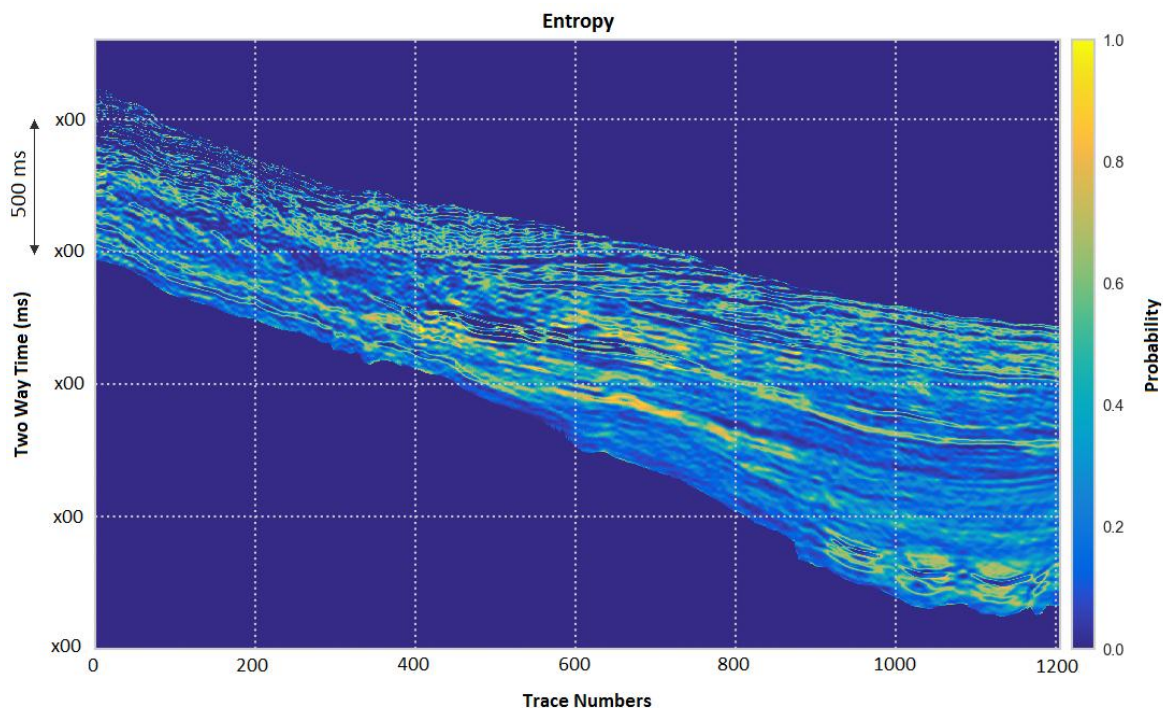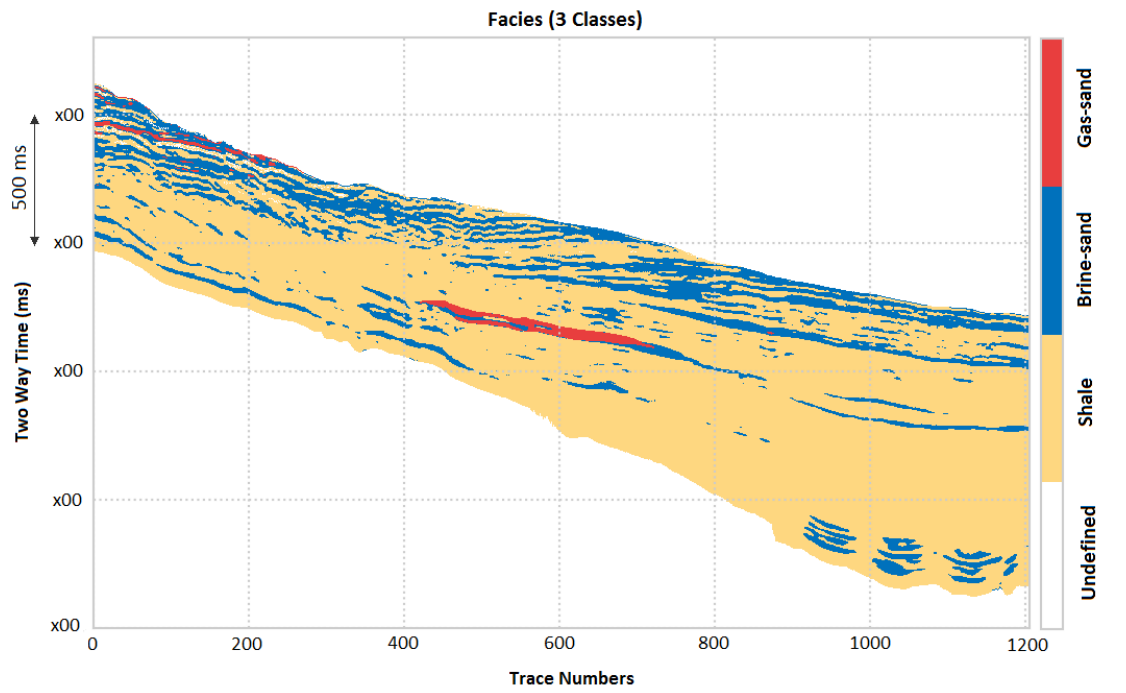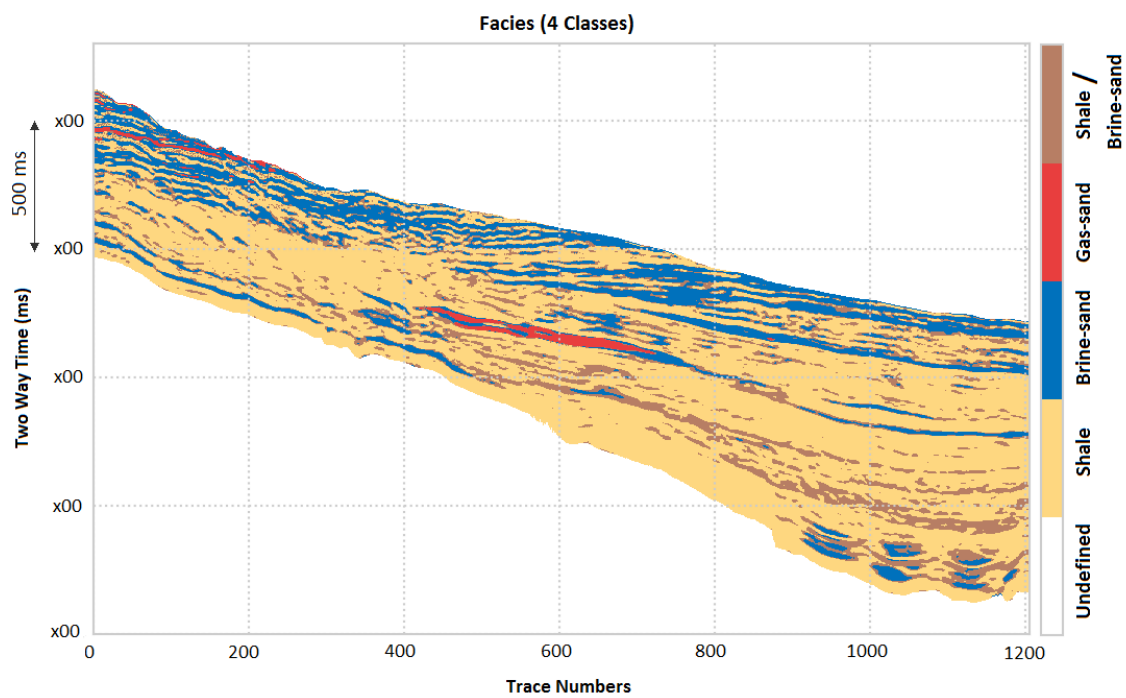
Figure 13: Cell-wise posterior marginal entropy of facies classification shown in Figure 12 scaled between 0.0 and 1.0. Yellow color represents high entropy (value=1.0) and dark blue color represents low entropy (value=0.0).

The similarity between the mean inversion results and the corresponding reference log curves is estimated in terms of Pearson's correlation coefficient, herein referred to simply as correlation. Excellent correlation of 0.91 and 0.93 is obtained for inverted V_{cl} and φ (compared to the measured reference log curves VCL and PHIT, respectively), while a relatively lower correlation of 0.81 is obtained between the inverted S_w and the measured SWT log curve. It shows that the elastic properties have a higher correlation with clay volume and porosity than with water saturation, which is also evident from Figure 10.

The success rate refers to the percentage of facies correctly predicted at the well location. The success rate is very good for shale (94%) and a bit low for brine-sand (76%), whereas the gas-sand has an excellent 98% predicted rate as the gas-sand properties are well discriminated from the rest of the two facies (Figure 9). As mentioned earlier, a 1D inversion with our method provides the best case results since the probabilistic inference is exact in this case, and minor discrepancies between predicted and actual properties are due to the approximations used in rock physics modeling. Since, the two wells are located quite close together (about 2.0 km apart), the reservoir properties are not expected to be too different and the assumption of stationarity appears to be valid.

(a)

(b)

Figure 14: Maps of the facies with maximum marginal distribution in each cell. (a) Map of the three inverted facies: Shale (SH: shown in yellow), brine-sand (BS: blue) and gas-sand (GS: red). (b) Map with an additional facie "Shale/Sand" (SS: brown) identified from high entropy layers in Figure 13.
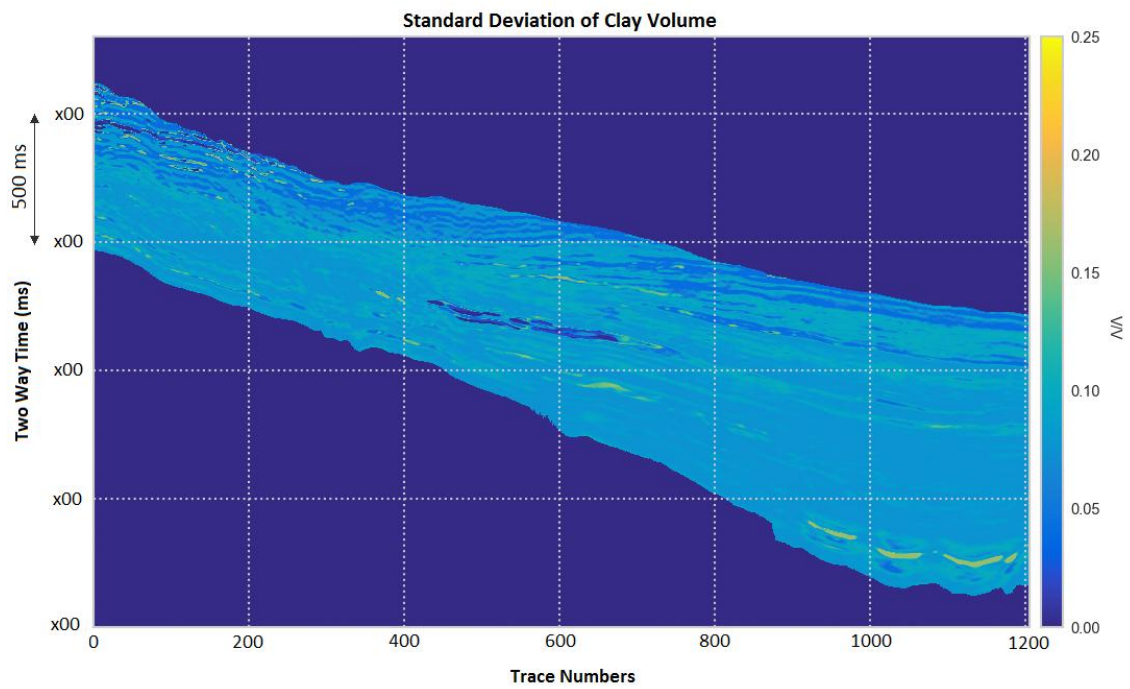
(a)

(b)

Figure 15: Cell-wise map of (a) clay volume (V_{cl}) and (b) its standard deviations (Std.). Yellow color represents high values and dark blue color represents low values of the respective properties.

(a)

(b)

Figure 16: Cell-wise map (a) water saturation (S_w) and (b) its standard deviations (Std.). Yellow color represents high values and dark blue color represents low values of the respective properties.

(a)

(b)

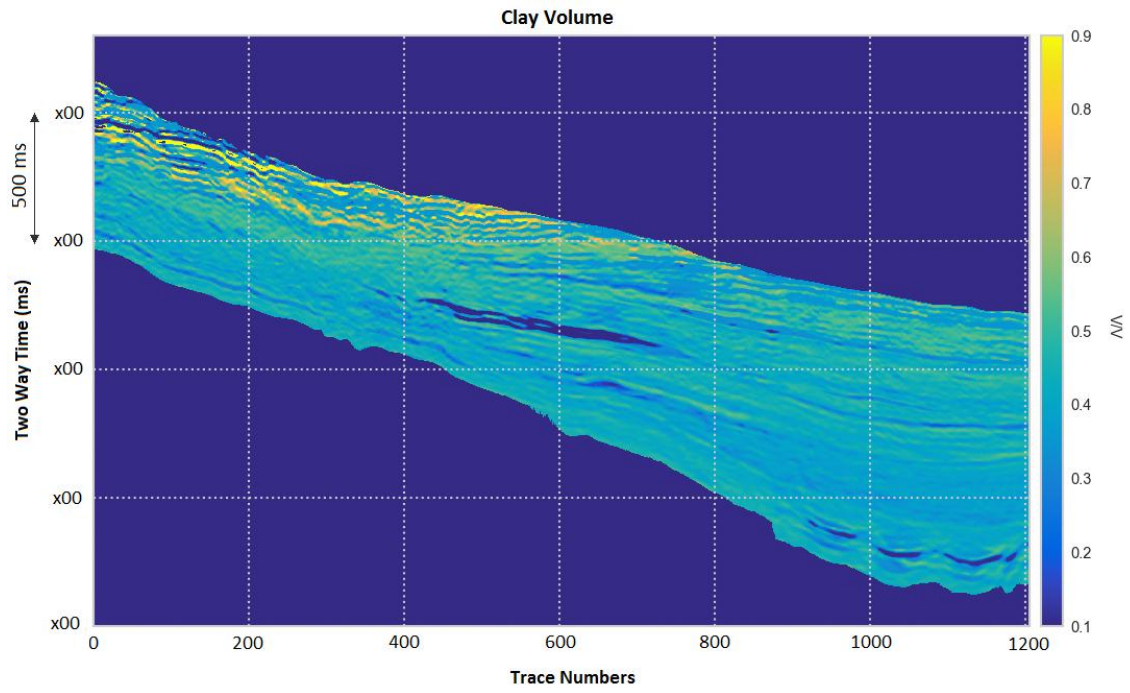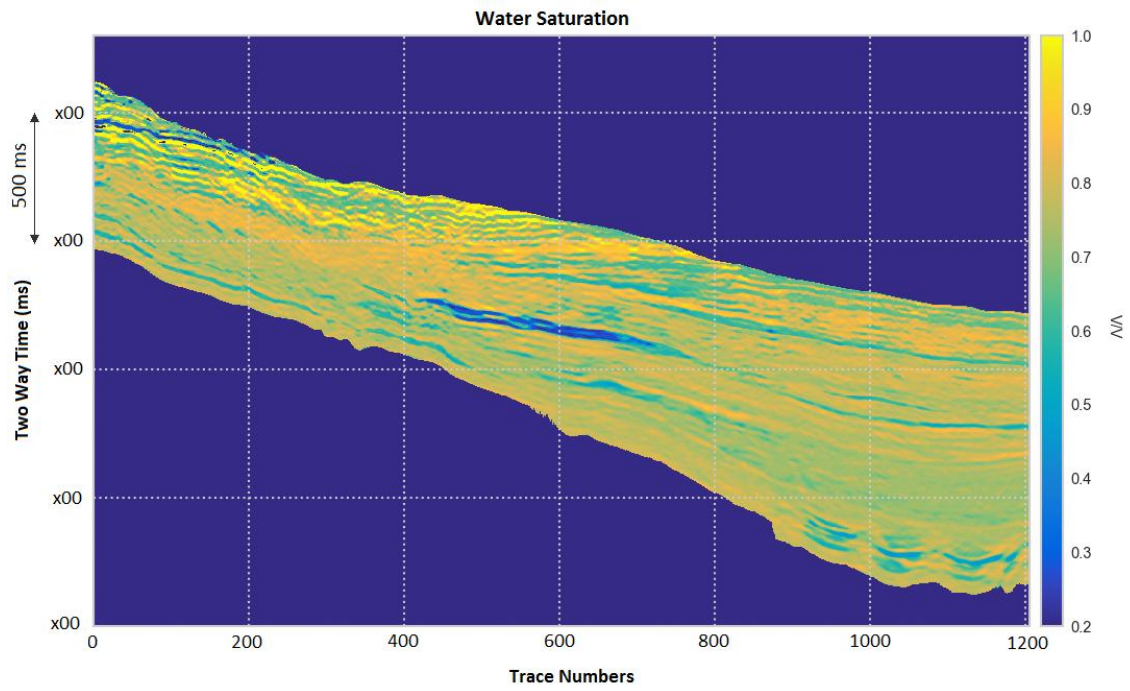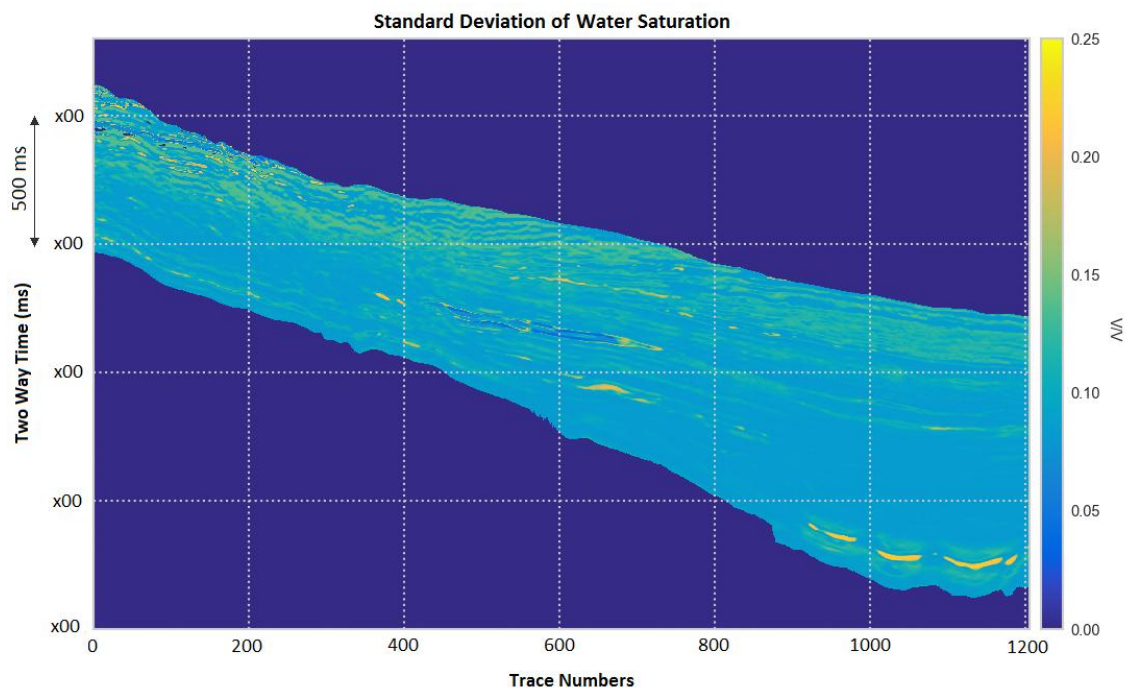*Figure 17: Cell-wise map (a) porosity (φ) and (b) its standard deviations (Std.). Yellow color represents high values and dark blue color represents low values of the respective properties.*
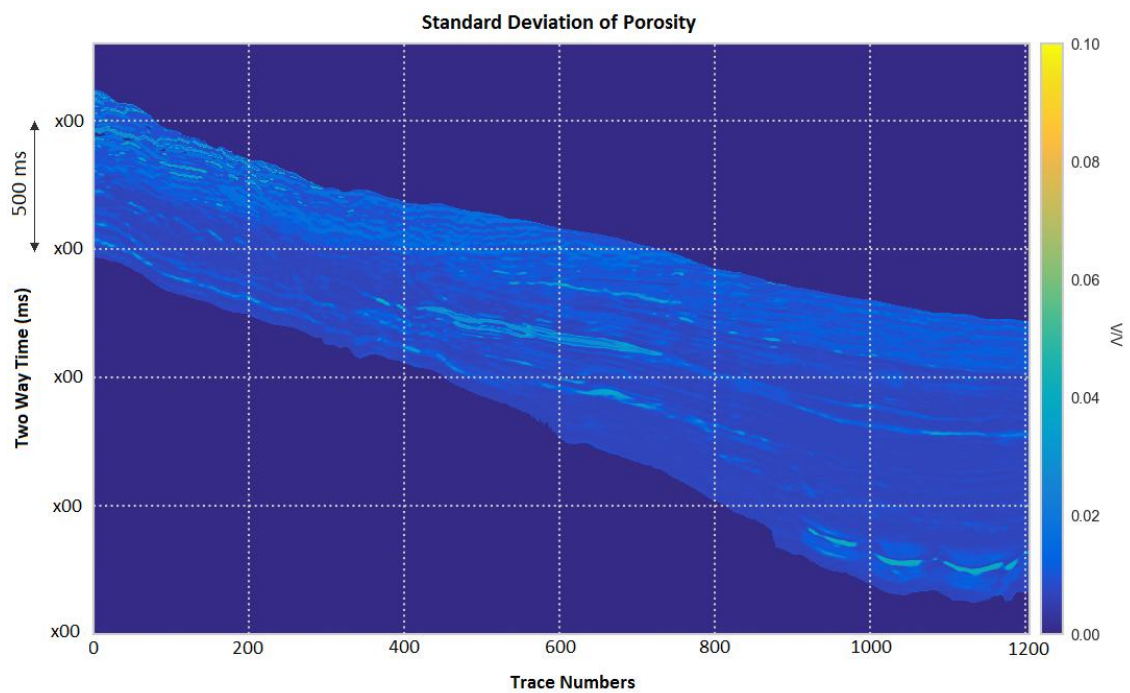
Figure 18: Seismic attributes inversion results at the (a) W1 and (b) W2 well locations. The first three columns display the elastic rock properties: P-wave impedance I_p (IP log), S-wave impedance I_s (IS log) and the P-wave to S-wave velocity ratios V_p/V_s (VPVS log), where the reference elastic well logs are shown in solid-black lines and the seismic attributes used as input to the inversion are shown in solid-red lines. The solid-black curves in columns 4-6 are the reference petrophysical well logs, and solid-red curves are the mean inverted petrophysical properties: clay volume V_{cl} (VCL log), water saturation S_w (SWT log) and porosity φ (PHIT log). Column-7 displays the reference facies interpreted from the well data and column-8 shows the inverted facies. The yellow shaded regions bounded by the dashed-red curves in columns 1-6 represent the 2nd standard deviation of the posterior marginal distributions of the respective rock properties.

34

After verifying the inversion results at the well log scale, we applied our method to invert the available elastic seismic attributes jointly for the spatial distributions of facies and petrophysical rock properties. The limited resolution of the seismic attributes is accounted for within the inversion framework using a boxcar averaging kernel (the regression coefficients in equation 10) whose length is determined by the dominant seismic wavelength. Figure 12 shows the marginal posterior distributions of the three facies and Figure 13 shows the entropy (a measure of uncertainty) of these distributions scaled between 0.0 and 1.0. The entropy is mostly low except at the transitions between different facies, but it appears to be high within some layers too. Since gas-sand has well discriminated properties as seen in the log data, high entropy within some layers indicates presence of mix brine-sand and shale lithology that is not well discriminated. Figure 14a shows the facies map with maximum marginal distributions in each model cell for the three inverted facies: shale, brine-sand, and gas-sand. Figure 14b shows the facies map with an additional facies defined as a combination of non-discriminated shale-sand identified to exist in the cells where entropy is greater than a cutoff value of 0.5 (i.e., 50% of the scaled entropy range from 0.0 to 1.0). Even though we inverted for three facies, the entropy of the marginal posterior distributions identifies that an additional facies may also be interpreted as shaly-sand or sandy-shale shown in brown color in Figure 14b.

The inverted petrophysical properties along with their standard deviations are shown in Figures 15 to 17. The gas reservoir consists of three sand layers (A, B and C), while only two layers are well identified which appear to be merging towards the right in the inversion results, possibly due to limited resolution of the input seismic attributes. The seismic attribute inversion results at the well locations are shown in Figure 18. The measured well logs are shown in solid-black curves for reference. The solid-red curves in columns 1-3 are the input seismic attributes along the boreholes in columns 1-3, and means of the posterior distribution of petrophysical properties in columns 4-6. The yellow shaded regions bounded by the dashed-red curves in columns 1-6 are the 2nd standard deviation of the posterior distribution of corresponding rock properties.

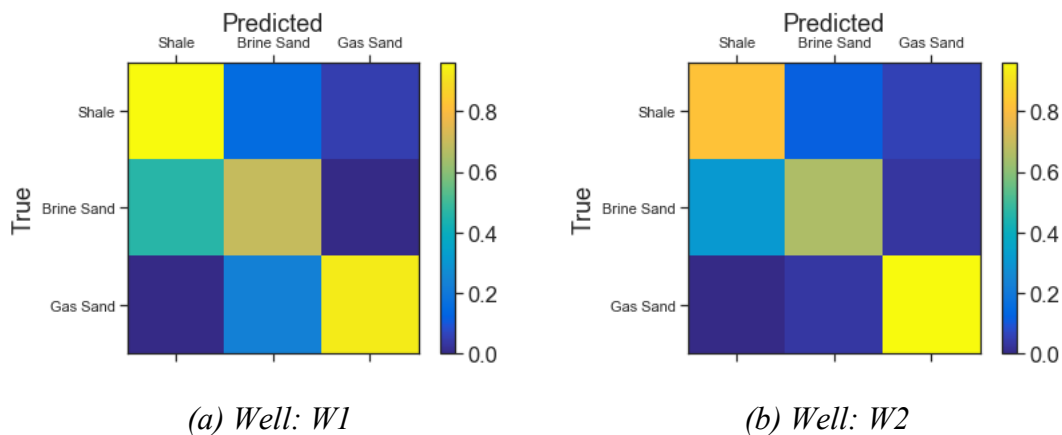(a) Well: W1   (b) Well: W2

Figure 19: Confusion matrix plots for facies prediction from seismic attributes at the locations of wells (a) W1 and (b) W2.

The quantitative analysis of seismic attributes inversion results is summarized in Table 2. The uncertainty is under-estimated for the inverted petrophysical properties in both of the wells (with CR ranging between 0.73 and 0.89). Acceptable correlations (ranging between 0.59 and 0.81) are found between the inverted petrophysical properties and the respective observed well logs. Lower correlations and coverage ratios are mainly due to a significantly lower resolution of input seismic attributes compared to the well logs. Facies prediction rates are very good for gas-sand and shale (between 80% and 96%) and are a bit low (60% and 66% in W1 and W2, respectively) for brine-sand because brine-sand exists mostly in the form of thin layers (Figure 4) which are below seismic resolution. Figure 19 shows the confusion matrix plot of facies predictions at the well locations of W1 and W2. The confusion matrix displays the percentage of predicted facies along columns with respect to the true facies along rows. For example, the element at index [1,1], i.e. top left square, represents the percentage of facies predicted as shale when the true facies is shale. Similarly, the element at index [1,2] (2nd box from left on the top row), represents the percentage of facies predicted as brine-sand when the true facies is shale, and so on. For a good prediction, the diagonal elements must have high values (shown as a color closer to yellow), and the off-diagonal elements must have a low value (shown as a color closer to dark blue).

DISCUSSION

A major contribution of this paper is the development of a computationally efficient inversion method for spatially correlated continuous (petrophysical) rock properties jointly with discrete rock properties (facies), using a sampling-free (i.e. without using McMC) yet fully probabilistic approach. The spatial correlations in continuous rock properties are governed by the spatial continuity of geological facies such that the inversion results honor both the data and the spatial prior information following the Bayesian philosophy.

The presented method avoids the common approach of petrophysical inversion that is based on an explicit use of a forward rock physics model (e.g., Bosch et al. 2009; Lang and Grana, 2018) that defines the relationship between data and model parameters. Contrary to that previous work, a pure data-driven approach does not require any models; the relationship between the data and model parameters is expressed in the form of a probability distribution. Both approaches have their merits and demerits. For example, forward modeling always requires some simplistic assumptions about rock composition and structures which govern their properties. Such assumptions are undesirable when sufficient well data is available, in which case a data-driven approach may perform better. On the other hand, rock physics models are more helpful in interpreting the inversion results.

Our method is primarily data driven; it builds facies dependent joint distributions of all of the continuous rock properties (elastic as well as petrophysical properties) and thus implicitly involves correlations between rock properties without requiring any forward model. However, a forward rock physics model may be used to augment the

existing well data by generating samples of potential reservoir scenarios that are not encountered in the existing wells, or in case of limited availability of well data. Augmenting the existing well data in this manner also ensures that the prior distribution does not over-fit the existing well data, which refers to the case when inversion might perfectly predict model parameters close to the well location but may fail at other locations. Explicit use of forward modeling for solving an inverse problem often requires further assumptions such as linearity of the relationship between data and model parameters (e.g., Grana et al., 2017) for computational efficiency. The presented method makes no such assumptions; it is fully nonlinear and is still computationally efficient.

An additional advantage of the presented method is that the prior joint distribution of elastic and petrophysical properties implicitly introduces prior information on the petrophysical properties. Only the prior information on the facies is separately required, which can be provided in the form of training images. A training image depicts the expected spatial continuity of geological facies which can be modeled using geological process modeling (Griffiths et al., 2001; Hill et al., 2009) or other methods (e.g., Mariethoz and Caers, 2014; Lindberg et al., 2015). Prior information on both facies and the petrophysical properties helps to regularize the nonlinear joint inversion problem.

Mixture density estimation has been widely used in the rock physics or petrophysical inversion literature. Grana (2018) used a data dependent non-parametric kernel density estimation (KDE) method. This approach may be computationally expensive in the case of a large dataset since it requires the fitting of a predefined kernel at each data point. Also, like any other data driven method, KDE is highly susceptible to over-fitting. Parametric distributions (e.g., Gaussian), on the other hand, are often too simple to reliably model a complex probability density function (PDF). In this paper, we used a semi-parametric Gaussian mixture (GM) distribution. A GM distribution is robust enough to capture any level of detail in any complex PDF provided a sufficient number of kernels are used, but it typically requires a much smaller number of parameters compared to a non-parametric distribution, and is therefore less prone to over-fitting.

Shahraeeni and Curtis (2011) used a GM distribution within a mixture density network (MDN) based inversion method for estimation of petrophysical parameters. They used a GM distribution with diagonal covariance matrices. A large number of kernels are required in such a case in order to reasonably represent a distribution with significantly nonlinearly correlated components. For example, P-wave and S-wave impedances are generally strongly correlated. In this work we used Gaussian components with full covariance matrices which capture any correlations among various variables. Such correlations are useful in regularizing an inverse problem in order to mitigate non-uniqueness of the solution. Although a GM distribution with full covariance involves more parameters per kernel, it requires a much smaller number of components to accurately model a given distribution.

A common approach in geophysical literature is to use a GM distribution with one component per facies to be inverted. In this paper, we generalized this approach by using multiple mixture components per facies. This allows the modeling of multimodal

distributions caused due intrinsic variability of rock properties within the same facies, e.g., due to patchy saturation, multiple types of porosity (pores, vugs, and fractures in carbonates), etc.

We presented an application of the method on a real dataset from the North Sea. We inverted a 2D seismic section with restricted depth range under the assumption of stationarity, i.e. the statistical relationship between the rock properties do not vary with location. If, however, a larger subsurface volume is to be inverted, non-stationarity may be a challenge which can be addressed by the introduction of spatial and depth trends in the rock properties, and zonation to account for changing patterns of facies (Mariethoz and Caers, 2014). In spite of such strategies, sufficient sampling of rock properties in the subsurface still remains a critical requirement for reliable inversion in any possible scenario.

In our real data example, the seismic attributes (P-wave and S-wave impedances and Vp/Vs ratios) were obtained deterministically from the seismic waveform data, which do not provide an uncertainty measure in the estimated attributes. Thus, the uncertainty in input attributes due to errors in their estimation process was not incorporated; only the uncertainty due to intrinsic variability of rock properties within each facies was incorporated. This resulted in under-estimation of the posterior uncertainty in petrophysical properties. This suggests that the ignored uncertainties should also be acknowledged for an improved estimation of posterior uncertainties in the petrophysical properties.

The presented method requires a predefined structure of the Markov random field (MRF) which means that the size of the neighborhood is fixed. This approach is similar to sequential simulation methods in Geostatistics that use a predefined template for spatial conditioning of neighboring variables (Strebelle, 2001; Mariethoz and Caers, 2014). A more general approach would invert the neighborhood structure and size along with the model parameters using a hierarchical Bayes approach (Luo and Tjelmeland, 2018). We leave this as a topic of future research.

CONCLUSIONS

We presented a Bayesian inversion method for joint estimation of geological facies and petrophysical rock properties and their associated uncertainties from seismic attributes. We showed that under a suitable set of assumptions that are less stringent compared to most previous research on this topic, we can devise an efficient method to solve the inverse problem. Our method is based on a variational optimization approach which is computationally efficient, allows reliable detection of convergence, and remains computationally efficient in high dimensions (e.g., when inverting 3D seismic data). We also demonstrated with the help of a real data example from North Sea that prior information about the spatial distribution of geological facies helps recover spatial correlation in petrophysical rock properties. Also, the use of a Gaussian mixture model (GMM) for joint distribution of petrophysical rock properties and seismic attributes, allowed us to capture complexity and multi-modality in the distribution of continuous rock

properties. The real data application showed reasonable accuracy of inversion results. However, like most other inversion methods, limited resolution of seismic data and lack of sufficient well data to provide prior information remain potential challenges for this method to produce reliable results.

REFERENCES

Balakrishnan, S., M. J. Wainwright, and B. Yu, 2017, Statistical guarantees for the EM algorithm: From population to sample-based analysis: The Annals of Statistics, **45**, no. 1, 77–120, doi: 10.1214/16-AOS1435.

Beal, M. J., 2003, Variational algorithms for approximate Bayesian inference: PhD. Thesis, University College London.

Berryman, J. G., 1999, Origin of Gassmann's equations: *Geophysics*, **64**, no. 5, 1627-1629. https://doi.org/10.1190/1.1444667

Besag, J., 1974, Spatial interaction and the statistical analysis of lattice systems: *Journal of the Royal Statistical Society, Series B (Methodological)*, **36**, no. 2, 192-236.

Bishop, C., 1995, Neural networks for pattern recognition: Oxford University Press.

Bosch, M., C. Carvajal, J. Rodrigues, A. Torres, M. Aldana, and J. Sierra, 2009, Petrophysical seismic inversion conditioned to well-log data: Methods and application to a gas reservoir: *Geophysics*, **74**, no. 2, O1-O15. https://doi.org/10.1190/1.3043796

Bosch, M., T. Mukerji, and E. F. Gonzalez, 2010, Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review: *Geophysics*, **75**, no. 5, 75A165-75A176. https://doi.org/10.1190/1.3478209

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm: *Journal of the Royal Statistical Society*, Series B (Methodological), **39**, 1–38.

Doyen, P. M., T. M. Guidish, and M. H. de Buyl, 1989, Monte Carlo simulation of lithology from seismic data in a channel-sand reservoir: *SPE paper* # 19588.

Dvorkin, J., and A. Nur, 1996, Elasticity of high-porosity sandstones: Theory for two North Sea data sets: *Geophysics*, **61**, no. 5, 1363-1370. https://doi.org/10.1190/1.1444059

Feynman, R. P., 1972, Statistical Mechanics: A Set of Lectures: Perseus, 1972.

Gardner, G. H. F., L. W. Gardner, and A. R. Gregory, 1974, Formation velocity and density – the diagnostic basics for stratigraphic traps: Geophysics, **39**, 770-780. Doi:10.1190/1.1440465.

Grana, D., and E. Della Rossa, 2010, Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion: *Geophysics*, **75**, no. 3, O21-O37. https://doi.org/10.1190/1.3386676

Grana, D., X. Lang, and W. Wu, 2016. Statistical facies classification from multiple seismic attributes: comparison between Bayesian classification and expectation–maximization method and application in petrophysical inversion: *Geophysical Prospecting*, **65**, no. 2, 544-562.

Grana, D., T. Fjeldstad, and H. Omre, 2017, Bayesian Gaussian mixture linear inversion for geophysical inverse problems: *Mathematical Geoscience*, **49**, 493, doi: 10.1007/s11004-016-9671-9.

Grana, D., 2018, Joint facies and reservoir properties inversion: *Geophysics*, **83**, no. 3, M15-M24. https://doi.org/10.1190/geo2017-0670.1

Griffiths, C., C. Dyt, E. Paraschivioiu, and K. Liu, 2001, SEDSIM in hydrocarbon exploration, *in* D. F. Merrian, J. C. Davis, eds., Geologic Modeling and Simulation: Kluwer Academic/Plenum Publishers, 71–99.

Hill, J., D. Tetzlaff, A. Curtis, and R. Wood, 2009, Modeling shallow marine carbonate depositional systems: Computers & Geosciences, **35**, 1862–1874.

Koller, D., and N. Friedman, 2009, Probabilistic Graphical Models: Principles and Techniques: *MIT Press*.

Lang, X., and D. Grana, 2018, Bayesian linearized petrophysical AVO inversion: *Geophysics*, **83**, no. 3, M1-M13. https://doi.org/10.1190/geo2017-0364.1

Lindberg, D., and H. Omre, 2014, Blind categorical deconvolution in two level hidden Markov models: *IEEE Transactions on Geoscience and Remote Sensing*, **52**(11), 7435-7447.

Lindberg, D., and H. Omre, 2015, Inference of the transition matrix in convolved hidden Markov models and the generalized Baum-Welch algorithm: *IEEE Transactions on Geoscience and Remote Sensing*, **53**, no. 12, 6443-6456.

Luo, X., and H. Tjelmeland, 2018, Prior specification for binary Markov mesh models: *Statistics and Computing*. Doi:10.1007/s11222-018-9813-7

Mariethoz, G., and J. Caers, 2014, Multiple-point Geostatistics: Stochastic Modeling with Training Images: Wiley-Blackwell.

Mavko, G., T. Mukerji, and J. Dvorkin, 2009, The rock physics handbook: tools for seismic analysis of porous media: Cambridge University Press, London

McLachlan, G., and D. Peel, 2000, Finite Mixture Models: Wiley Interscience.

Meier, U., A. Curtis, and J. Trampert, 2007a, Fully nonlinear inversion of fundamental mode surface waves for a global crustal model: Geophysical Research Letters, **34**, L16304, doi:10.1029/2007GL030989.

Meier, U., A. Curtis, and J. Trampert, 2007b, Global crustal thickness from neural network inversion of surface wave data: Geophysical Journal International, **169**, 706–722.

Meier, U., J. Trampert, and A. Curtis, 2009, Global variations of temperature and water content in the mantle transition zone from higher mode surface waves: Earth & Planetary Science Letters, **282**, 91–101.

Mukerji, T., A. Jørstad, P. Avseth, G. Mavko, and J. R. Granli, 2001, Mapping lithofacies and pore-fluid probabilities in a North Sea reservoir: Seismic inversions and statistical rock physics: *Geophysics*, **66**, Special Section, 988-1001.

Murphy, K. P., Y. Weiss, and M. I. Jordan, 1999, Loopy belief propagation for approximate inference: An empirical study, *in* Proceedings of Uncertainty in AI, **9**, 467–475.

Nawaz, M. A., and A. Curtis, 2017, Bayesian inversion of seismic attributes for geological facies using a hidden Markov model: Geophysical Journal International, **208**, 1184–1200.

Nawaz, M. A., and A. Curtis, 2018, Variational Bayesian inversion of seismically derived non-localized rock properties for the spatial distribution of geological facies: Geophysical Journal International, **214**, 845–875. doi: 10.1093/gji/ggy163.

Nawaz, M. A., and A. Curtis, 2019, Rapid discriminative variational Bayesian inversion of geophysical data for the spatial distribution of geological properties: *Journal of Geophysical Research: Solid Earth*, **124**, no. 6, 5867-5887, doi: 10.1029/2018JB016652.

Neal, R. M., and G. E. Hinton, 1998, A view of the EM Algorithm that justifies incremental, sparse, and other variants, *in* M. I. Jordan, eds., Learning in Graphical Models: *NATO ASI Series* (Series D: Behavioural and Social Sciences), **89**. Springer.

Pearl, J., 1982, Reverend Bayes on inference engines: A distributed hierarchical approach: *Proceedings of the Second National Conference on Artificial Intelligence*, AAAI-82: Pittsburgh, PA. AAAI Press. 133–136.

Pearl, J., 1988, Probabilistic reasoning in intelligent systems: networks of plausible inference: *Morgan Kaufmann Publishers Inc.*, 2nd ed., San Francisco, CA, USA. ISBN 1-55860-479-0.

Rimstad, K., and H. Omre, 2013, Approximate posterior distributions for convolutional two-level hidden Markov models: *Computational Statistics & Data Analysis*, **58**(C), 187–200.

Shahraeeni, M. S., and A. Curtis, 2011, Fast probabilistic nonlinear petrophysical inversion, *Geophysics*, **76**, no. 2, E45-E58.

Shahraeeni, M. S., A. Curtis, and G. Chao, 2012, Fast probabilistic petrophysical mapping of reservoirs from 3D seismic data, *Geophysics*, **77**, no. 3, O1-O19.

Shannon, C. E., 1948, A mathematical theory of communication: *Bell System Technical Journal*, **27**, no. 3), 379-423.

Strebelle S., 2001, Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, **34**, no. 1, 1-21.

Walker, M., and A. Curtis, 2014a, Spatial Bayesian inversion with localized likelihoods: an exact sampling alternative to McMC: *Journal of Geophysical Research: Solid Earth*, **119**, 5741-5761.

Yedidia, J. S., W. T. Freeman, and Y. Weiss, 2001a, Bethe free energy, Kikuchi approximations and belief propagation algorithms: Technical report, Mitsubishi Electric Research Laboratories. TR-2001-16.

Yedidia, J. S., W. T. Freeman, and Y. Weiss, 2001b, Understanding belief propagation and its generalizations: *Technical report, Mitsubishi Electric* Research Laboratories. TR-2001-15.

Zhang, X., and A. Curtis, 2020a, Seismic tomography using variational inference methods: Journal of Geophysical Research: Solid Earth, 125, e2019JB018589, doi: 10.1029/2019JB018589.

Zhang, X., and A. Curtis, 2020b, Variational full-waveform inversion: Geophysical Journal International, ggaa170, 10.1093/gji/ggaa170.

LIST OF FIGURES

Figure 1: (a) Representation of a standard gridded (cellular) model, and (b, c and d) a probabilistic graphical model (PGM), where vertices in the latter (shown as circles) represent random variables and the edges (links between vertices) indicate probabilistic dependence between the connected vertices (or the associated random variables). Red circles represent hidden vertices or unobserved variables (model parameters) and the blue circles represent observed vertices (data). (c) A typical HMRF with localized likelihoods (LL) where each unobserved variable is conditioned on the observed variable at the same location only. (d) A HMRF with the quasi-localized likelihoods (QLL) assumption of [Nawaz and Curtis (2018)](#), where the hidden variable at each location is conditioned on the observed variables within a pre-specified neighborhood around that location. In this paper we use the QLL assumption which is a relaxation of the LL assumption. The neighborhood of any hidden vertex (red circle) in (b)-(d) consists of the four hidden vertices that share an edge with that vertex.

Figure 2: A flow-chart summary of the inversion method. Inputs are shown in green color: well data, rock physics model and seismic attributes. Prior information about facies is shown in red color, and estimated quantities and distributions are shown in white color. Rock physics modeling and the corresponding synthetic rock properties for each of the

geological facies are enclosed in a blue box to emphasize that these are optional and may not be required if sufficient well data is available.

Figure 3: Seismic attributes (a) P-wave impedance, (b) S-wave impedance, and (c) Vp/Vs ratios, derived from a selected 2D section of waveform seismic AVO data using a deterministic inversion method. These attributes are used as inputs to our method for the joint inversion of geological facies and petrophysical rock properties.

Figure 4: Well-log data and facies profiles in a well W1 in the study area. Standard well-log pneumonics are used for the well log curves as shown in the headers above the display columns. The color codes for three facies, i.e. yellow for shale, blue for brine-sand and red for gas-sand, are used as standard in all of the subsequent figures in this paper. The well log data from W1 are used as input for modeling the facies dependent prior joint distribution of elastic (seismic attributes) and petrophysical rock properties. Three reservoir layers encountered in W1 are marked with labels 'A', 'B' and 'C' for correlation.

Figure 5: Well-log data and facies profiles in a well W2 in the study area. These data were not incorporated in the inversion process, and were used only for cross-validation (testing) of the results. Standard well-log pneumonics are used for the well log curves as shown in the headers above respective columns. Three reservoir layers encountered in W2 are marked with labels 'A', 'B' and 'C' for correlation.

Figure 6: Crossplots between various combinations of P-wave impedance (I_p) and S-wave impedance (I_s) and the P-wave to S-wave velocity ratios (Vp/Vs) observed in the well log data: (a) I_p versus I_s, (b) I_p versus V_p/V_s, (c) I_s versus V_p/V_s, (d) I_p versus I_s, (e) I_p versus V_p/V_s, (f) I_s versus V_p/V_s. The crossplots are color coded with respect to the volume of clay (V_{cl}) in (a)-(c) and with respect to the interpreted facies (d)-(f). The gas-sand points are well separated from the other facies, while the brine-sand and shale points have a significant overlap.

Figure 7: The training image used to model the spatial prior distribution of facies that is constructed from histograms of various facies configurations found in this image.

Figure 8: Porosity (φ) vs. P-wave velocity (V_p) crossplots with color codes based on the facies interpreted from the well data. The overlaid rock physics template (lines with different shades of grey) correspond to trends for different Net-to-Gross (N:G) ratios predicted using (a) the soft-sand and (b) the stiff-sand model. Each of the two rock physics models are calibrated using different set of parameters: the coordination number $C_n = 13$ *and the* critical porosity $\varphi_c = 0.5$ *for the soft-sand model, and* $C_n = 5$ *and the* critical porosity $\varphi_c = 0.4$ *for the stiff-sand model. This shows that the reservoir can be modeled*

using the Intermediate stiff-sand model ([Mavko et al., 2009](#)), i.e. either by a stiffer soft-sand model or a softer stiff-sand model.

Figure 9: (a)-(c) I_p versus V_p/V_s crossplots in the first row, (d)-(f) I_s versus V_p/V_s crossplots in the second row, and (g)-(i) I_p versus I_s crossplots in the third row. The first column (a, d and g) displays the crossplots using log data from W1. The second column (b, e and h) displays the crossplots using the original well data together with the well data after replacing gas with brine in the sand layers using Gassmann fluid substitution modeling to show the effect of brine on the elastic properties of reservoir layers (A, B and C). The third column (c, f and i) displays the crossplots using Monte Carlo (MC) simulated data using the soft-sand model with intermediate rock physics parameters as shown in Table 1 to simulate a wide range of possible values that might not have been sampled in the well data.

Figure 10: Matrix-plot of samples from components of the prior joint distribution of elastic and petrophysical rock properties. The first three components are the elastic properties: P-wave impedance I_p (IP log), S-wave impedance I_s (IS log) and the P-wave to S-wave velocity ratios V_p/V_s (VPVS log), and the last three components are the petrophysical properties: clay volume V_{cl} (VCL log), water saturation S_w (SWT log) and porosity φ (PHIT log). The diagonal plots represent smoothed histograms of each of the components, and the off-diagonal plots show facies dependent correlations between the respective components.

Figure 11: Well logs inversion results. The first three columns display the input elastic rock properties: P-wave impedance I_p (IP log), S-wave impedance I_s (IS log) and the P-wave to S-wave velocity ratios V_p/V_s (VPVS log), shown in the solid-black lines estimated from the sonic (DTP and DTS) and density (ZDEN) logs shown in Figures 4 and 5. The solid-black curves in columns 4-6 are the reference petrophysical well logs, and solid-red curves the mean inverted petrophysical properties: clay volume V_{cl} (VCL log), water saturation S_w (SWT log) and porosity φ (PHIT log). Column-7 displays the reference facies interpreted from the well data and column-8 shows the inverted facies. The yellow shaded regions bounded by the dashed-red curves represent the 2nd standard deviation of the posterior marginal distributions of the petrophysical rock properties in columns 4-6, and the 2nd standard deviation of the conditional marginals of the joint distribution of rock properties obtained by conditioning on the estimated posterior mean petrophysical properties and integrating out the elastic properties other than the one that is plotted in columns 1-3.

Figure 12: Cell-wise posterior marginal distributions of (a) shale, (b) brine-sand, and (c) gas-sand. Yellow color represents high probability (value=1.0) and dark blue color represents low probability (value=0.0).

Figure 13: Cell-wise posterior marginal entropy of facies classification shown in Figure 12 scaled between 0.0 and 1.0. Yellow color represents high entropy (value=1.0) and dark blue color represents low entropy (value=0.0).

Figure 14: Maps of the facies with maximum marginal distribution in each cell. (a) Map of the three inverted facies: Shale (SH: shown in yellow), brine-sand (BS: blue) and gas-sand (GS: red). (b) Map with an additional facie "Shale/Sand" (SS: brown) identified from high entropy layers in Figure 13.

Figure 15: Cell-wise map of (a) clay volume (V_{cl}) and (b) its standard deviations (Std.). Yellow color represents high values and dark blue color represents low values of the respective properties.

Figure 16: Cell-wise map (a) water saturation (S_w) and (b) its standard deviations (Std.). Yellow color represents high values and dark blue color represents low values of the respective properties.

Figure 17: Cell-wise map (a) porosity (φ) and (b) its standard deviations (Std.). Yellow color represents high values and dark blue color represents low values of the respective properties.

Figure 18: Seismic attributes inversion results at the (a) W1 and (b) W2 well locations. The first three columns display the elastic rock properties: P-wave impedance I_p (IP log), S-wave impedance I_s (IS log) and the P-wave to S-wave velocity ratios V_p/V_s (VPVS log), where the reference elastic well logs are shown in solid-black lines and the seismic attributes used as input to the inversion are shown in solid-red lines. The solid-black curves in columns 4-6 are the reference petrophysical well logs, and solid-red curves are the mean inverted petrophysical properties: clay volume V_{cl} (VCL log), water saturation S_w (SWT log) and porosity φ (PHIT log). Column-7 displays the reference facies interpreted from the well data and column-8 shows the inverted facies. The yellow shaded regions bounded by the dashed-red curves in columns 1-6 represent the 2nd standard deviation of the posterior marginal distributions of the respective rock properties.

Figure 19: Confusion matrix plots for facies prediction from seismic attributes at the locations of wells (a) W1 and (b) W2.

LIST OF TABLES

Table 1: Prior Uniform distribution ranges used for the intermediate rock physics parameters.

Table 2: Accuracy measures for the petrophysical properties and facies inverted at well locations computed with respect to the actually measured (reference) log-curves and facies interpreted from well data. Confidence ratio and success rate are defined in the text.